一、SQL优化：

1. 索引（优缺点）

优点：

查询快，提高查询性能

缺点：

 占用存储空间，并不是越多越好

 占用维护的开销。 每一次往表中 修改(增删改)数据的时候，索引都需要自动更新，影响到更新的性能。

2.导致索引失效的原因；

1. 给索引用了函数 2、作了判断 3、使用了like 模糊查询 4、隐式转换等 5、考虑是否重建索引？ 因为操作频繁，导致表的索引失效。

3.简单说说视图：

1. 可以限制用户只能通过视图检索数据。这样就可以对最终用户屏蔽建表时底层的基表，具有安全性。（优点）
2. 可以将复杂的查询保存为视图，屏蔽复杂性。（优点）
3. WITH READ ONLY说明视图是只读视图，不能通过该视图进行增删改操作。现实开发中，基本上不通过视图对表中的数据进行增删改操作

Sql语句优化：

（这个一般都是当执行慢的时候就要考虑优化）

1、检查关联条件：看关联有没有找对，值是不是唯一的

2、去重的时候尽量使用group by 这样跑的会快一些

······

用hint优化

查看执行计划，耗费超过1000的就优化。

有用到索引的先把索引关掉（no\_index（表别名）)、嵌套式连接改hash ：use\_hash(表别名）

小表驱动大表leading(表别名）

在大数据中使用过哪些语言？

简单说一下内部表和外部表

Delete 和truncate操作对表有什么影响？

truncate table命令将快速删除数据表中的所有记录，但保留数据表结构。这种快速删除与delete from 数据表的删除全部数据表记录不一样，delete命令删除的数据将存储在系统回滚段中，需要的时候，数据可以回滚恢复，而truncate命令删除的数据是不可以恢复的

注意:这里说的delete是指不带where子句的delete语句
相同点
truncate和不带where子句的delete, 以及drop都会删除表内的数据

1. 数仓怎么搭建？



## **业务数据生成**

### **建表语句**

1. 创建数据库
2. 设置数据库编码
3. 导入建表脚本，创成表
4. 导入数据插入脚本，函数脚本，存储过程脚本

### **业务数据生成**

通过函数脚本生成业务数据

## **业务数据导入数仓**

ODS层

完全仿照业务数据库中的表字段，一模一样的创建ODS层对应表。

DWD层

积压ODS层对数据进行判空过滤。对商品分类表进行维度退化(降维)。

DWS层之用户行为宽表

为什么要建宽表？

需求目标，把每个用户单日的行为聚合起来组成一张多列宽表，以便之后关联用户维度信息后进行，不同角度的统计分析。

创建用户行为宽表

导入用户行为数据

三、之前有用到哪些关联，怎么操作的，存储过程结构如何

1.内关联、左右关联、全外关联（）平时左右使用左关联比较多。

Select \*

From 表

Left join 从表

On 关联字段

2.创建存储过程的语法结构：

create or replace procedure 存储过程名[(参数1 [参数类型] 数据类型,....)]

is / as

 --声明/定义部分

begin

 --逻辑操作部分

end;

--开发规范：存储过程名以 sp\_ 开头

ETL做过吗，知道增量全量区别吗？

四、博彦科技——浦发

请分享一次您在团队共同合作完成的任务中遇到棘手问题的经历。具体讲讲当时遇到了什么样的问题是?遇到问题之后您做了什么?最后这项任务的结果如何?

在你过往的项目经验中，有没有为了更好地完成任务而付出了很多额外时间或努力的经历?请具体讲一讲当时的任务内容和目标是什么?您做了哪些额外的努力?任务结果如何?

在您的过去工作中有没有主动承担过不是自己分内之事的任务?请描述一下当时的场景以及你做了什么?结果如何?当时又是什么原因驱使你去做这件事情。

有的，项目组人员调动，人手不够，我们都需要额外负责其他部分的工作，为了项目进度

我们时不时需要对曾经写过的代码或现有的工作流程进行优化，以提升效率，或避免未来出现错误可能性。请举例说明最近你为提升工作质量/效率所采取的优化方案。原来的代码或工作流程存在哪些纰漏或不足?你做了哪些优化?这些优化带来了哪些成果?

