面试总结：

第一场面试：

1.问项目内容，问我在项目中干了什么工作:

答：银行贷款项目及其大数据改造。我主要从事ETL工作。

1. （忘记了问的啥）

第二场面试：（腾讯会议没录音，晚上想不起来问啥了）

1. 谈谈你对表空间的理解：

表空间是一个逻辑概念，它统一管理空间中的数据文件。数据库（房间外壳），表空间（房间里面的空间），数据库文件（空间中的箱子），表（箱子里的东西）。因为主要存放的是表，所以叫表空间。

表空间的分类：

永久性表空间、临时性表空间、撤销表空间。

1. 数据库备份：

逻辑备份和物理备份。物理备份分为：冷备份和热备份。

逻辑备份是指利用exp命令进行备份，简单易行，不影响正常数据库操作。经常使用。

冷备份：数据库关闭的状态下进行的备份。

热备份：数据库开启的状态下进行的备份。要先保证数据库运行处于归档模式，然后备份表空间文件，最后备份控制文件。

1. Oracle数据库怎么优化

--合理使用索引，建分区表，修改执行计划

可以进行SQL语句优化，也可以进行索引优化，分区表优化，执行计划优化。请问您想听哪一块呢？

第三场：问来源表数据发生变化，你是怎么做的？怎么做风险评估？

答：

1. 怎么做：（1.如果到达我上线的前三天，代码都是不会动的，有问题留到下个版本在更新。

2..如果是来源表数据类型发生变化，根据mapping文档，改ODS层对应列的数据类型。评估其对DW层和DM层的影响。

3.如果来源表字段发生变化，如这个关联字段被删除了，那影响挺大的，数据字典，mapping文档要改，ODS层里与原表发生变更对应的表要改，DW层相关的存储过程什么的要改，DM层也得相应的改。

4.如果某张来源表被删除了，那得流程就得重做）

风险评估：

通常情况下，遇到来源表数据发生变化时，我会考虑我的剩余工作量和剩余工作时间和变更的紧急性，首先考虑时间上允不允许我接受这个变更。如果时间上接受不了，我会找组长阐明状况，组长或者项目经理会进行风险评估，出解决方案。如果时间上能接受，我会考虑下这个变更的紧急性，然后和组长汇报状况，按任务优先级处理。至于发生变更带来的风险，需要组长或项目经理判断。

问：拉链表怎么建设?

1. 创建用户姓名变化拉链表
2. 创建最新用户姓名表
3. 找出发生变化的数据，比如ID相等，名字不等，结束时间等于默认最大时间。在拉链表把这部分数据的结束时间改成当前时间减一。
4. 把发生变化数据插入拉链表，开始时间为当天，结束时间为最大默认时间。
5. 找出拉链表没有的数据，插入拉链表。（不存在exists id相等的情况下就插入拉链表，开始时间为当天，结束时间是最大默认时间。）

问：拉链表，找某一个时间点，对应的拉链表中所有的信息。

Select \* from 拉链表 where 时间 between 开始时间 and 结束时间。

问：Linux 查hive中找脚本是否正在运行

答：**JPS**

问：SQL中常见的函数

答：聚合函数、分析函数、排名函数、位移函数、单行函数

问：rank和row\_number的区别

答：比如rank 是1224排，row\_number 是1234排。

问：数据整合层什么内容

答：根据DW层的事实表和维度表整合的大宽表，根据业务需求，写SQL语句，通过帆软或tebleau 出报表。

问：数仓架构

答：ODS层，主要是做数据的抽取，转换，加载，通过调度工具（CTM）按照要求，定期从源系统中讲数据抽取到ODS层。

DW层，从ODS 层同步数据到数据仓库，根据业务需求对事实表和维度表进行建模（星型模型或者雪花模型）。

DM层，主要为报表展示提供数据支撑，会从DW层同步数据到DM层，同时对各项指标提供数据或者通过函数计算得到报表需要的数据指标。