# 1有一张银行的记账表，我希望是早上8点到晚上6点这个时间段内进行增删改，非这个时间段内不准进行增删改等操作，我该用脚本怎么实现？

答：用存储过程＋条件判断实现，条件判断例如IF ELSE或CASE WHEN

面试官回答：其实用触发器就可以了，在一段特定时间内触发就行

# 2 索引有哪些？一般你对什么索引进行优化？

唯一索引

普通索引

组合索引

函数索引

位图索引

面试官回答：一般对组合索引进行优化

# 3 DW层相当于是做一张事实表是吗？

维度表（客户方面，机构，产品，账户之类的），事实表（交易信息表，存款信息表，信用卡开户信息等，贷款信息表等）

# 4 增量周期一般是多少

七天跑一次，实时性没有这么高

应该是每天晚上12点更新一次

# 5优化的话有哪几种方式，具体是怎么优化的？

**重点：SQL语句优化、索引优化、表分区优化、并行优化！**

/\*+ PARALLEL(表名,并行数) \*/ --指定开启多少个并行

# 6通过什么方式来实现这个拉链表的？怎么缓慢保存历史变化？

拉链表是保存历史数据的。有变化但变化不频繁的数据适合建拉链表。

比如我有100w用户，用户表记录了他们的状态，如果部分用户状态每天都有更新，我采用每天存全量用户数据的方式来记录历史和更新的状态，这样极大浪费存储空间，我用拉链表的话，每天把跟新用户的数据插入到用户表，这样既保留了历史数据，并且使存储最为节省。

# 7表分区有哪些类型？怎么进行分区？

**范围分区** （RANGE）：用指定的**分区键**决定的范围进行分区，最为常用，分区键常采用日期。

**列表分区**(LIST)：某列的值只有几个，可以采用列表分区。

**散列分区** (HASH)：通过指定分区数量或编号来均匀分布数据的一种分区类型，分区数量常采用2的N次方；当列的值没有合适的条件时，采用散列分区；

**组合分区**:范围分区和列表分区组合；范围分区和散列分区组合，分区中的分区被称为子分区；

分区表：-- PARTITION BY

CREATE TABLE 表名 (列名 数据类型 [,列名 数据类型]...)

PARTITION BY RANGE (HIREDATE)

# 8分区的优缺点有哪些？

优点：改善查询性能；增强数据可用性、可靠性；维护方便；

缺点：已经存在的表无法直接转化为分区,需要维护。

# 9存储过程实现的东西是不是也可以用自定义函数可以实现？

答： 可以

# 10行列转换是用什么方式来做的？

**列转行：**用分析函数lead/lag；用DECODE；用CASE WHEN;部分关联；pivot

注意：

pivot函数：行转列函数：
　　语法：pivot(任一聚合函数 for 需专列的值所在列名 in (需转为列名的值))；

unpivot函数：列转行函数：
　　语法：unpivot(新增值所在列的列名 for 新增列转为行后所在列的列名 in (需转为行的列名))；

# 11 SQL优化怎么优化？

**常用的SQL语句调优？**

1.避免索引失效；

2.多使用COMMIT提交数据；

3.in,not in 用exist,not exist 替换，少用JOIN；

 4.临时表；使用WITH..AS语句，对某段SQL建立临时表；拆分步骤；

5.DISTINCT 性能最差的去重，尽量用group by替代

6.尽量把like模糊查询语句替换成精确查询；

7.使用SQL内置函数加快查询速度，例如使用decode函数替换case when；

8.避免在select 后使用'\*' 符号；

9.连接和分组时先使用where条件过滤

10.避免笛卡儿积；

11.删除全表数据时TRUNCATE替换DELETE;

