中软国际 华为外包 华为坂田基地E区东门

面试题:

1. 自我介绍
2. 介绍项目流程
3. 每天的数据量多大
4. 集群配置，从数据采取一直到数据存储
5. 100G数据怎么清洗，怎么优化
6. 对Hadoop的理解
7. Hadoop的优缺点
8. 介绍一个熟悉的项目
9. 围绕项目展开业务上的问题，几个人写的，什么时候写的，周期多长（好像和简历上的时间核对）
10. 项目中遇到的难点，怎么解决的
11. 说说你项目中最熟的位置
12. TopN的实现
13. 临时出的面试题，用spark实现，图片上，由左边输入得到右边输出。
14. 数据倾斜遇到过么，怎么产生的，怎么解决的？
15. OOM发生过么，怎么解决的？
16. 对JVM的理解，调优
17. 应用过机器学习吗？
18. 除了调库，还做过其他的吗？
19. Spark什么版本
20. Spark2.0的新特性
21. Flink和Spark Streaming的区别
22. 一个URL文件50亿行，一行64K，32G内存的服务器，怎么查到只出现一次的URL,如果只有2G内存呢？
23. Hive的理解
24. 项目业务的实际意义你觉得在哪儿

面试经过：

一共有4个面试官，两个外包的，两个华为的。轮流面试的，对业务这方面问的比较多，会让你一个项目的架构，中间不断提问，有技术的，有业务的。最后一个面试官级别有点高，很和蔼，但问得很细，问三年工作都做了什么，听到感兴趣的就会打断往细了问，最后也问了项目流程，项目的价值在哪儿？可能是我的项目都是通信业务场景，所以业务问的比较多，他们也是做通信场景业务的。

每个面试官聊完之后，觉得可以的话都会和你介绍项目组的情况，集群有1000多台，数据量300T，20多人的小组，总共是200多人大数据团队。之后也会让你问他问题。

面试总结：

面试题大概是这么多，技术和业务问的时间大概各占一半，项目介绍，实现过程，都会问的，建议大家多熟悉熟悉自己的项目，自我介绍也尽量完善点，问项目业务被拆穿的话，

不要慌，我们公司就这样的；问到技术不懂的，懂的就把自己懂的说出来，真不懂的话就直接说是盲点，乱扯一通，估计印象会更不好。以上是个人见解，仅供参考。

