1，项目中怎么实现数据同步，你涉及的存储过程是怎么写的，还有用过什么函数（写过的自定义函数有哪些），

根据项目你所负责的指标来介绍（比如指标净息差=利息净收入/生息资产，从哪取数，用到哪些表，那些字段，怎么写存储过程及自定义函数，写完后怎么去验证等）。

 项目中怎么实现数据同步 ，

 增量 和 全量。 在在此基础上可以通过SP的参数进行切片的增量或者全量同步 。

 写过的自定义函数有哪些 ？

 自定义函数写的不多 ，主要还是开发存储过程 。

2，工作这么多年或者你在这个项目中，遇到过的难题有什么，发现怎么错误，怎么解决？

 最困难的事情主要是 业务方面吧 ，每次接触新的项目 对新需求不熟悉的时候 ，需求理解会有点慢，导致开发效率开始的时候较慢，BUG也比较多。

 但是随着我对项目的业务的熟悉程度不断提升，这些问题都会消失 。

3，如果在执行/编写存储过程中，出现问题，怎么定位错误？

 1 先保证 存储过程 编译通过 。

 2 保证调用正常。

 3 如果编译失败 或者 调用失败 ，要看错误提示 。比如 oracle的异常编码 ora-xxx

 解决的办法，直接在 SP 的编辑界面执行以下代码 一般报错的部分会用 黄色高亮显示 。

 找问题 一般是从报错的位置 向上找问题 。

4，如果在同步过程中出现数据错误，或 数据中出现 null值，怎么去处理 ？

 出错 查日志 分析问题的原因直接修改 。

 对于 null值 ，我们项目之前的开发规范要求 必须给 null 值一个 默认值 比如通过 nvl() 函数赋值 。

 建表的时候 加约束 ，计算的时候 维度字段不允许为空 。

5，dm层存储的表，是不是每个指标一个表，还是存储一定量的表，然后所有指标都在这些表表中关联提取进行展示？

 看需求吧 ，有些报表 对应一张表 ， 有些是 一张表 对应不同的报表。

6，dm层取数是不是一定从dw层取？ dw层跟dm层的区别？

 不一定 ，如果要做历史数据和最新数据的对比 ，那么 DM层的数据 可能来自 ODS 和 DW.

