**赛意现场面试：**

1. 存储过程写过多少行？创建过多少个包？

看需求，简单的三四百行，复杂的上千行。

看项目，一个项目大概一个包。

包可以放很多东西，比如存储过程，自定义函数，包的作用就是方便调用。

1. 存储过程里用到的异常有哪些：除了主键冲突，others，还有其他吗？（重点问我还有其他吗）

存储过程里遇到的异常：主键冲突；内存溢出；除数为零；程序试图打开一个已经打开的游标；系统等待某一个资源 ，时间超时

1. 给一个数据同步的题目，怎么去重，怎么更新数据

在Kettle里数据同步时，有个去除重复数据的组件可以去重，可以用全量或者增量更新，如果数据量小可以用全量更新，如果数据量大用增量

1. 两个表关联，怎么性能好

如果两个表都是数据量很大的表

1. 游标怎么用

Begin -> 打开游标open -> 取值 -> 关闭游标close -> end

1. 动态视图是什么，有用过吗？

动态视图属于数据字典的一种，很像普通的数据库表，但不允许修改，大多数只能特权用户和dba用户查询，名字一般用v$开头。

锁表之后可以查动态视图看锁表的进程。

7.业务方面的，有哪些指标，有哪些表，需求给到你了，怎么去做？

8.怎么知道自己的多个索引有没使用？

9.索引失效有哪些？怎么看执行计划，执行计划里有什么？

索引失效的原因：索引列进行运算；针对索引使用函数；对索引列使用隐式转换；用！=和null判断；like ‘%\_’百分号在前；

看执行计划： 执行一条SQL语句之后，按F5可以查询该SQL语句的执行计划。

可以查看基数（当前操作返回结果集的行数），

字节（执行该步骤后返回的字节数），

CPU耗费（执行该步骤的执行成本），

时间（当前操作所需的时间）。

**妍丽（自研）**

1、什么是事实表，什么是维度表

2、帆软和tableaue的区别

两者各有千秋，帆软是OLTP联机事务处理产品，定位于企业数据采集和处理软件，满足企业个性化报表需求，同时具备强悍的列印和填报等功能，但是数据分析能力较弱。Tableau是OLAP联机分析处理产品，具有很强的数据分析能力，能够灵活的在前端进行数据分析操作，无需编写编程码，劣势是无法满足企业复杂报表，列印等需求。

3、海量数据用增量更新很慢吧，怎么不在etl中直接实现

4、oracle怎么重启

**Windows：**

（1） 启动两个Oracle服务，一个是监听程序，一个是Service

（2） cmd-->sqlplus /nolog

（3） cmd-->set ORALCE\_SID=HRKF

（4） sql-->connect / as sysdba

（5） sql-->startup

5、存储过程怎么写

Create or replace procedure sp-过程名 (参数1 数据类型，参数2 数据类型。。。)

AS

V-变量名 数据类型；

Begin

End；

6、常用的函数：数字函数，字符函数，转换函数，日期函数，rownumber , rowid等

7、会建模吗？会，在dw数仓层，取出我们需要的数据按照事实表和维度表进行建模，雪花模型或星型模型。

**软通（电话）：**

1、性能优化；

2、表分区及其用法；

表分区分为列表分区，哈希分区，组合分区，范围分区

列表分区：某列的值只有几个，基于这个特点我们采用列表分区

哈希分区：通过指定分区编号来均匀分布数据的一种分区类型

组合分区：基于两种分区的组合，分区之中的分区被称为子分区

范围分区：最常用的一种分区，经常使用日期分区，比如年月日。。

3、分析函数：采用partition by，，在数据中进行分组然后基于组的某种统计值，并且每一组的每一行都可以返回一个统计值。种类：row\_number(),rank(),dense\_rank(),count()….

