

# 第5章-1：数据库基础

Python数据科学：全栈技术详解

讲师：Ben

# 自我介绍

- 天善商业智能和大数据社区      讲师 – Ben
- 天善社区 ID - Ben\_Chang
- <https://www.hellobi.com> – 学习过程中有任何相关的问题都可以提到技术社区数据挖掘版块。

# 主要内容

- 背景介绍
- 数据库基础知识
- 数据介绍

# 背景介绍

- 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;
- 一个账户可以办理多张信用卡
- 一个账户只能有一笔贷款

# 数据库基础知识



- 关系型数据库：
  - 将世界抽象为实体和关系，实体包括物和事件。一张二维表中存放着实体的集合，“行”为一个实体，“列”为属性和特征。表与表之间通过逻辑关系相连。
- 主键与外键：
  - 主键：指该表中唯一确定每条记录的一个变量或多个变量的组合，它具有无缺失、无重复的特征。经常用于数据清洗和数据横向合并。
  - 外键：是另一张表中与这张表的某个字段的类型，字段名相同的字段，一般用作关联两张或两张以上的数据表。

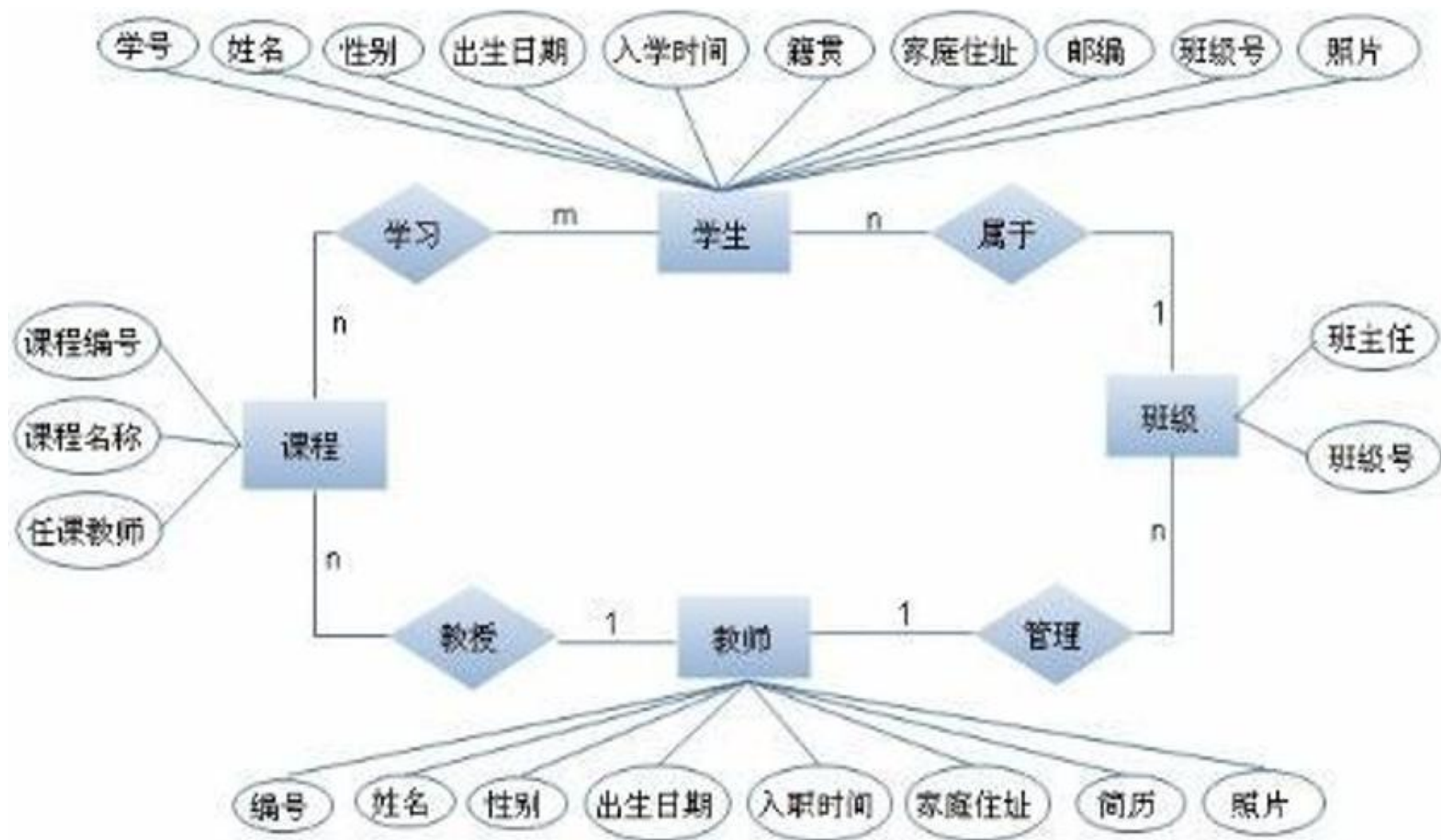


- E-R图

- **实体型**(Entity)：具有相同属性的实体具有相同的特征和性质，用实体名及其属性名集合来抽象和刻画同类实体；
- **属性**(Attribute)：实体所具有的某一特性，一个实体可由若干个属性来刻画。
- **联系**(Relationship)：联系也称关系，信息世界中反映实体内部或实体之间的联系。

