自我介绍

### ES 相关面试

ES的查询怎么进行性能优化？

（1）调优Filesystem Cache性能的方法：

要让 ES 性能好，最佳的情况下，就是机器的内存，至少可以容纳总数据量的一半。索引数据占用的内存最好控制在留给Filesystem Cache的内存容量中。

要做到上面这一点，可以从两个方面着手改造：

给 Filesystem Cache 更多的内存，尽量让内存可以容纳所有的 IDX Segment File 索引数据文件。

ES 中仅写入要用来检索的少数几个字段就可以了，这样可以缩小单条数据的数据量， Filesystem Cahce 能缓存的数据就越多。可以将其他的字段数据存在 MySQL/HBase 里，然后你从 ES 检索可能就花费 20ms，然后再根据 ES 返回的 id 去 MySQL/HBase 里查询，查 20 条数据，可能也就耗费个 30ms。

（2）数据预热方法：

对于热门的搜索词，后台系统可以定时自己去搜索一下，这样，热门数据就会刷到Filesystem Cache中，后面用户实际来搜索热词时就是直接从内存中搜索了。

（3）冷热分离方法：

将热门数据写一个索引，冷门数据单独写一个索引，这样在热门数据被预热（即写入Filesystem Cache）后，不会被冷数据给冲刷掉。

（4）document模型设计方法：

ES对复杂关联查询不是很擅长，不要在ES中做复杂关联查询，最好在Java中完成关联后直接写入ES。

ES写入怎么进行性能优化？

ES的深度分页问题？

分页性能优化方法：

现在要分页查询，每页10条，要查第100页。若ES有5个shard，那么ES是从每个shard上查出1000条数据，协调节点对这5000条数据进行合并，然后根据需求进行排序、筛选等操作后，再次分页，拿到里面的第100页数据。所以翻页越深，shard返回的数据就越多，协调节点处理时间越长，性能越差。

解决办法：

1、不允许用户深度分页。ES翻得越深，性能越差，深度分页性能很差。

2、使用Scroll API 或者 search\_after来操作翻页。

### Redis 相关面试题

Redis的常用数据结构有哪些？你们项目里面分别用来做什么？

Redis的数据结构类型

Redis有很多种数据结构，包括string、hash、list、set、sortset、hyperloglog、GEO。这些数据结构都可以设置键值的存在时间。

string介绍

string是以一种纯字符串作为value的形式存在的。也是这几种之中使用最多的数据结构。value可以存储json格式、数值型等。

string使用场景

string使用场景一般是存储简单的键值类型。比如用户信息，登录信息，配置信息等。还有一种用得比较多的是string的incr/decr操作，即自减/自增操作。调用它是原子性的，无论调用多少次，都一一计算成功。例如需要增减库存的操作。

尽管string的value可以存储很大，甚至500多MB的容量。但是在性能上来说，我们尽量存储value的值不要过1MB。

hash介绍

hash是一个集合，使用过hash的人都知道，hash的读取性能都一样快。在redis中，hash因为是一个集合，所以有两层。第一层是key：hash集合value，第二层是hashkey：string value。所以判断是否采用hash的时候可以参照有两层key的设计来做参考。并且注意的是，设置过期时间只能在第一层的key上面设置。

hash使用场景

使用hash，一般是有那种需要两层key的应用场景，也可以是‘删除一个key可以删除所有内容’的场景。例如一个商品有很多规格，规格里面有不同的值。

如果需要删除商品时，可以一次性删除‘商品id’的key，则商品里面的所有规格也会删除，而不需要找到对应的规格再做处理。如果查找商品id与规格id1的商品时，则通过两个key查找即可。或者查找所有商品的规格，查找商品id即可。

需要注意的是，经过测试，在性能上来说一般hash里面的第二层key，不要超过200个为佳。尽管hash里面的key-value能达到500多MB的存储容量。

list介绍

list是一个集合，而在redis中，list的使用场景多种多样。在redis中，插入list中的值，只需要找到list的key即可，而不需要像hash一样插入两层的key。list是一种有序的、可重复的集合。

list使用场景

list可以使用左推、左拉、右推、右拉的方式。所以你可以使用list作为集合存储，比如存储某宝商铺里面的所有商品。

也可以用作轻量级别的队列来使用。左推左拉、右推右拉。

需要注意的是尽管redis可以使用推拉的队列模式，但是一定要注意场景。因为redis的队列是一种轻量级别的，没有队列重试、队列重放机制。消费完队列消息在redis代表已经删除了。

消息队列之发布订阅模式

redis除了list的推拉队列模式之外，还有一种是发布订阅的模式。发布订阅模式有是publish/subscribe的命令。与推拉队列模式不同之处在于，订阅者可以订阅多个频道，发布者可以向指定的频道或者所有频道发送消息。