# 12怎么进行调度，用什么来调度

CTM调度

# 13跨库不用kettle怎么进行同步数据？

DBLINK

# 14存储过程包括什么？

包括过程的名称、过程使用的参数以及过程执行的操作

# 15存储过程你们是完成了一个什么样的功能呢？可以具体一点吗？举个例子

主要是用来同步数据，后面还说了结合项目具体业务的的主要业务

**根据具体的项目给我具体描述一下存储过程？**

答案：客户信息表，电话、地址拆分；性别补充；拆分另外一张表

# 16什么情况用游标，什么情况下不能用游标？

游标是指向查询结果集的一个指标，通过游标可以将查询结果集的记录逐一取出，并在PL\SQL程序块中进行处理。

隐式游标通过select into对变量进行赋值（只能是单行值）

显示游标必须先申明，然后打开游标，遍历数据，最后关闭游标

# 17索引失效的场景是哪些？

针对索引使用函数；

索引列算数运算：

索引列隐式转换；

！= 和NULL判断；

LIKE “%\_”百分号在前；

# 18全局索引和分区索引有什么区别？

**全表扫描和索引扫描的区别是什么？**

读取全表和部门的数据区别：提高效率 数据量少的时候，可以全表扫描

# 19数据仓库是怎么搭建的？有没有搭建数据仓库的经验，部署数据库

# 20 CHAR和VARCHAR2有什么区别

1．CHAR的长度是固定的，而VARCHAR2的长度是可以变化的， 比如，存储dao字符串“abc"，对于CHAR (20)，表示你存储的字符将占20个字节(包括17个空字符)，而同样的VARCHAR2 (20)则只占用3个字节的长度，20只是最大值，当你存储的字符小于20时，按实际长度存储。

2．CHAR的效率比VARCHAR2的效率稍高。

1)varchar2所有字符都占两字节处理(一般情况下)，varchar只对汉字和全角等字符占两字节，数字，英文字符等都是一个字节；

2)VARCHAR2把空串等同于null处理，而varchar仍按照空串处理；

hql等价改写为SQL：

# 21数仓每一层的作用是什么？

ODS层，主要是做数据的集成，转换和同步，通过调度工具按照每天／每月，定期从源系统中将数据抽取到ODS层。

DW层，主要是从ODS层同步数据到数据仓库，根据业务的需要对事实表和维度表进行建模（星型或雪花模型）

DM层，土要是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。

# 22开发过程中肯定会遇到比较慢的SQL，你SQL优化的步骤或者思路是什么？

数据量比较大的表查询很慢，我会进行分表分区，建立索引，SQL语句进行优化

**程序跑得慢怎么知道是哪一块跑得慢？**

分段注释代码，进行验证，直到找出执行那段SQL耗时最长。

每个存储过程在进行数据同步时都会记录日志，日志中会记录数据同步的开始和结束时间，找到最慢的程序，然后一点点发现。

# 23大概有多少张报表？

一个月十几张，包含新增和修改的，因为是一个月一个版本

# 24你们的数据是怎么对接的？业务系统中的查询接口是怎么接过来的？

# 25左外连接时，100条数据变为200条，数据增多的原因是什么？

# 26成绩表中，有姓名、科目、成绩三个字段，查询出每一个人的最高分是什么科目？

SELECT CC.\*

FROM (SELECT C.\*,ROW\_NUMBER()OVER(ORDER BY SCORE DESC) RM

FROM 表名

GROUP BY 科目 C) CC

WHERE CC.RM = 1;

# 27单张表或多张表怎么对数据进行清洗？

# 28怎么进行表与表之间的优化？

**有两张很大的表关联，怎么做才能关联性能好？**

1.首先要建立适当的索引。SQL在索引字段不要加函数，保证索引起效。如果是复合索引注意在SQL的顺序。如果已经存在索引，建议你先重建索引，因为大数据表的索引维护到了一个阶段就是乱的，一般建议重建。建立好的一般可以获得几十倍的速度提升。

2.最大数据量的表放在最前，最小的表放在最后面。SQL是从最后面开始反向解析的。

3.其次是要把最有效缩小范围的条件放到SQL末尾去。尤其是主键或者索引字段的条件。

4.为两个表设计合理的表分区，然后分别对应关联两个表的分区数据，再用union all把各个连接结果叠加起来；

**大表和小表连接如何进行优化？**

将小表作为驱动表，大表作为匹配表，关联字段建索引。