

数据库技术-ER模型

赵亚伟

zhaoyw@ucas.ac.cn

中国科学院大学 大数据分析技术实验室

2017.11.26

目录

- **E-R模型**
- 约束
- **E-R图**
- **E-R设计问题**
- 弱实体集
- 扩展**E-R**特性
- **E-R模型与关系模式**

关于概念模型

- 概念模型是现实世界到计算机世界的一个中间层次，用于信息世界的建模，是现实世界到信息世界的第一层抽象，是数据库设计人员进行数据库设计的有力工具，也是数据库设计人员与用户之间进行交流的语言。因此，概念模型应具备两个基本特征：
 - 较强的语义表达能力，表达语义知识
 - 简单，清晰，易于用户理解

E-R模型

- 概念模型有很多，其中最为著名的是 **P.P.S.Chen**（陈品山）于**1976**年提出的实体-联系方法（**Entity-Relationship Approach**）
- 该方法用**E-R**图描述了现实世界的概念模型，称为**E-R**模型。**E-R**模型满足上述概念模型特征，是一种概念模型，也是一种语义模型。
- （关系模型，**1970**年，**E.F. Codd**）



E-R模型的关键词：语义

E-R模型的三要素

- E-R模型包括：实体集、联系集、属性三个要素。
 - 实体是现实世界中可区别于其他对象的事物
 - 实体集是具有相同性质(或属性)的实体集合.
 - 属性是实体集中每个成员具有的描述性性质，
将一个属性赋予某实体集表明数据库为实体集中每个实体存储相似信息
 - 联系是指多个实体间的相互关联
 - 联系集是同类型联系的集合

实体及实体集

- 实体通过一组属性来表示，实体可以是实实在在的，如人或书；也可以是抽象的，如贷款、假期或概念。
- 实体集之间可以相交，如银行的员工实体集 **employee** 和所有客户的实体集 **customer**，一个 **employee** 的实体也可能是 **customer** 的实体

联系及联系集

□ 联系的描述:

联系集是 $n(n \geq 2)$ 个实体集上的数学关系, 如果 E_1, E_2, \dots, E_n 为 n 个实体集, 那么联系集 R 是:

$\{(e_1, e_2, \dots, e_n) | e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$

的一个子集, 其中 (e_1, e_2, \dots, e_n) 是一个联系, 对比广义笛卡尔积与关系

□ 联系是通过一组实体来描述的, 不是实体属性。如联系集***borrower***的一个元素（联系）为

$((\underline{321-12-2123}, \underline{Jones, Main, Harrison}), (\underline{L-17}, \underline{1000})) \in \text{borrower}$

□ 原则: 将所有有联系的实体都表达出来。

属性

- ❑ 与关系模型中的属性具有相似含义，每个属性都有一个可取值的集合，称为该属性的域(**domain**)、或者值集(**value set**)
- ❑ 在**E-R**模型中，属性可能不是原子的，即属性可以进行划分，如复合属性可以化为简单属性
- ❑ 属性的值也可以取**null**

