
抖音 2020 面经汇总

目录

字节跳动抖音部门内推名额, 4 面通过 (3 轮技术面+1 轮 HR 面)	1
字节跳动抖音一面二面面经	2
抖音后台开发社招 2 面凉经	4
字节跳动抖音社招后台开发工程师面经	5
抖音 iOS 客户端面经	7
抖音视觉岗 9.9 面经	9

字节跳动抖音部门内推名额, 4 面通过 (3 轮技术面+1 轮 HR 面)

来源: <https://zhuanlan.zhihu.com/p/85800038>

抖音内推一面

自我介绍

项目介绍

用到了什么 rpc 框架? 那 coral 是基于什么协议的? http 协议的 header 是什么结构的?

TCP 和 UDP 的区别

TCP 怎么保证可靠传输的?

如果 tcp 连接断了 (比如 server 宕机), client 如何处理?

介绍一下链表

单链表和双链表的区别及应用场景

MySQL 索引

操作系统和数据库了解么?

HashMap 的底层实现?

数组存储的元素是什么?

详细介绍一下 put 这个方法的执行流程

hashCode 返回 32 位的 int 结果

与数组长度取模 (按位与), 计算出 index, 插入 index 下的链表

volatile 的作用

如何保证可见性?

算法题: 输入一个数据流 (4,3,10), 请依次输出当前数据的中位数, 声明一个链表去存储这些数据, 并返回链表的中位数 (ON)

给一个很大的 log file, 形式是 user id, login time, logout time, 如何找到峰值。

抖音内推二面

聊项目

Java 的 hashmap

Java 中用到的锁

一个 class 中有两个加锁的方法, 可以同时分别被不同的线程调用么?

如何保证缓存与数据库读写一致性?

针对千万次每秒的访问, 分布式缓存如何实现? 主从换从如何保证一致性

Git rebase 与 Git merge 的区别

算法题: 打印二叉树的左边界

抖音内推三面

Java Servlet 和 Filter 的比较

100 层楼, 给两颗鸡蛋, 找到鸡蛋在哪层会破。要求最坏情况下次数最少的方法。

MySQL 的 SQL

多个服务器间共享 session 的解决方案

你学过 redis? 介绍下 redis

二叉树转双向链表

逻辑题: 岛上 n 个人, 有人红帽有人白帽, 每人只能看到别人头上的帽子, 且彼此不能交流, 如若确认自己是红帽子的话可以离开岛。有经过的渔民说了一句“你们中至少一人是红帽!”

请问接下来发生什么情况?

你有什么想问的吗?

抖音内推 HR 面

自我介绍

未来的职业规划

说一下自己平时的学习方法

你认为这些学习方法里最有效的是哪一种?

评价一下之前的面试官, 或者说之前的面试官有没有给你留下印象最深刻的一点

为什么想要来字节跳动?

你平时都用字节的哪些产品? 有什么好的建议吗?

期望薪资

你有没有什么想问的?

字节跳动抖音一面二面面经

作者: 闯江湖

链接:

<https://www.nowcoder.com/discuss/363421?type=post&order=create&pos=&page=1>

来源: 牛客网

抖音一面

说一下二叉树，满二叉树，bst avl b+树 b 树，红黑树的区别

说一下数组 链表 哈希表的使用场景，时间复杂度，优劣

说一下操作系统用户态和内核态

说一下操作系统虚拟内存

说一下 Linux fork()系统调用的过程，要求讲到内存过程

说一下 Linux 无名管道 命名管道的内存过程，底层原理

无名管道产生的过程，原理

Java 说出你知道创建字符串的方式，比较性能优劣

说一下 jvm 线程调度过程，什么情况会发生抢占

说一下线程通过继承 thread 和实现 runnable 接口实现的线程，底层有什么区别

说一下 cookie 和 session 的区别 场景的选择

cookie 跨域

说一下 xss 攻击

算法：

一个长度大小为 n 的数组，数组中的每个元素的取值范围在 $[1, n]$ ，且为正整数。

问：如何在时间复杂度为 $O(n)$ ，空间复杂度为 $O(1)$ 的条件下，统计数组中不同元素出现的次数。

<https://blog.csdn.net/hjwang1/article/details/81048396>

抖音二面(90 分钟)

