2019/4-2020/11**广东华兴银行管理驾驶舱**

所属公司：

深圳市火星时代数据服务有限公司

项目描述：

建设华兴银行管理驾驶舱系统，主要是为银行管理层分析决策提供数据指标方面的有力支持，该系统主要对日常业务进行统计、分析和监控，从多种维度通过仪表盘、柱状图、 折线图、报表、地图等形式将经营情况直观地展示，为领导层提供决策支持。

责任描述：

1.参与需求分析，分析源数据以及取数逻辑；
2.参与模型设计、数据字典编写等方案的设计；
3.根据项目需求进行DM层开发，通过PL/SQL开发存储过程进行数据同步，完成数据字典的编写及维护，编写自定义函数进行部分报表指标计算，检查SQL性能；
4.使用CTM配置依赖和定时任务；
5.通过SQL和FineReport报表开发工具根据主题进行报表开发；
6.独立完成UT测试，验证数据的逻辑和指标计算正确性，并协助完成SIT测试，解决发现的BUG更新环境；
7.输出上线脚本清单，上线定位修复BUG , 形成闭环。

**题目在第二面**

1. 你们项目组维度表多少张，数据量大概多少

 我负责部分的维度表有客户，产品，账户，币种，渠道和机构等，客户大概有200多万

 2、你在工作中遇到过哪些报错，什么原因导致

  向目标表里放数据的时候类型不匹配或者值过多、过少，原因是没有对应目标表去插入数据，

3、查两千万的数据 得多久

 这个要看取数规则，限制条件少，关联表少的时候很快，几秒钟就查询出来了，条件复杂，关联条件多的话，最多跑半个小时

 4、你们表最大的数据有多少条？

  借记卡交易明细数据很大，差不多2个亿

1. 每天增量同步的表有哪些表，增量数据有哪些

 一般事实表都是增量的，借记卡开户信息表、贷记卡合同信息表、交易信息表、存款信息表等，每天同步新开户、新合同、新的交易流水、新存款等

 6、几十亿的数据你们怎么同步？每天更新的数据量

  增量同步，对于交易明细每天有几百万的增量，这个比较大

1. 整个项目每个周期大概要出多少张报表。

 十几张，包括新增的和修改的需求

1. 取数部分是按什么取的，怎么保证抓的数据尽量少又不漏

 时间戳来取的，我们是取最近7天的数据作为增量

1. 大表的数据有多大，性能这块儿怎么样？报表一个月大概会做多少张？

 大表的话就是历史表了，上亿了，性能可以，因为我们会做多层汇总，不同粒度会放到不同汇总表里；一个月十几张，包含新增和修改的，因为是一个月一个迭代。

1. 项目数据的来源是什么？需求是怎么来的，拿到需求以后怎么做的？

 数据来源于业务系统，包括个人贷款系统、公司贷款系统、票据系统、总账系统、基金理财系统等，需求是客户先提到需求分析师那里，然后需求分析师会对需求进行拆解和澄清，确认取数逻辑，然后我们会开会讨论如何实现，包括可行性分析，设计好模型表，然后我们进行开发的。

1. 存储过程写的多吗，一个版本写多少个，一个有多少行，写过最复杂的存储过程是什么呢，存储过程中怎么记录日志，找出异常？

 嗯，我主要就是写存储过程，一个版本大概要写2、3个，每个7、8百行，还有一些对历史存储过程修改的，最复杂的是统计卡交易明细，因为粒度比较细，涉及表很多，还要对交易类型和机构进行过滤等等，写了1000行左右。存储过程每个步骤都会记录日志表，也会记录异常（**这部分到讲义里复习一下！**）

1. 说说项目中做过的指标有哪些？用到的表有哪些？

 客户余额、日均余额统计表：用到了客户交易明细和客户信息、币种、机构、贷款合同等

 如存款账户交易月汇总指标，即根据存款账户不同的交易类型（交易码）来统计交易笔数和交易金额：客户交易明细和客户信息、币种、机构。

1. 设计模型的时候有没有发生变更的情况，为什么发生变更？划分主题的方法论是什么，怎么保证划分主题是正确的，怎么保证来了几个需求之后不会发生很大的改变？之前的DM层有哪些模型，具体的业务是什么？

 主要分析数据仓库的数据如何满足集市的需要，即确定需要数据仓库的那些表数据，如果数据仓库没有，则需要数据仓库进行数据补充。先分析数据应用系统的需求，再提炼数据集市的数据接口需求。

 1、各机构的客户数，可以按支行、分行、总行进行逐层统计

 2、根据维度组合（卡、产品、机构、客户、币种）来统计交易笔数和交易金额

 3、客户日均余额

 4、贷款通过率

 5、等

1. DM层如何保证数据质量，如何保证数据是正确的？

 开发完成要经过多轮测试才能上线，包括UT、SIT、UAT测试，还会对作业进行监控，确保数据流程的正确

1. 怎么设计表的，怎么设计mapping的，表是谁设计的，DW层提供给报表那边的表涉及关联吗？

 这个主要是设计人员设计的，设计好目标表的字段以及来源于哪些表，DW表会向上关联汇总做成DM表然后提供给报表端。