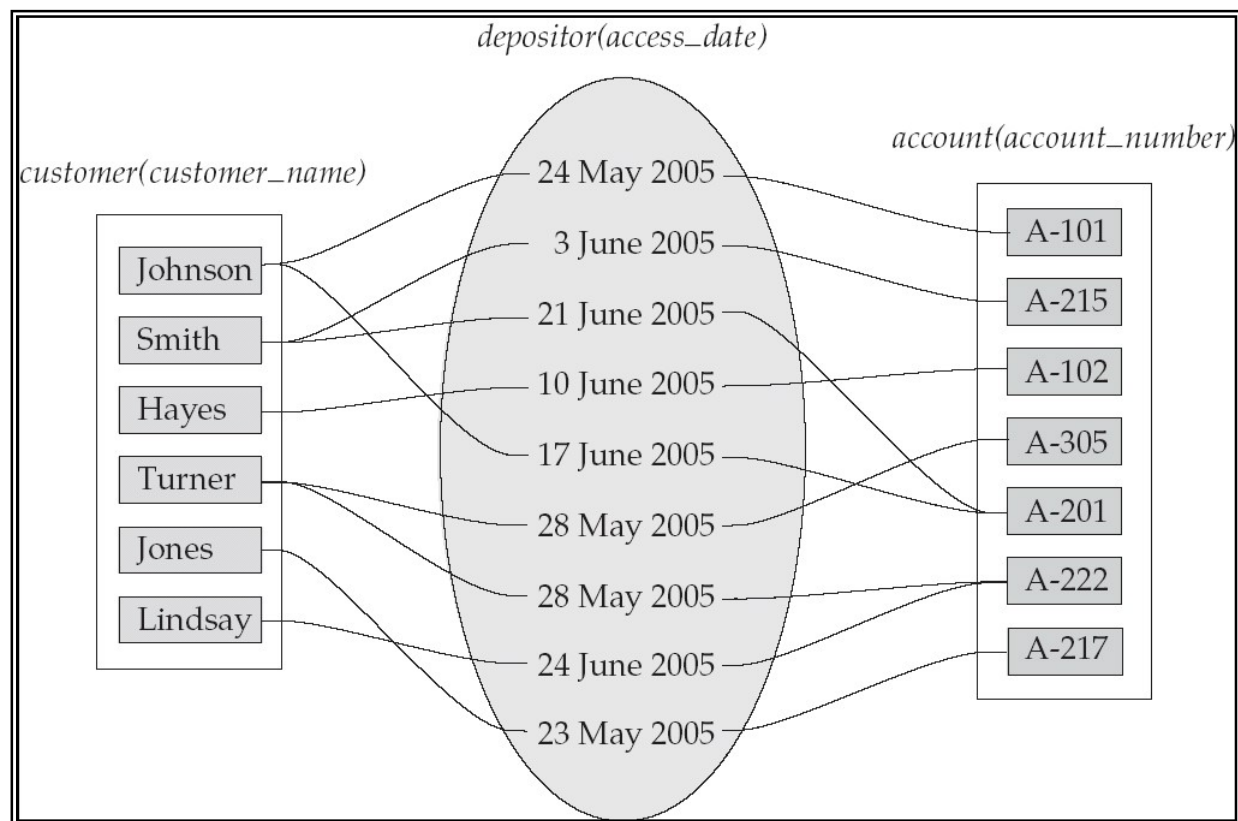
常见属性类型

- ❑ 简单属性: 不能划分为更小的部分的属性
- ❑ 复合属性: 可以再划分为更小的部分的属性, 如地址, 可以有层次
- ❑ 单值属性: 对一个特定的实体只有单独一个取值的属性, 如学号
- ❑ 多值属性: 有多个取值的属性, 如电话
- ❑ 派生属性: 从其他属性或实体派生出来的属性, 如某实体具有“出生日期”属性, 根据该属性可以计算出“年龄”, “年龄”则为“出生日期”的派生属性

联系集的描述属性

- 联系也可以有属性 (例如夫妻关系 有个登记号或日期) For instance, the *depositor* relationship set between entity sets *customer* and *account* may have the attribute *access-date*
- 这个属性称之为描述性属性，描述性属性要显示地标识出来，而参与联系的实体集的属性在联系集中不被显示地标识

例子：联系集的描述属性



access_date属性描述了客户最近一次访问的时间

目录

- E-R模型
- 约束
- E-R图
- E-R设计问题
- 弱实体集
- 扩展E-R特性
- E-R模型与关系模式

E-R模型-约束-映射基数

- 映射基数（势）：一个实体集的实体通过一个联系集与另一个实体相联系的实体的数目。
- 映射基数用于描述二元联系集，也用于多个实体集的联系集的描述，但二元联系是基础，即多个实体集的联系集可以转换为二元联系集。
- 对于实体集A与B之间的二元联系集R来说，映射基数必然是下列情况之一：
 - 一对一
 - 一对多
 - 多对一
 - 多对多

E-R模型-约束-映射基数

- 一对一：A中的一个实体至多与B中的一个实体相联系
- 一对多：A中的一个实体可以与B中的任意数目的实体相联系
- 多对一：A中的一个实体至多与B中的一个实体相联系,而B中的一个实体可以与A中的任意数目的实体相联系
- 多对多：A中的一个实体可以与B中的任意数据的实体相联系，B中的一个实体也可以与A中的任意数目的实体相联系

E-R模型-码

- ❑ 码用于区别实体, 也可以唯一地标识联系
- ❑ 超码是一个或多个属性的集合, 这些属性的组合可以唯一地标识一个实体
- ❑ 候选码: 最小超码, 任意真子集都不能构成超码的超码
- ❑ 主码: 用来在同一实体集中区分不同实体的候选码

码是实体集的性质, 而不是单个实体的性质

E-R模型-主码描述联系

- 有了码后，可以用码描述联系而不用实体的方式（太麻烦）
- 如果没有属性同联系集**R**相联，那么属性集合 $Primary-key(E1) \cup Primary-key(E2) \cup \dots \cup Primary-key(E_n)$ 描述了联系集**R**中的某个联系
- 如果属性 $a1, a2, \dots, a_m$ 同集合**R**相联，那么属性集合 $Primary-key(E1) \cup Primary-key(E2) \cup \dots \cup Primary-key(E_n) \cup \{a1, a2, \dots, a_m\}$ 描述了联系集**R**中的某个联系
- $Primary-key(E1) \cup Primary-key(E2) \cup \dots \cup Primary-key(E_n)$ 构成了联系集的一个超码