说一下 Java 内存模型

Java 内存交互操作

如何保证数据一致性？

volatile 如何保证数据一致性？

volatile 的变量两个线程一起操作会发生什么？联系对象头。

Intel cpu 的数据一致性协议是如何保证数据一致性的？

为什么要有高速缓存，寄存器，内存？

高速缓存的作用？

寄存器发生了什么？

操作系统系统的读者写者问题？（怼了 15 分钟）读会发生什么，多个线程读怎么样，如果有线程要写，如何保证写线程不被饿死？

地址空间分配的时机？

操作系统内存管理？

什么时候进行地址空间分配？

函数调用局部变量？内存分配发生了什么？

如果向操作系统请求一个 4g 的连续空间会分配吗？过程？

数据库如何保证数据一致性，如何解决？

redis 分布式锁？

B 树和 B+ 树的区别性？

innodb 索引原理？

myisam 的索引原理？

为什么 myisam 支持压缩表？
myisam 索引的组织方式？
稠密索引是个啥
最左前缀使用：
以下哪个能用索引
（列了四个说说为什么）
说一下 icmp 用来干嘛？
ping 原理，ping 一个地址，通了会怎么样，不通会怎么样（现象）

算法
相交链表 多种解法如何优化？
bst 的第 k 个节点？中序遍历。
最长无重复子串？如何记忆优化，写一下后面用 hashset 写了 面试官不满意 问我你不是说要用 hashmap 优化吗
看着面试官的笑容逐渐消失。。感觉不太行了

抖音后台开发社招 2 面凉经

作者：DEE +
链接：
<https://www.nowcoder.com/discuss/346051?type=post&order=create&pos=&page=1>
来源：牛客网

一面
CtrlC 可以终止程序，我们是否可以在程序终止的时候做一些别的操作？
tcp 四次挥手
进程和线程的区别
进程间通信的方法都有什么
100Mbps 的带宽三个人使用，每人 50Mbps，tcp 怎么保证速度的
1000qps 单机怎么处理？
算法题：一个超级长的数字（有几十位那种），允许进行任意多次的任意两位交换，得到的比他大的最小的数字是多少

二面
为什么想来抖音
简述分页分段机制
http 的 501, 502, 503, 504 什么意思
操作系统进程状态有什么
nginx 的 location 是什么
http 和 https 的区别，为什么不能从 http 升级为 https
pyc 是什么，什么时候生成
算法题：实现一个 LFU***，要求实现 set 和 get 方法，复杂度为 $O(1)$ 。

(该题的 LFU***是一个存储空间, 有容量, 是整数 n , 存 key-value 对, 每访问一次这个 kv 对的频率就加一, 如果新 set kv 对 的时候, 已经达到容量上限, 删除频率最低的 kv 对)

然后告诉我我挂了

字节跳动抖音社招后台开发工程师面经

作者: loginin

链接:

<https://www.nowcoder.com/discuss/327679?type=post&order=create&pos=&page=1>

来源: 牛客网

1. 一面 (1 小时)

一面主要是基础考察, 包括简历中提到的以及一些标准的基础问题。

VUE 和 React 区别 (因为简历中提到了 VUE)

VUE 的内部机制 (问一下前端技术是因为我简历提到自己是全栈工程师, 正常情况下是不会出现前端问题的)

CRSF 的机制

Spring IOC 和 BEAN 循环依赖

Kafka 和 Cassandra 内部机制 (简历中提及)

算法题: 有 $2n$ 个人, 序号为 0 到 $2n-1$, 要求两两握手, 但是握手不能存在交叉线, 求最后一

共存在多少种握手可能, 写出 $f(2n)$ 表达式。

思路: 假设 0 号和 i 号点握手, 那么整张图就分为了 $0 \sim i-1$ 和 $i+1 \sim 2n$ 这两部分。

算法题: 给定一个二维整数矩阵, 找出最长递增路径的长度。对于每个单元格, 你可以往上, 下, 左, 右四个方向移动。 你不能在对角线方向上移动或移动到边界外 (即不允许环绕)。

在另一个事业部一面时, 遇到了三个水壶的问题: 三个水壶, 容量为 $8L$, $3L$, $5L$, 给一个 num , 判断能否量出这个指定 num 的水。

2. 二面 (1 小时)

项目介绍 (半小时)

准备好一些工作中解决的问题, 尽可能详细的阐述给面试官, 告诉他你主要解决了什么问题。

系统设计题：上海地铁线纵横交错（线路与线路之间存在交叉，并且每一条线路站与站之间的发车间隔可能是不一样的），现在做一个 app，输入是起点和终点已经出发时间，输出一条耗时最短的线路。

从数据库表，到缓存，到算法的设计。

3. 三面（40 分钟）

三面是抖音上海负责人的面试，因此着重在系统设计。

项目介绍（15 分钟）

系统设计：抖音里每秒都产生大量视频，现在要求持续计算一段时间内的 top k 个最热门的视频，你怎么设计。

这个问题类似于下面这个问题：

实现前 5 分钟，1 小时，24 小时内分享最多的 feed

从算法的角度，可以简单的称之为 Top K Frequent Elements in Recent X mins。算法的角度，本质就是设计一个数据结构，支持给某个 key 的 count+1（有一个 post 被分享了），给某个 key 的 count-1（有一个分享的计数已经过期了），然后查询 Top k。