4、索引的用法和优化；

索引分为唯一索引，组合索引，位图索引，基于函数的索引。

唯一索引：当某列任意两行数据不相同，建立主键和唯一约束时索引自动建立

组合索引：当多列经常一起出现在where条件时，创建索引

位图索引：当一列有大量重复数据时建立

基于函数的索引：当where条件语句中包含函数或者表达式时建立

5、ds的用法；

6、数仓分层和各自的功能；

7、项目指标（问的很细）；

**天阳（电话面试）**

1、表怎么分区

 1.范围分区：RANGE

 2.列表分区：LIST

 3.散列（哈希）分区：HASH

 4.组合分区

2、写过哪些脚本：

3、银行业务问题

**赛意**

1自我介绍，项目介绍。

2性能优化

3自定义函数写过哪些

4出了个题目需要把表中工号取出来，工号不规则 提示了正则法则

易宝面试

自我介绍

1、性能优化

2、union  和 union all

3、行转列

4、进程慢  卡死的情况   怎么处理

5、表之间的关联的方式

6、表分区的建立

7、表学生和学科和成绩 ，找出所有的成绩大于80 的成绩

8、Linus熟悉吗

赛意技术面试：

1. 项目中 用到哪些技术
2. Sql性能优化
3. 怎么看执行计划，有哪些内容

在sql窗口执行完一条select语句后按F5查询该条查询语句的执行计划。**基数ROWS：**返回结果集行数；**字节Bytes**：返回的字节数；**耗费COST、CUP耗费：**SQL执行成本；**时间time：**当前操作所需的时间。

1. Merge into 用法
2. 常用函数有哪些
3. 建模类型有哪些
4. 怎么查看数字字典
5. Hints用法，怎么用
6. 怎么删除表分区数据？（只删除数据）

佰钧成电话面试

1. 问做哪些工作内容
2. 事实表和维表数据有什么不一样
3. Datastage 跟kettle有什么不一样
4. 星型模型和雪花模型有什么不一样
5. 写过存储过程吗？是在plsql工具上写还是在kettle上写
6. 写存储过程想要达到什么目的，想要实现什么功能？
7. Linus用过吗？

华为，中软12.8 复试

1.离职原因

2.介绍上个项目做的什么事

3.需求多不多？

4.多长时间一个版本

5.一个版本大概怎么运作

6.有哪些角色？有什么分工？（项目人员分工）

7.pg库与oracle有什么不一样？

8.decode 在pg 库中有什么不一样？

9.数据库的性能调优，除了索引和sql的调优，还要什么方法调优？（我答，还要分区表和hints优化器调优，面试官说这也是属于sql调优的一种）数据库调参

性能调优： 首先要对sql语句进行优化，多用group by ,

10.数据库调参有用过吗？这也是数据库调优的一种。

11.举一个工作中实际做过的性能调优的例子

我在我的表中选取一个经常使用或者说重复率高的字段，建一个索引以达到优化性能的目的。当我需要做增量时，我会先把该索引设为失效状态，等数据更新完以后再把索引设为有效状态。

中软国际 华为 12.8 初试

1.说一下以前负责的项目流程

2.常用的函数、分析函数

常用的函数：字符函数，数字函数，日期函数，转换函数，nvl , decode, case when...等等

分析函数：是oracle数据库为了解决复杂报表统计需求的功能强大的函数，它可以在数据中分组然后计算基于组的某种统计值，并且每一组的每一行都可以返回一个统计值。种类：

聚合分析函数Sum(),min(),max(),avg(),count(),

排序分析函数rank(),dense\_rank(),row\_number()

3.星型模型和雪花模型的原理

星型模型是所有维表都能直接与事实表关联，存在冗余数据，一般来说性能会更好；

雪花模型：一个或多个维表没有直接与事实表关联，需通过其他维表才能关联到事实表，雪花模型是星型模型的扩展，去除了冗余数据，因与维表要关联多次，所以效率没有星型模型好；

事实星座模型：多个事实表共享维表，可以看作是星型模型的汇集。

软通 深交所 12.8 初试

1.自我介绍

2.项目流程

3.不同来源的数据，有相同字段，你怎么确定哪个来源的数据是正确的？

我回答：这种情况就当作无效的数据清理掉。面试官表示惊讶。

4.同样的场景，不同来源的数据，有相同字段，a来源的有aa字段，没有bb字段，b来源的有bb字段，没有aa字段，这种情况怎么办？

我答：没有数据的字段设置空值。面试官点头

5.怎么理解拉链表？什么数据才需要创建拉链表？

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。

12.9 天阳 厦门

1.实例

实例是访问Oracle数据库所需的一部分计算机内存和辅助处理后台进程，是由进程和这些进程所使用的内存(SGA)所构成一个集合。
其实就是用来访问和使用数据库的一块进程，它只存在于内存中2.truncate 与 de le te的区别

