

写此贴主要是为了感谢在面试中从牛客得到的经验，反馈给后来者。

本人是两年大数据平台开发经验跳槽的。~~

一面（大约6.1，一个小时）：

1. 自我介绍~
2. 介绍下第一个项目（大数据平台的升级 hadoop的升级）
3. hadoop为什么要从2.x升级到3.x?
4. 不停机升级过程中有哪些不兼容的地方（namenode的editlog、datanode的块布局等），你是怎么解决的，社区是怎么解决的，版本回滚会有啥问题？
5. hdfs写数据过程，写的过程中有哪些故障，分别会怎么处理？dataqueue和ackqueue如何交互的？ackqueue中的数据如何重新放回到dataqueue中？如何保证有序？写数据的过程中是一个个package写入并确认成功后再写下一个吗？不过不是，是怎么写的，除了错误如何恢复的？
6. 改过yarn和hdfs的哪部分源码吗？为什么这样改？聊聊你的改的是哪个类？效率怎么样？（我这边改过hdfs中datanode退役相关模块的代码和yarn中parent类中一些代码）
7. 聊聊QJM吧，这边paxos算法怎么用的，你知道为什么字节不用QJM而是使用bookkeeper吗？
8. 你知道字节在hdfs和yarn上有哪些改进吗？
9. 聊聊hashmap和concurrenthashmap的区别和细节
10. 说下G1垃圾回收器的原理
11. 算法 LRU ----不能使用LinkedHashMap，要自己实现底层结构。

总结：1面问的东西比较细，主要是考察了hdfs的源码细节和自己改动的地方，没有超过简历的范围，但是深度比较大。算法题不让使用现成的结构，写起来比较麻烦，面试官还是比较有耐心的。

二面（大约6.10 1个小时）：

1. 自我介绍
2. 讲下第二个项目（prometheus监控相关的），问的很细，很多
3. 聊聊 yarn是如何保证数据一致性的（CAP），主要问的resourcemanager
4. 聊聊hdfs的数据一致性和租约管理以及RBF
5. 说下kafka你知道的东西
6. 说下raft、zab和paxos

7. 说下JVM中复制算法和标记清除算法、标记整理算法的区别，你分析下为什么复制算法不能用到老年代，有啥问题等
8. 说下想来字节的原因
9. 算法题：凑零钱和最大正方形面积（leetcode有）

总结：面试官很nice，很有激情，问题还好，准备足了，应该不难，只是挺考验对知识的掌握度以及对社区最近发展的关注情况。

三面（6.20左右，一个小时）

1. 不用自我介绍，请说下你的三个项目，以及你的思考。
2. 你在当前公司的工作，以及时间安排。
3. 说下kafka你平时杂用的，都了解哪些
4. 算法（随口出的字符串处理题，不难，然后再调优下）
5. 你有啥想问的，前两面面试官考察的挺细的了，我就不问啥了，你有啥想问我的。

总结：面字节问的都很基础，也很深，简历的东西写上去了，一定要会，而且要有自己的思考。平时要总结，以及了解组件的版本特性以及社区发展的方向等。还有，一定要刷题，不然真的过不去。