做法是维护一个有序序列（用链表来维护），每个链表的节点的 key 是 count，value 是 list of elements that has this count，也用 linked list 串起来。比如 a 出现 2 次，b 出现 3 次，c 出现 2 次，d 出现 1 次。那么这个链表就是：

1

{3: [b]} --> {2: [a -> c]} --> {1: [d]}

然后另外还需要一个 hashmap，key 是 element，value 是这个 element 在链表上的具***置。因为每一次的操作都是 count + 1 和 count - 1，那么每次你通过 hashmap 找到对应的 element 在数据结构中的位置，+1 的话，就是往头移动一格，-1 的话，就是往尾巴移动一格。总而言之复杂度都是 O(1)。当你需要找 Top K 的时候，也是 O(k)的时间可以解决的。

算法实现请参考：

<https://github.com/cheergoivan/leetcode/blob/master/src/leetcode/problem460/LFUCache.java>

工程角度的优化：

如果我要去算最近 5 分钟的数据，我就按照 1 秒钟为一个 bucket 的单位，收集最近 300 个 buckets 里的数据。如果是统计最近 1 小时的数据，那么就以 1 分钟为单位，收集最近 60 个 Buckets 的数据，如果是最近 1 天，那么就以小时为单位，收集最近 24 小时的数据。那么也就是说，当来了一个某个帖子被分享了 1 次的数据的时候，这条数据被会分别存放在当前时间(以秒为单位)，当前分钟，当前小时的三个 buckets 里，用于服务之后最近 5 分钟，最近 1 小时和最近 24 小时的数据统计。

你可能会疑惑，为什么要这么做呢？这么做有什么好处呢？这样做的好处是，比如你统计最近 1 小时的数据的时候，就可以随着时间的推移，每次增加当前分钟的所有数据的统计，然后扔掉一小时里最早的 1 分钟里的所有数据。这样子就不用真的一个一个的+1 或者-1 了，而是整体的 +X 和 -X。当然，这样做之后，前面的算法部分提出来的数据结构就不 work 了，但是可以结合下面提到的数据抽样的方法，来减小所有候选 key 的数目，然后用普通的 Top K 的算法来解决问题。

另外，在分布式环境下，哪些帖子被分享了多少次这些数据，首先在 web server 中进行一次缓存，也就是说 web server 的一个进程接收到一个分享的请求之后，比如 tweet_id=100 的 tweet 被分享了。那么他把这个数据先汇报给 web server 上跑着的 agent 进程，这个 agent 进程在机器刚启动的时候，就会一直运行着，他接受在台 web server 上跑着的若干个 web 进程(process) 发过来的 count +1 请求。

这个 agent 整理好这些数据之后，每隔 510 秒汇报给中心节点。这样子通过 510s 的数据延迟，解决了中心节点访问频率过高的问题。

还可以通过数据抽样进行优化。因为那些 Top K 的 post，一定是被分享了很多很多次的，所以可以进行抽样记录。如果是 5 分钟以内的数据，就不抽样，全记录。如果是最近 1 小时，就可以按照比如 1/100 的概率进行 sample。

最后是使用 Cache 进行缓存计算结果。对于最近 5 分钟的结果，每隔 5s 才更新一次。对于最近 1 小时的结果，每隔 1 分钟更新一次。对于最近 24 小时的结果，每隔 10 分钟才更新一次。用户要看结果的时候，永远看的是 Cache 里的结果。另外用一个进程按照上面的更新频率去逐渐更新 Cache。

总结：算法方面使用 LFU 算法来解决。而工程方面使用分段统计，抽样法和 Cache 缓存进行优化。

以上方案参考自：<https://www.jiuzhang.com/qa/219/>

4. 总结

网络和操作系统需要复习，复习一些常见的网上都能搜到的字节考题就行，不用太深入。

算法 leetcode 刷 200~300 道差不多了，另外面试前搜一下字节考过的算法题，很有可能遇到相同的。

系统设计题就看平时积累，有想法就说出来，即使不是最优解也要说出来。

抖音 iOS 客户端面经

作者：一只程序媛酱

链接：

<https://www.nowcoder.com/discuss/286762?type=post&order=create&pos=&page=1>

来源：牛客网

抖音（意向）

=== 一面（50~60min） ===

算法

会一行一行讲，也会问时间复杂度，以及优化

最近公共父节点（两个节点，都可能为 null，都可能不在树上。我最开始写的没考虑不在树上的情况，面试官还吓唬我说“你以为你这算写对一半吗？没做出来就是 0 分哦”）