 DW 层主要做数据汇总 ，包括历史数据，比如 拉链表就在 DW 层。 建模的主要工作也是在 DW 层。

 DM 层主要是偏业务的指标计算 做可视化的。

7，上线后出现问题怎么解决？

 现在测试环境解决并验证通过后 ，再跟新到生产环境进行验证，通过后关闭问题单。

 解决问题一定要在 上线的时间窗内解决 。

8，代码质量如何？出现错误的次数是多少，频率？

 这个肯定没问题，我特意查过测试提交的BUG，也看过测试报告，我发现大部分问题都是 前端的问题 。

 一般情况下 ，一个存储过程的BUG 在 5~10个内。

9，dm层怎么建数据模型，以什么方式储存，怎么使用？

 通过 PowerDesigner 这个工具建模，具体的不是我负责的。

10，mysql 跟oracle 的 区别？

 性能 ，成本 ，数据类型 和 函数不同的。

11， 你用过哪些脚本

 主要是 SQL 脚本 。ETL 脚本 。

12 ，有没有用到过代理键？

 了解，代理主要主要用在 拉链表保留历史数据时使用 。

13，迁移表怎么做的，如何确认迁移表的数据完整性

 迁移就是数据迁移 。

 指 将 A 数仓的数据 迁移到 B 数仓 ，迁移的过程 可以使用 导入导出实现 也可以用 ETL工具实现 。

 步骤：

 1 迁移表结构 和 表的数据 。

 2 指标的计算和展示的数据源 也要切换 到B ，同时也不能直接把 A 断了 。

 3 原来的脚本将数据存储到 A 的，全部要修改 成 B 。

 如果A 和 B 数据库的类型 不同的，那么里面所有的脚本 要重新修改 。

1.有A\B\C三列，按C排序，取20-30行，怎么在Oracle、mysql里面写代码

 oracle中使用 rownm 分页查询实现 。

 mysql中可以使用 limit 。

2.怎么捕获异常

 exception 捕获异常 。

3.捕获到的异常种类，捕获到异常之后怎么处理？

 非运行是异常和运行时异常 。

 预定义异常 和 自定义异常 https://www.cnblogs.com/yunduo/p/4251930.html

 捕获异常后记录到日志表中 。

4.数据调优经验，怎么定位性能瓶颈问题，怎么解决

 性能瓶 分为三类 ：

 1 硬件的性能瓶颈

 这个需要查服务器的资源占用率 。

 CPU ,内存 ，硬盘的 I/O

 2 代码的性能瓶颈

 执行计划 ，资源消耗 ，耗时等 。

 3 网络带宽的性能瓶颈

 要做带宽的监控和测试 。

5.说说 索引 和 HINTS，怎么定位哪里有问题，问题是什么，怎么解决

 索引 是为了优化性能而创建的 ，参考索引的种类和 索引的注意事项 。

 HINTS 是修改执行计划的命令 ，通过执行计划 可以看出 扫表的方式 和 表的 关联机制 以及 消耗 ， 耗时 等等 。

6.没有异常，以前正常，现在跑得特别慢，怎么找问题？

 检查是否锁表 。

 详见见 《oracle中的锁》

7.如果建索引，你一般找那些列建索引？

主键已经自带唯一索引

8.分区选那些进行分区？

9.Oracle里面你知道常用的DDL语句吗？

 数据定义语言（DDL），包括CREATE（创建）命令、ALTER（修改）命令、DROP（删除）命令等。

10.一个表，已经建好了，怎么增加一个列？改表结构。说一下建表语句

 ALTER（修改）命令

11.TRUNCATE直接放在存储过程吗？要怎么使用？

 不行 ，必须要使用 动态SQL 执行 。

12.ETL工具，KETTLE常用的组件？DATASTAGE常用组件？DATASTAGE常用客户端？

 背一下面试宝典中的答案 。

13.星型和雪花模型的优劣？底层逻辑有什么不同？

 两种模型的区别。

 https://www.cnblogs.com/jiangbei/p/8485122.html

14.LINUX的使用，有一个.SH文件我要调用它，怎么调？怎么调SHELL脚本？

15.创建文件夹，查看文件，编辑文件vi，平时看文件怎么看？

16.如果文件有几百万行，我想检索里面的某些关键字，怎么检索？

17.HINT强制走索引，强制检表扫描？

 --常用HINTS

 1、/\*+ PARALLEL(表名1,并行数)[(表名2,并行数)……] \*/ --指定开启多少个并行|并发（一般为2、4、8……）

 2、/\*+ INDEX(表名,索引名) \*/ --指定索引

 3、/\*+ FULL(表名) \*/ --指定全表扫描

 4、/\*+ USE\_NL(表名1，表名2) \*/ --指定用NESTED LOOP连接

 5、/\*+ USE\_HASH(表名1，表名2) \*/ --指定用HASH连接

 6、/\*+ USE\_MERGE(表名1，表名2) \*/ --指定用SORT MERGE JOIN

 7、/\*+ LEADING(表名1，表名2) \*/ --指定表1作为驱动表

 8、/\*+ APPEND \*/ --数据直接插入到高水位上面(与insert连用)直接往后面插，无视前面的空位置 。

18.版本控制用的是什么？GIT有用过吗？

 SVN 跟 GIT 是同一类工具 。

19.怎么识别一个客户是一个人，并不是两个人？给过来的ID是已经识别好的吗？

 在下游的目标表中创建主键（ 主键是约束 ） 。

20.ETL仓库做的那一层？

 ODS 贴源层 。

 主要做汇总吗？

 ODS 不做汇总。

 有一个表有姓名、性别，我要统计这张表男女各有多少人，性别存的是女是i，男是1。最后想要的是两列，男多少人，女多少人。

 通过 CASE WHEN 对男性 和 女性 分别进行判断 然后再通过 聚合函数统计人数。

 SELECT

 COUNT( CASE WHEN 性别 = '1' THEN 1 END ) AS "男多少人" ,

 COUNT( CASE WHEN 性别 = 'i' THEN 1 END ) AS "女多少人"

 FROM 表 ;

1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？

 不行 。

 自定义函数通常是用来计算的 ，不做数据更新/同步 。 但是从语法的角度来说 可以写。

2什么情况用游标，什么情况下不能用游标？

 数据增量同步的时候使用游标 。

 游标 一般结合循环一起使用 。

 全量同步时 不能使用游标。

3 全局索引和分区索引有什么区别？

 全局索引是再表中创建的索引 。

 分区索引 也叫做 局部索引 实在 分区表的 某个分区内创建的。

4 数据仓库是怎么搭建的？面试官问的是不是怎么搭建成数仓3层架构？

5 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？

 我们所有的需求 都是直接跟 产品经理对接 ，不直接跟 业务对接 。

 如果有需要的话 我也愿意尝试一下 。

6 单张表或多张表怎么对数据进行清洗？

 单表 数据清洗是为了确保数据的有效性和完整性 ，这个在 ODS 做数据清洗 ，可以用 SQL 或者 ETL 工具实现 。

 多张表 一般是数仓中 清洗掉 不需要的 列 。关联后 只查询我们需要的列 不需要的列 不查询/清洗掉 。

Oracle表空间、段、区和块简述

http://blog.itpub.net/17203031/viewspace-682003/