1. 数仓架构

答：面试宝典

1. 你们搭建数仓肯定有模型和主题嘛，那你们的主题有哪些啊？

答：主题愣是没想起来

1. Oracle数据库，没有借助工具，我直接让你去写像运行shell脚本，比如说导入的命令，导出的命令。

答：不清楚，我们更多的是用PLSQL写存储过程，然后用调度工具去调度存储过程。

1. 我想了解一下，你写存储过程有一步是那个，统计信息的收集嘛，你对统计信息有了解吗？

答：我可能没有get到你的点..

面：统计信息的话相当于你在写存储过程，写步骤的时候，现在存储过程可能有N步，每一步的话你都会对这一步的，像执行语句后+信息收集，像数据库的表，索引的大小和规模，然后像一些数据的分布块下，这些其实都是方便你在后续的时候，你在用子计划的时候可以存本，执行成本最小的计划，所以你们在写存储过程的时候没有去考虑这一个吗？

答：瞎扯，我们会看存储过程的整体的执行计划。

面：斯，你们考虑这个优化是因为它执行的太慢了是吧？

答：对啊，它执行慢我们就会考虑优化，第一步是SQL语句的优化，然后索引优化，分区优化，执行计划优化。

面：那你就是对你的主表或者驱动表进行优化，那你这些表的话，加索引的话，哦，加索引也可以。额，你对那个游标有使用的多吗？

答：主要是在存储过程时候使用，看需求吧....

面：（打断）你用过哪些啊？

答：显性游标和隐性游标

面：它好像还有分一个静态游标和动态游标

答：这个没有用过，我用过的只有显性游标和隐性游标。

1. 那你对大数据那个文件系统，现在文件已经上传到HDFS中了，现在我要在数据库使用的时候，一般是使用外部表还是自己建一个，比如叫ORC表这边的，你们是怎么用的。

答：没有理解到，然后带偏了

面：不是，你在使用的时候比如说去查询一些日志啊，查看它的date文件吗？

答：不会，答非所问，说我们是会常见的Linux命令，如......

面：（不屑一笑）那你有些过shell脚本吗？

答：我们是把shell脚本当容器，把以前的存储过程，放shell脚本里跑，实现数据同步。

面：（不屑一笑）那你这其实也没啥工作量，只是搬迁就是了，就是表名相应的做下变更。

答：附和下

面：那你更多的是做kettle这边调度这一块吧，那你在使用kettle时候有遇到一些问题吗？像你使用kettle时一次提交的行数是有多少行?

答：瞎编，就瞎了......然后面试官救场

面：你设置这么小是不合理的，其实一般的话是五千行以上