- 联系可分为以下 3 种类型：
  - (1) 一对一联系(1 : 1)
    - 例如，一个部门有一个经理，而每个经理只在一个部门任职，则部门与经理的联系是一对一的。
  - (2) 一对多联系(1 : N)
    - 例如，某校教师与课程之间存在一对多的联系，即每位教师可以教多门课程，但是每门课程只能由一位教师来教
  - (3) 多对多联系(M : N)

# E-R图示例



# 数据介绍



## 账户表(Accounts)

- 每条记录描述了一个账户的静态信息
- 条数: 4500

## 客户信息表 (Clients)

- 每条记录描述了一个客户的特征信息
- 条数: 5369

### 账户表(Accounts)

名称	标签
account_id	账户号(主键)
district_id	开户分行地区号
date	开户日期
frequency	结算频度 (月, 周, 交易之后马上)

### 客户信息表 (Clients)

列名	标签
client_id	客户号(主键)
Sex	性别
birth_date	出生日期
district_id	地区号 (客户所属地区)

# 数据说明

## 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
- 条数: 5369

## 支付命令表 (Orders)

- 每条记录代表描述了一个支付命令
- 条数: 6471

## 权限分配表(Disp)

名称	标签	说明
disp_id	权限设置号	(主键)
client_id	顾客号	
account_id	账户号	
type	权限类型	只用"所有者"身份可以进行增值业务操作和贷款

## 支付订单表 (Orders)

名称	标签	说明
order_id	订单号	(主键)
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
account_to	收款客户号	
amount	金额	
K_symbol	支付方式	

# 数据说明

## 交易表 (Trans)

- 每条记录代表每个账户上的一条交易
- 条数:1056320

## 贷款表(Loans)

- 每条记录代表某个账户的上的一条贷款信息
- 条数： 682

### 贷款表(Loans)

名称	标签	说明
disp_id	权限号	(主键)
loan_id	贷款号	
account_id	账户号	
date	发放贷款日期	
amount	贷款金额	
duration	贷款期限	
payments	每月归还额	
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。

### 交易表 (Trans)

名称	标签
trans_id	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
operation	交易类型
amount	金额
balance	账户余额
K_Symbol	交易特征
bank	对方银行
account	对方账户号