TRUNCATE 是DDL命令，可以把表中的所有数据一次性删除，删除的数据不能恢复；

Delete 是DML命令，删除后的数据可通过日志文件恢复。

如果一个表中的数据记录很多，truncate 相对于delete速度块。

3.高水位线怎么理解

高水位线 是一个标记，是一个容纳数据的上限，原则上只会增大不会缩小。delete删除数据之后，高水位线不会重置为0，还是为原值。即使高水位线下都是空的，当插入数据以后，高水位线就会上涨。

4.怎么回收这个高水位线

1）、用truncate 命令回收由delete命令产生的空闲空间

 Truncate table big\_emp1 reuse storage

2)、复制要保留的数据到临时表T，drop原表，然后rename 临时表T为原表

3）、执行表重建命令 alter table 表名 move

5.存储过程用什么捕捉异常

用exception捕捉异常

6.捕捉异常以后怎么处理异常

在exception中把过程数据记录到错误日志表中，然后对主过程的事务进行回滚，并针对异常我们会记录在异常日志表中，方便后续的问题分析。

7.存储过程有用过游标吗？有

8.游标的概念

游标是 指向查询结果集的一个指针，通过游标可以将查询结果集中的记录逐一取出，并在pl/sql程序块中进行处理。

9.什么时候关闭游标

当最后一行结果集被处理完后，close就会关闭游标。

10.分区表

当表中的数据量不断增大时，查询表中的数据速度就会变慢，性能下降，这个时候就应该考虑对表进行分区，表进行分区后，逻辑上表还是一张完整的表，只是将表中的数据在物理上存放在一个或者多个表空间，这样查询数据时，不至于扫描整张表。

11.分区表建的索引和单表建的索引有什么不一样

分区表： 建的索引在这个区域才有效，在其他分区无效

单表： 建的索引对整张表有效。

12.知道什么是shell脚本吗？

Shell脚本就是用各类命令预先放在一个文件中，方便一次性执行的一个程序文件，主要是方便管理员进行设置和管理用的。它跟windows的bat批处理文件类似，但是比windows的批处理更强大，比用其他编程程序编辑的程序效率更高，它使用了linux下的命令。

软通12.11 技术面试

1. 项目流程
2. 物化视图？物化视图占不占用存储空间？

物化视图是当时建立视图时固定下来的数据，如果原表数据有更新，物化视图的数据是不变。

占用存储空间。

1. 内关联和外关联

内关联也叫内连接，是从结果表中删除与其他被连接表中没有匹配行的所有行，所以内连接可能会丢失信息；

外关联，分为左外关联和右外关联，全外关联

左外关联，以左边的表为主表，右边的表为从表，主表的记录全部显示，从表表的关联字跟主表相同的就全部显示，不相同的就自动补null。

右外关联，以右边的表为主表，左边的表为从表、、、、

全外关联，完整外部连接返回左表和右表的所有行。

1. 模型设计
2. 星型模型和雪花模型
3. 表中有重复的记录，怎么找出来？用group by，rowid伪列
4. 分析函数有哪些：

row\_number， rank, dense\_rank, lead/lad, 函数（）over partition by

1. 怎么排名？

 不考虑并列，用row\_number,考虑并列用 rank，dense\_rank

1. 性能优化

首先我会先对sql语句进行优化，比如select语句不用\*星号，多用commit提交数据，多使用表的别名，用where子句代替having子句等。其次，还可以在重复率高的字段建立索引，如果表中数据量很大的话，可以考虑建分区表。查看执行计划，看看耗能多少，跑了多长时间。如果还不行就考虑hints优化器进行优化。

1. 分区表有哪些？

分为 ：范围分区，散列分区（哈希分区），组合分区，列表分区

1. 缓慢变化维

处理缓慢变化维的三种方式：1、直接覆盖原值2、添加维度行3、添加属性列

1. 数仓架构：分为ods层，dw层，dm层

moon:

增量方式：

1、时间戳

2、没有时间戳就看有没有记录数据版本号标记，如果有就按数据版本号取增量

3、触发器，来源表上创建触发器，如果发生数据变化就插入到目标表

4、物化视图，通过定期刷新物化视图的方式记录新增

5、数据库事务日志，通过解析数据表操作日志获取变化的数据

6、其他数据实时同步组件，来源发生变化就插入到目标，类似从库，订阅库

宽表切片：

指从包含了很多维度的宽表里按一个或几个用户想看的维度把数据统计出来，也就是切出来，把这个提供给用户查看，避免用户访问整张宽表，达到优化体验的目的。