1. 之前工作中遇到什么困难，如何解决

问题：使用KETTLE工具同步数据出现中文乱码问题

原因：因为2个数据库的编码格式不一致 导致出现中文乱码 。

解决办法 ：

1 找到 相应的DB（连接目标库）链接 修改 编码格式为 *utf8*  或则 GBK ;

 在选项中 设置

参数名 ：characterEncoding

值：utf8

或者 gbk

六、使用过informarica吗？Informatica了解吗

————跟kettle是ETL的抽数工具

七、索引失效的原因？

 隐式转换导致索引失效

 对索引列进行运算导致索引失效

 使用Oracle内部函数导致索引失效

使用 <> 、not in 、not exist、!=

八、ORACLE：简单解释冷备份和热备份的不同点以及各自的优点？

热备份针对归档模式的数据库，在数据库仍旧处于[工作状态](http://www.so.com/s?q=%E5%B7%A5%E4%BD%9C%E7%8A%B6%E6%80%81&ie=utf-8&src=internal_wenda_recommend_textn" \t "https://wenda.so.com/q/_blank)时进行备份。

而[冷备份](http://www.so.com/s?q=%E5%86%B7%E5%A4%87%E4%BB%BD&ie=utf-8&src=internal_wenda_recommend_textn" \t "https://wenda.so.com/q/_blank)指在数据库关闭后，进行备份，适用于所有模式的数据库。

热备份的优点在于当备份时，数据库仍旧可以被使用并且可以将数据库恢复到任意一个[时间点](http://www.so.com/s?q=%E6%97%B6%E9%97%B4%E7%82%B9&ie=utf-8&src=internal_wenda_recommend_textn" \t "https://wenda.so.com/q/_blank)。（用alter database bachup controlfile命令来备份控制文件）。

冷备份的优点在于它的备份和恢复操作相当简单，并且由于冷备份的数据库可以工作在非归档模式下,数据库性能会比归档模式稍好。

冷备份也有如下不足：

1、 单独使用时，只能提供到“某一时间点上”的恢复。

2、 再实施备份的全过程中，数据库必须要作备份而不能作其他工作。也就是说，在冷备份过程中，数据库必须是关闭状态。

3、 若磁盘空间有限，只能拷贝到磁带等其他外部存储设备上，速度会很慢。

4、 不能按表或按用户恢复。

热备份的不足是：
1． 不能出错，否则后果严重
2． 若热备份不成功，所得结果不可用于时间点的恢复
3． 因难于维护，所以要特别仔细小心，不允许“以失败告终”。

九、使用过powerbi吗？之前报表用什么工具做的？

 帆软、tableau

帆软用了多长时间？

几个星期

1. 基于岗位需要涉及到相关大数据基础，有使用过HIVE吗？
2. kettle数据同步用的是增量还是全量的（增量）

2.kettle抽数怎么做？

1. 数仓是怎么分层的？

4.之前有用过什么数据工具？

5.索引失效的原因？

6.工作原因需要时常出差，能接受吗？

7.一张数据量很大的表（上千万）怎么存储，才能顺畅的进行同步？

1. BI跟建模有经验吗？

 ETL架构师的回答是：就是根据分析内容，建立起一张一张的事实表与维度表，并建立相互关系。使其满足数据分析的要求。说白了，就是表结构。

业务模型定义非常清晰，数据结构定好后，在实施过程中，特别是数据核对过程中，不断的需要修改表结构调整，原有的指标定义出现问题等。（银行和电信项目除外，因为都是文本文件抽取，其不会有这种感受，只是数据量处理而已。）

1. 什么是执行计划？
2. 脚本执行报错了，怎么做？

查日志

有个用户登陆表，包含用户id和用户登陆时间，表名users

求：用户连续登陆天数，和用户最大连续登陆天数



先把每个用户按照登陆时间排序rn，在用登陆时间减rn得到一个值，如果值一样，说明是一段的连续登陆，然后把不连续的天数去掉（上一步得到的值只有一个，说明是不连续的），然后得出每段连续的登陆

1. 说一下hint优化
2. 介绍一下近期做的项目
3. 你在项目中主要负责那一块？
4. 能否举例你在项目中遇到的困难？怎么解决的？
5. 维度建模有什么优缺点
6. 未来职业规划