# 西安雷格泰克软件有限责任公司 2018年-8月--2020年-8月 数据开发

1. 长安银行历史数据迁移项目 2018年-8月--2018年11月

项目描述：

由于银行进行整合以往数据，使用kettle把Oracle中的数据转移到Hive中，同时进行存储过程和函数的重新编写。

项目职责：

1. 需求分析
2. 表格的设计和文档的编写
3. 使用kettle进行oracle数据库和hive数据库的连接和数据迁移
4. 使用hql和hplsql进行存储过程的重新编写
5. 使用hive数据库重新编写之前旧的oracle数据库的需求
6. 对脚本进行自测
7. 长安银行保险数据分析项目 2018年-11月--2019--10月

项目描述：

根据长安银行提供的保险客户数据，对客户身份，客户需求，购买期限，保险类型等进行数据的分析，以便银行进行相关保险业务的推广，对不同用户身份进行不同保险业务的推荐。

项目职责：

1. 进行需求分析和相关表格的设计
2. 参与开发文档的编写
3. 使用hive数据库进行hql的语句编写
4. 对数据库进行性能调优
5. 使用shell脚本，对数据库的sql语句进行批量的数据
6. 对脚本进行自测
7. 长安银行信用卡用户数据分析项目 2019年-10月--2020-9月

项目描述：

对长安银行发行信用卡的用户信息进行收集和分析，主要是对用户对于信用卡开发、激活、使用等不同的数据，对新用户开卡、老用户回归、用户刷卡数量、频率、刷卡渠道等进行数据的分析，目的是使用不同的渠道的数据用于银行信用卡活动的推广，提高用户对于信用卡的使用率。

项目职责：

1. 参与到具体业务的需求分析过程中，进行项目需求和表格的设计
2. 参与开发设计文档的编写，日常维护数据字典
3. 使用hive数据库进行hql的语句编写，进行DM集市层的数据计算
4. 对数据倾斜产生的问题进行优化
5. 使用shell脚本，对数据库的sql语句进行批量的数据
6. 对脚本进行自测，并且对上线后的版本进行跟踪和维护

面试时候的项目介绍：

我先介绍下我最近做过的项目，这是个长安银行信用卡的项目，我在这个项目从事的是从DW层中取数，并且进行各种汇总和计算，最终写入到DM集市层中的工作，这个项目是使用的hadoop平台，里面是采用sqoop进行数据的全量和增量的抽取，使用hdfs进行数据的存储，使用mapreduce进行数据的计算，我的主要工作，是使用hql编写脚本，完成各项需求和业务指标的计算，在这个项目中，我主要负责的模块是事件主题库和用户主题库的部分表格和数据，平时做的多的事情例如不同渠道开卡的用户比例计算、每个月/季度开卡环比和同比计算、用户刷卡消费的计算、每个支行的信用卡开卡/激活/消费/还款/逾期的一些数据计算等，在项目中常用的工具例如sqoop、dbeaver、Python等等。

在你的这个项目中，一共有多少表？你所在的层有多少表？你的工作经常涉及的有多少表？

900多张表 ODS 100多200左右 DW 100多200左右 DM 300多左右 DA 300多左右

经常会用到的可能就是20多张表的样子

你经常操作的表的名字？

dw\_detail\_customer\_cards

用户id 用户姓名 手机号码 身份证 卡号 开卡日期 是否激活状态 激活日期 额度 开卡银行支行 联系人 联系人号码 公司 当前工资 公司号码...

dw\_detail\_customer\_cost

用户id 卡号 刷卡费用 手续费 是否跨境 剩余额度 刷卡币种 刷卡地点 刷卡消费类型 刷卡时间...

dm\_customer\_rc

主键id 年份 月份 数量 增长比例

这个表有哪些字段？

经常计算的指标有哪些？

用户相关（新增用户量：新增行数 用户活跃度：每日登录和消费的用户数量/总用户数量 PV：页面点击数量 UV：每日不同ip访问数量

不同渠道新增用户比例：不同渠道的用户数量/总新增用户数量）

费用相关（手续费计算：0.02%-0.3%之间，过户的手续费万分之3 个税计算：20%）

利润相关（利息收入 手续费收入 毛利润：某个时间段以内收入的数据 净利润：某个时间段以内收入的数据-支出数据）

财务相关（各部门的收入支出的预期结果和实际结果）

销售相关（销售团队或者个人以周或者月为单位，进行销售的预期和实际结果计算与对比）...

项目组有多少人？

数据开发 4-10个人

项目中数据量有多大？

我平时操作的表格中，比较大数据量的表，例如用户的零售表格，差不多在八九千万的样子。