反转链表（常规题）

其他

HTTP 了解吗？HTTP2.0 了解下

iOS9-13，每个版本讲一个特性

怎么实现多个任务执行完后，再统一处理（应该是同步阻塞、栅栏函数、调度组。中间那个没答）

夜间模式怎么实现的？除了通知的方式，有什么更优雅的方式吗？

图片怎么加载的？（SDWebImage）

不让用 SDWebImage，如何实现类似逻辑（自己设计图片请求和缓存，简单讲了 SDWebImage 的思路）

每个 TableViewCell 都有一个 UIImageView 要加载图片，如果没有缓存，有五个 Cell 请求一个 url，同时发请求肯定是不实际的，如何处理？

微信 TableView 滑动的时候，动图是不动的，为什么？（RunLoop 的 Mode 决定的）

有了解过 WebSocket 吗？（没有…）

=== 二面（1.5h） ===

算法

如何不用递归，打印所有一个 UIView 所有的 subviews（包括孙子 view 等）

设计一个图片缓存系统，并代码实现

清理缓存应该用 LRU，当时的我并不会，但是被一步步引导着写出来了，感谢面试官没放弃我 hhh

其他

大部分是计网+多线程

介绍研究生做的基于 LLVM/Clang 的静态分析（面试官最后默默来了句，还算是讲清楚了）

七层协议栈

HTTP 是哪一层的协议

HTTP 常见的返回码有哪些？

HTTP 除了 GET、POST 还有哪些？（UPDATE、DELETE、HEAD）

HTTPS 介绍一下

HTTPS 能防止中间人攻击吗？（答了可以，答案应该是可以）

TCP 和 UDP 的区别

线程和进程的区别

线程之间如何进行通讯

不使用 gcd，如何保证线程安全？

怎么实现多个任务执行完后，再统一处理（重复）

iOS 有哪些多线程方式？（NSOperation、NSThread、GCD）

同步和异步有什么区别？

串行队列和并行队列有什么区别？

队列和线程的关系

死锁是怎么产生的？（四个条件只说了三个…）

你有什么要问我的（第一次跟面试官聊规划，感恩面试官没有敷衍我）

=== 三面（40+min） ===

算法

合并排序链表

其他

HTTPS 用的对称加密还是非对称加密？

TCP 的慢启动，快恢复介绍下

客户端发送网络请求，到拿到响应的整个过程（类似打开一个网页的过程）

接上题，如果 DNS 解析的时候，某个 ISP 服务器没有了，这时候逻辑是怎样的？（不会，面试官说会通知其他节点，还提到了 ARP 协议…）

介绍了下简历上写的字典转模型的设计+夜间模式的设计

为什么不用成熟的字典转模型的三方？有了解过吗？（了解一点，有 YYModel，但没用过）

夜间模式有什么缺点，能怎么优化？（从初始化的时候设置值，改为用配置文件。再就不会了）

锁有哪些？（只答出了 NSLock、SpinLock…）

遇到哪些小坑？内存相关的有吗？

我看你简历写了数据存储，怎么存的（FMDB）

自己写的 LoadingView 怎么做的（用 UIImage 的 animationImages…）

线程安全错误是怎么出现的？（面试官解释：先读到寄存器…就出现了错误）

你有什么要问我的（聊了很多，感恩）

我加下你微信有事可以随时找我（是团队 leader，后续还找他帮忙了，一万个感恩）

字节跳动提前批，最早的一场面试啦，凭运气+还算比较好的基础+良好的沟通能力通过了。

小伙伴们，沟通真的很重要~

抖音视觉岗 9.9 面经

作者：death&love&robot

链接：

<https://www.nowcoder.com/discuss/252124?type=post&order=create&pos=&page=0>

来源：牛客网

2019/9/9 抖音一面

(1) 项目相关：

主要介绍了实习项目基于 GAN 做图像补全的项目：包括网络结构、损失函数、实际效果、指标

问：网络损失函数的超参数是怎么调的、这个指标是什么意思、

解释一下谱归一化、

你是否用过除了上采样以外的超分辨率的方法？Sub-pixel Convolution、其中 Sub-pixel Convolution 输出的结果是否改变了 channel 维度（是的因为替代反卷积的一个操作）

解释一下 WGAN 解决了什么问题，原理是啥

你常用的数据扩增方法

如果给你一个新的场景，你会如何使用图像算法进行补全？之前训练的是否又效果

(2) 有关 Pytorch 实战的问题：

`torch.Tensor` 和 `torch.tensor` 有什么区别

`torch.no_grad` 和 `required_grad=False` 的区别

(2) 代码题

<https://leetcode.com/problems/longest-duplicate-substring/>

等了几分钟没有思路，换了写快排

(3) 机器学习

高斯混合模型和 K-means 的区别和联系

(4) 计算机基础

栈和堆的认识

如果有一个函数 f 他调用了另一个函数 $p(a,b)$ 如果让你来写这个编译器你怎么实现（其实就想让你说这个参数咋传递，压栈的过程）

交换 a 和 b 两个数，不借助第三个变量