面试总结

**3.30 德高行（专利这块的服务，属于创业型公司，数据来源爬虫，去买数据，所有的数据都是历史数据，数据量大概是1TB,所用的数据库是NoSQL,岗位偏数据分析）**

1 首先自我介绍

2 优化实例

 a SQL语句的优化

 回答了去重不用distinct，那你们用什么去重？回答group by。group by 怎么去重？回答字段里面重复的值知会返回一个。

 b 建立索引

 c 表分区

 d hints，并行优化

3 如果说你们按照时间范围来进行分区,你们多少数据量进行表分区，如何进行分区的？聊聊hash分区

4 有用过组合索引吗？（我说的唯一索引根据用户ID）

5 游标和自定义函数，说说你的理解

6 说说你对大数据的理解？（回答数据量增大，逐渐由传统关系型数据转型成Hadoop以hdfs分布式存储方式...）那你知道阿里是用的mysql的数据库吗？为什么一定要替换呢？

7 死锁知道吗？

8 python里面的浅拷贝和深拷贝区别在哪？

9 python里面bin包知道吗？

10 MySQL和Oracle里面的分页操作分别是什么？

11 数据同步方式？多久同步一次？