之所以在上面的list中结束这种队列模式，主要为了对比这两种的区别。

set介绍

set是一种无序的，不能重复的集合。并且在redis中，只有一个key。

set使用场景

如保存一些标签的名字。标签的名字不可以重复，顺序是可以无序的。

需要注意的是使用set一定不要存储大量的数据。value的值不宜过大，并且集合数量不宜过大。几百个集合的值，value不超过1MB为佳。

sortset介绍

sortset在redis中是有序的，并且不能重复。既有list的有序，又有set的不可重复性。

sortset使用场景

sortset的使用场景一般是排行榜之类的场景。

clipboard

sortset需要注意的地方与上面的set一致。

hyperloglog介绍

hyperloglog，使用得不多。hyperloglog从创建一开始存储12KB的容量。

hyperloglog使用场景

一般用于统计使用，例如统计页面PV/UV等数据。

GEO介绍

GEO，使用得也不多。用于处理地理位置的信息。

GEO使用场景

在redis中，GEO可以保存地理位置的信息，并且可以计算地理位置的距离等。场景就在于使用地理位置时，并且需要计算，快速的场景，可使用。

Redis的缓存过期和淘汰策略了解吗？LRU和LFU具体有什么区别？Redis是怎么实现LRU和LFU的？

你们怎么解决Redis缓存和DB数据不一致的问题？延迟双删的原理

Zset的底层数据结构了解吗？Zset是怎么实现的？

HashMap源码了解吗？讲讲get和put的过程？HashMap怎么解决Hash冲突？java8的HashMap中为什么要引入红黑树？

HashMap是线程安全的吗？ConcurrentHashMap怎么保证线程安全的？

MySQL隔离级别？不可重复读和幻读的区别？

给了几种情况让MySQL判断索引是不是命中？

新建MySQL索引的时候依据什么原则呢？

有没有遇到MySQL主从的数据一致性问题，怎么解决的？

Jvm垃圾回收算法了解吗？

能讲讲CMS和G1收集器的具体原理吗？哪些垃圾收集器是吞吐量优先的？

你们项目的限流怎么做的？

滑动窗口了解吗？

MySQL事务的隔离级别？

数据倾斜问题了解吗？

你们MySQL的架构是什么？

你了解JVM吗？如果项目死了怎么办？怎么分析与优化？

你对多线程熟悉吗？

多线程

redis在项目中应用及部署方式

mysql索引原理，以及应用场景，如何做sql优化的

rabitteMQ使用场景，在项目中如何使用的。

了解dubbo和zookeeper吗

了解哪些设计模式，平时有实际应用吗？

两个有序链表，如何合并到一个链表中并保证有序

Spring bean的生命周期

Spring beanpostprocesser的意义

SpringCloud Gateway的原理

如果让你设计服务注册中心你会怎么做

ElasticJob的任务分片实现原理

Hashmap的底层实现原理

ConcurrentHashMap的锁机制

线程池的核心参数详解，future的异步通知机制

Integer的缓存

Ngnix的负载均衡策略

谈谈你对CAS的理解

MVCC的原理

Spring MVC是如何初始化的（没答上来，说了下一个请求到来的流程）

说一下Spring中的事务管理，@Transactional注解失效的场景（说了下注解，底层原理，传播行为，失效场景说了两个）

Spring和nginx进程模型有什么区别（具体描述记不清了，Spring不了解，说了下nginx）

nginx进程之间如何通信的（没答上来）

nginx是如何控制并发数量的（没答上来，说了下整个配置文件的三大模块内容，好像是通过lua脚本，因为也问了下是否了解lua）

Dubbo和Spring Cloud的注册中心，注册中心之间的区别（我答的ZK和Eureka的区别）

Dubbo和Spring Cloud使用的协议上的区别（大概是这么问的）

ZK是如何保证高可用的，Eureka是如何保证高可用的（回答了下集群和master选举，Eureka乱说的）

Dubbo的服务端和客户端是如何启动注册的，以及如何通过@Reference注解进行调用的（我答的课上讲的整个调用的过程）

说下Redis的数据类型和底层结构。（答得大概，说的自己也有些乱，当时有些记不清了）

Redis是如何保证高可用的，如果进行主从复制的

用了哪些消息中间件（课程中的，也说了下区别，吞吐量，topic数量影响，时效性，可用性）又继续追问了，为什么topic对kafka影响大，没答上来

消息如何保证不重复发送的（我答的ACK机制，简单说了下，但说不太对，感觉是业务上数据的不重复，也没在解释和追问了）

最后要求写两道题1、有一些字符串，找出其中含有公共前缀，并且最长的那个前缀（我想了下说没太好的思路，提示说用树结构）2、并发十个线程去发送请求，需要接收返回结果，如果响应时间超过2s，就新起线程重新调用（我没写，说了下通过信号量类来实现思路）

