T&D 大数据的面经

1. java有哪些常用的类库？Java的多线程方面的问题

2. 口述scala版本的wordcount代码，流程？scala的匿名函数？

3. scala函数的返回值能不能是函数，如果可以，举个例子

4. secondly namenode的作用

5. hive和hadoop的区别

6. 我项目中的一些具体细节问题

如Spark的提交参数有哪些，资源怎么配？项目做了什么，负责什么模块

7.业务场景的问题，怎么做怎么优化之类，1T文件怎么排序，ABCDE等多个列球最大值等？

8. rdd分区的问题： dataset一个2个分区，一个3个分区，ID为key有多少分区？100m文件需要多少partition处理？

9. kafka的文件存储格式

10.

spark.table(\*\*.sql).drop("cloumnA=value").filter("cloumA").count();

spark.table(\*\*.sql).filter("cloumA").drop("cloumnA=value").count();

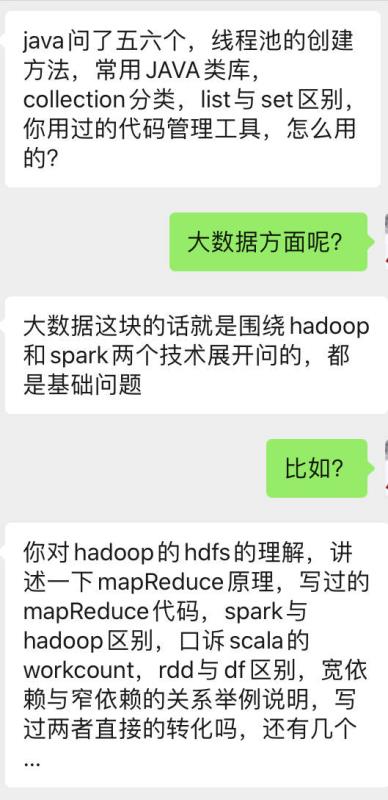
这两句有没有报错？

11.例如2T的文本文件只有一列数据，如何排序?

技术问题（难度大一点）：

1、服务器里面有2Tb的数据，用spark计算得到排序后的结果文件，怎么做？说两个方案的思路；

2、执行spark-shell 做count操作报OOM Error 为什么，如何解决？



sparksql大表加载内存溢出如何解决，spark加载两个表，一个表2个分区，一个表3个分区，两表join之后总共多少个分区，sql行转列问题

全栈的面经

项目中遇到的困难，如何解决（了解问题的解决思路），根据我候选人说的问问题，比较随机。

分布式定时任务，多线程使用，(用多线程进行单元测试如何保证数据一致）

redis哨兵模式以及redis的参数配置

Melody D D Melody组

后端技术类（多线程，微服务设计，jdk特性），数据库调优（索引），前端开发的一些了解，项目开发流程，项目分支冲突管理

EQ的面经

技术主要是问了我jvm内存调优，垃圾回收算法。分布式事务，多线程并发，还有系统异常相关的问题

Emily面试经验

Raymond 组：会考察代码的重构，单元测试，SQL数据库的优化，Linux/Unix的命令，前后端的框架的使用及原理

刚开始问的操作系统这一块，然后还有 通信协议，微服务，项目相关技术

HSS D&D Johnny

