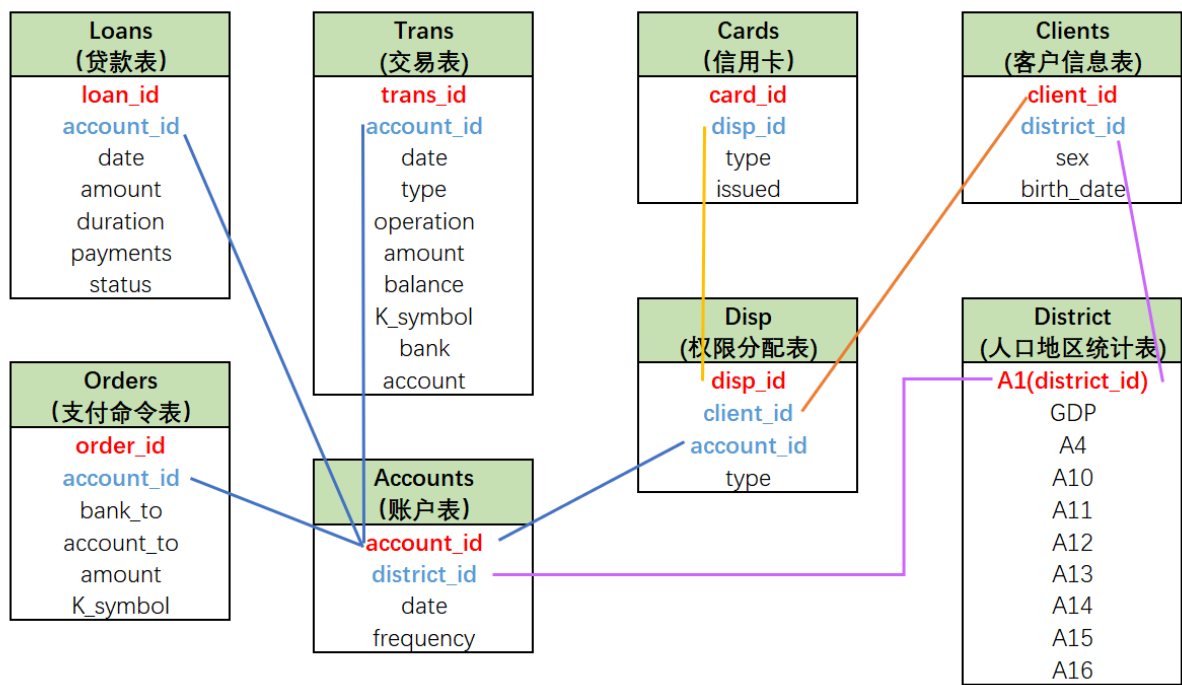


# KDD99 数据集

来自于美国1999年数据挖掘竞赛的金融数据部分；  
截取自一家银行的真实客户与交易数据；  
涉及客户主记录、帐号、交易、业务和信用卡数据；  
可用于做客户贷款违约预测、客户信用卡精准营销、客户细分等多种数据挖掘主题。

一个人可以有多个账户，一个account账户号可以对应有多个client顾客号，即多个人可以共管一个账户，账户与客户号的对应关系，在disp中进行列示。  
“loan” 和 “credit card” 为银行提供给客户的服务；  
一个账户可以多张信用卡，一个账户只能一笔贷款

实体-关系图（ER图）



## 表与字段说明

账户表(Accounts)	每条记录描述了一个账户的静态信息（4500条记录）
名称	标签
<i>account_id</i>	<i>账户号(主键)</i>
district_id	开户分行地区号

账户表(Accounts)	每条记录描述了一个账户的静态信息（4500条记录）
date	开户日期
frequency	结算频度（交易之后马上，周结，月结）

客户信息表（Clients）	每条记录描述了一个客户的特征信息（5369条记录）
列名	标签
<i>client_id</i>	<b>客户号(主键)</b>
Sex	性别（男，女）
birth_date	出生日期
district_id	地区号（客户所属地区）

信用卡(Cards)表		每条记录描述了一个账户上的信用卡信息（892条记录）
名称	标签	说明
<i>card_id</i>	<b>信用卡id</b>	<b>(主键)</b>
disp_id	账户权限号	
type	卡类型	普通卡 金卡 青年卡
issued	发卡日期	

权限分配表 (Disp)		每条记录描述了客户和账户之间的关系，以及客户操作账户的权限（5369条记录）
名称	标签	说明
<i>disp_id</i>	<b>权限设置号</b>	<b>(主键)</b>
client_id	顾客号	
account_id	账户号	
type	权限类型	所有者（可以进行增值业务操作和贷款） 用户

人口地区统计表 (District)	每条记录描述了一个地区的人口统计学信息 (77条记录)
列名	标签
<i>A1</i>	<i>地区号 (主键)</i>
GDP	GDP总量
A4	居住人口
A10	城镇人口比例
A11	平均工资
A12	1995年失业率
A13	1996年失业率
A14	1000人中有多少企业家
A15	1995年犯罪率 (千人)
A16	1996年犯罪率 (千人)

支付命令表 (Orders)		每条记录代表描述了一个支付命令 (6471条记录)
名称	标签	说明
<i>order_id</i>	<i>订单号</i>	<i>(主键)</i>
account_id	发起订单的账户号	
bank_to	收款银行	每家银行用两个字母来代表,用于脱敏信息 AB CD EF GH IJ KL MN OP QR ST UV WX YZ
account_to	收款客户号	
amount	金额	
K_symbol	支付方式	保险支付 贷款偿还 日常支出 租赁支付

交易表 (Trans)		每条记录代表每个账户上的一条交易 (1,056,320条记录)
名称	标签	说明

交易表 (Trans)		每条记录代表每个账户上的一条交易 （1,056,320条记录）
<i>trans_id</i>	交易序号	主键
account_id	发起交易的账户号	
date	交易日期	
type	借贷类型	借， 贷
operation	交易类型	从他行收款 信用卡借方 信贷资金 支取现金 汇款到另一家银行
amount	金额	
balance	账户余额	
K_Symbol	交易特征	保险费 养老金 利息所得 房屋贷款 支付贷款 支票
bank	对方银行	每家银行用两个字母来代表,用于脱敏信息 AB CD EF GH IJ KL MN OP QR ST UV WX YZ
account	对方账户号	

贷款表(Loans)		每条记录代表某个账户的上的一条贷款信息 （682条记录）
名称	标签	说明
<i>loan_id</i>	贷款号	(主键)
disp_id	权限号	
account_id	账户号	
date	发放贷款日期	
amount	贷款金额	
duration	贷款期限	以“月”为单位
payments	每月归还额	

贷款表(Loans)		每条记录代表某个账户的上的一条贷款信息（682条记录）
status	还款状态	A代表合同终止，没问题； B代表合同终止，贷款没有支付； C代表合同处于执行期，至今正常； D代表合同处于执行期，欠债状态。