## 信用卡(Cards)

- 每条记录描述了一个账户上的信用卡信息
- 条数: 892

## 人口地区统计表 (District)

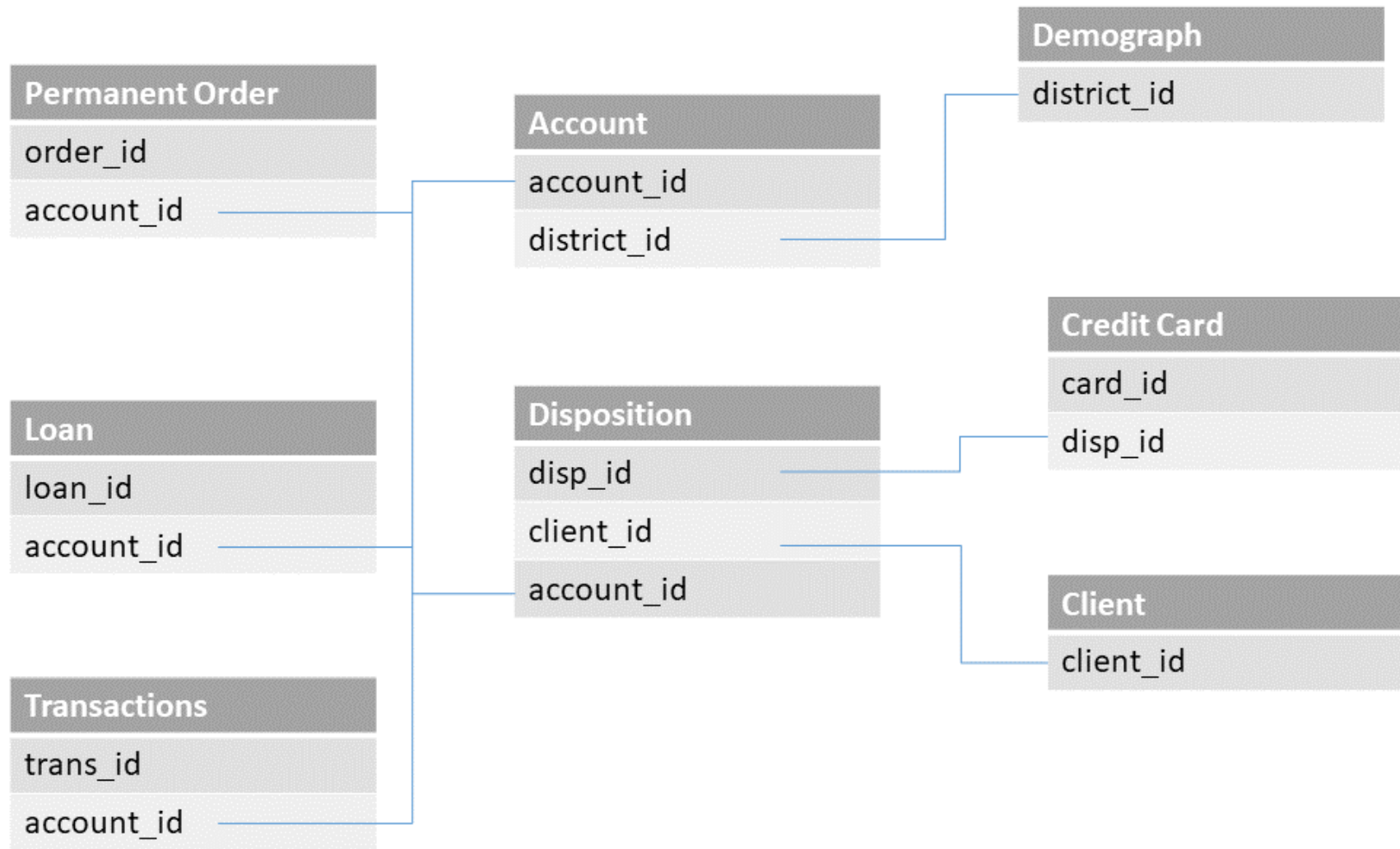
- 每条记录描述了一个地区的人口统计学信息
- 条数: 77

信用卡(Cards)表	
名称	标签
card_id	信用卡id(主键)
disp_id	账户权限号
type	卡类型
issued	发卡日期

人口地区统计表 (District)	
名称	标签
A1 = district_id	地区号(主键)
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犯罪率(千人)
A16	1996犯罪率(千人)



# 表的关系：数据的实体-关系图(ER图)



—— 秦路主讲 ——  
**七周成为数据分析师**  
七周为期，Get一条数据分析师职业黄金通道！



—— Python ——  
**数据分析与挖掘**  
集Python爬虫、数据采集、数据处理、数据分析与数据挖掘于一体，打造Python全栈工程师  
主讲老师: 韦玮  
VIP会员群+在线答疑+录播复习+1年反复观看



**案例为师, 实战为王**  
开启Python机器学习之路  
科学规划全套课程体系，从入门到进阶，从理论到技巧，嵌入丰富课程案例讲解，逐步推进  
讲师: 唐宇迪 深度学习领域多年一线实践研究专家



**独一无二的  
数据仓库**建模指南系列教程升级版  
• 从企业视角进行数据规划以及数据仓库模型的搭建  
• 高质量的数据库模型和技巧，以及丰富的例子  
• 数据仓库架构理论和实践要领  
资深讲师: BAO胖子 15年+BI从业经验  
涉足电力、快消品、医药、信息服务行业的BI老兵



**业务知识一站通**  
技术+业务，挣钱有门路！  
—— 讲师: 陈文 ——



自己动手 丰衣足食  
**Python3网络爬虫实战案例**  
— 循序渐进，案例为王，诠释全面，思路制胜 —  
讲师: 崔庆才 北航硕士，百万级热度爬文博主



讲师 丘祐玮  
**人人都爱数据科学家**  
Python数据科学精华实战课程



**数据分析报告制作**  
秘籍升级版  
讲师: 陈丹奕 知乎大神，前百度资深数据分析师



**先机致胜 破冰AI**  
—— 深度学习模型/框架与实战 ——  
讲师: 唐宇迪 同济大学硕士  
深度学习领域多年一线实践研究专家



BI、商业智能  
数据挖掘 大数据  
数据分析师  
R语言 Python  
机器学习  
深度学习  
人工智能  
Hive Hadoop  
Tableau  
BIEE ETL  
数据科学家  
PowerBI