jdk1.6升到1.8 还有 tomcat6升到8分别都有什么变化？

2.jdk1.8新特性？

3.为什么要用stream？并发场景使用stream有做压测吗？

4.线程池参数怎么配置？为什么这么配置？做过压测吗？遇到过线程池拒绝的情况吗？怎么处理的？

5.mysql索引类型？

6.需要对索引优化，你会怎么做？

7.explain解释字段含义？

8.项目中如何发现慢sql的？在项目中你是怎么优化的？

9.linux指令?maven指令？git指令？

10.在使用springboot遇到了什么问题？为什么推出springboot,使用springboot的好处？

11.springmvc是怎么解决并发问题的？

12.有经历过jvm调优吗？

1.jdk并发包

2.eureka如何治理服务的？是AP还是CP?redis是AP还是CP?

3.HashMap

4.redis 什么时候出现数据不一致的情况

5.mysql 索引

6.范围查找走聚簇索引还是非聚簇索引？

7.jvm内存结构，栈、堆、方法区

8.项目中QPS多少 RT多少

9.redis用的什么模式？哨兵模式是如何选举的？

10.聚簇索引和非聚簇索引

1.写查询语句 觉得如何创建索引来走索引

2.spring循环依赖说一下

3.gc清理java依赖，和c中是否一样？

4.队列解决堆积和可靠性

5.如果消费者都挂了，重复回队重试发送怎么解决

6.缓存穿透解决

7.zset底层原理

8.跳跃表如何查询

9.mvcc原理

10.update全流程

11.mysql中二阶段提交讲一下

12.索引失效情况

13.索引创建原则

14.redis分布式锁中redission看门狗如何实现续锁

15.手写循环依赖伪代码

16.ioc启动流程

17.数据库如何使用锁，间隙锁是否互斥，间隙锁引起的死锁情况

18.项目突出的成果

19.一千万 数据 a b 联合索引 select . from table where a=1 and b=2 and c=3讲一下查询过程，是否索引有优化，为什么

20.可重复读如何实现

21.幻读如何解决，除了序列化

22.hashmap原理

23.guava cache缓存失效底层

24.线程池聊一下放不下了有什么处理方式

25.hashmap负载因子为什么是0.75

1. 如何进行优雅停机

2. 响应式编程

3. JVM如何实现热加载

4. Cglib动态代理的原理

5. 数据库千万级大表新增字段如何进行平滑升级？

6. 数据库分表的设计？订单系统，日订单量500W如何设计？

7. Redis，1W人在线，进行10人抽奖？Set，List

8. 定义一个countdownlatch，多线程启动之后，计数减少，进入wait状态，直到计数为0后，多线程再同时启动。

一面：

1. redis五种基本数据类型

2. redis为什么那么快

3. es和solr区别

4. 搜索项目qps、数据量、架构说下

5. 说一下tfidf和bm25

6. hashmap jdk8做过哪些优化

7. 抓取动态页如何失效，怎么区分动态静态页，讲一下你们抓取项目，如何维护种子页？rpush,lpoll

8. 多线程了解？然后我分别介绍了Future、threadpool、Forkjoin、semaphore、countdownlatch

二面：

1. JVM内存模型

2. 垃圾回收器用的哪个

3. gc各个算法

4. redis基本数据类型，应用场景

5. zset底层数据结构

6. lucene底层similary、FunctionQuery原理提供哪些方法，如何实现，你项目怎么使用的

7. lucene的索引文件后缀名介绍，分别有什么作用、查询为什么快？跳跃表？

8. kafka如何保证顺序消费

9. 你们搜索项目qps?每天请求量？数据量占用文件多大？内存？

10. 你java看的哪本书？

11. 手写代码：平衡二叉树，给你一个节点怎么查到节点的父节，插入思想、写查询代码。

2、 自定义注解实现原理

3、线程池的使用

4、threadLocal的原理

5、springbean的生命周期

6、CompletableFuture的实现原理

7、线程池的启动原理

2021-0313 KEEP面试 二面

项目介绍

画出系统架构图和主要担任的部分

jvm优化 如何排查

2021-0313 KEEP面试 三面

1、项目介绍 和介绍自己

2、公司的qps 如何排查线上问题

3、jvm优化 如何排查

4、介绍RPC通信原理 画出流程图

5、DUBBO和SPRING CLOUD的区别 是一个东西吗