联系集的主码结构

- 联系集的主码结构依赖于联系集的映射基数，即不同的映射基数主码结构也不同。
- 如多对一（一对多）的联系集的主码选择“多”一方的实体集的主码作为主码，多对多选两个实体集的主码的并，一对一则选哪一个实体集的主码都可以
- 原则：将所有有联系的实体都表达出来。

目录

- E-R模型
- 约束
- **E-R图**
- E-R设计问题
- 弱实体集
- 扩展E-R特性
- E-R模型与关系模式

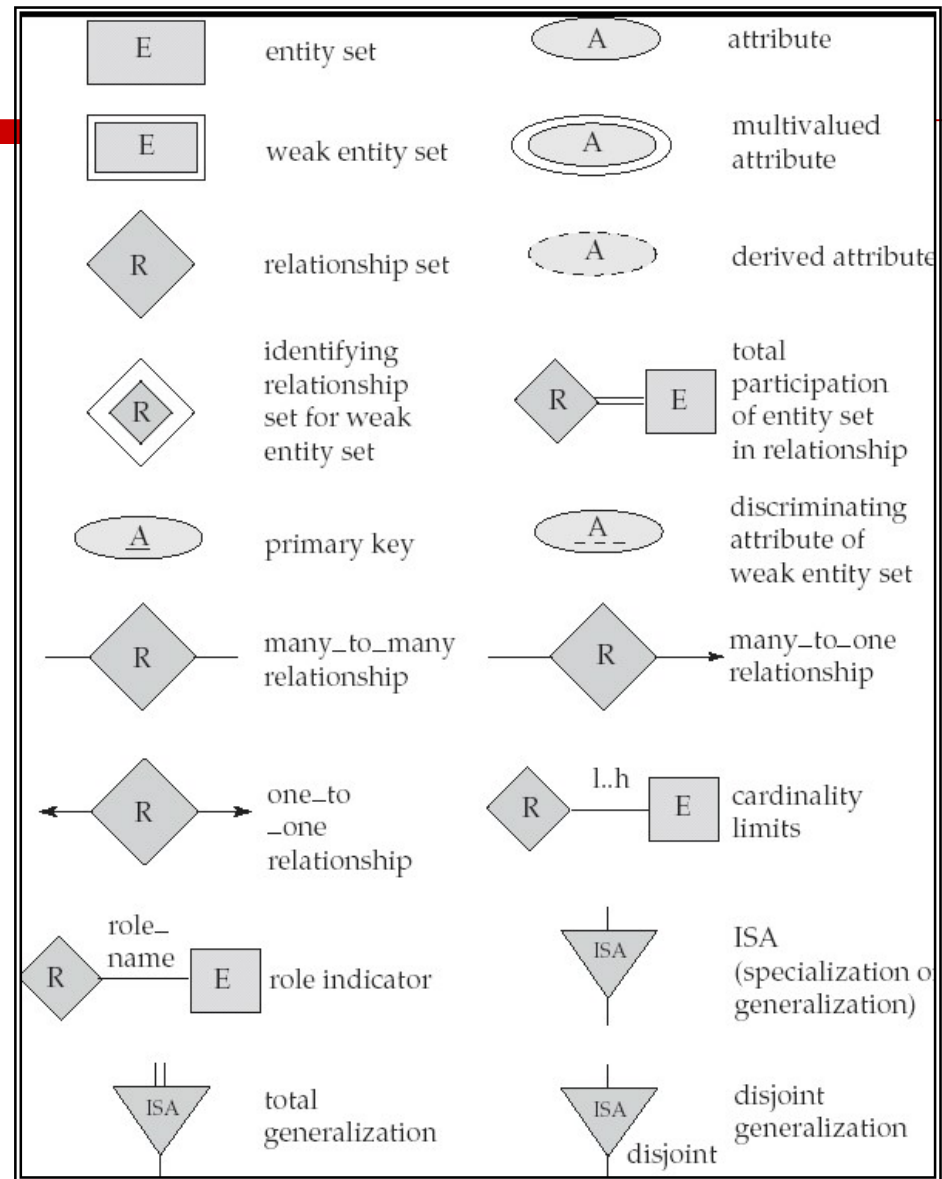
E-R图类型

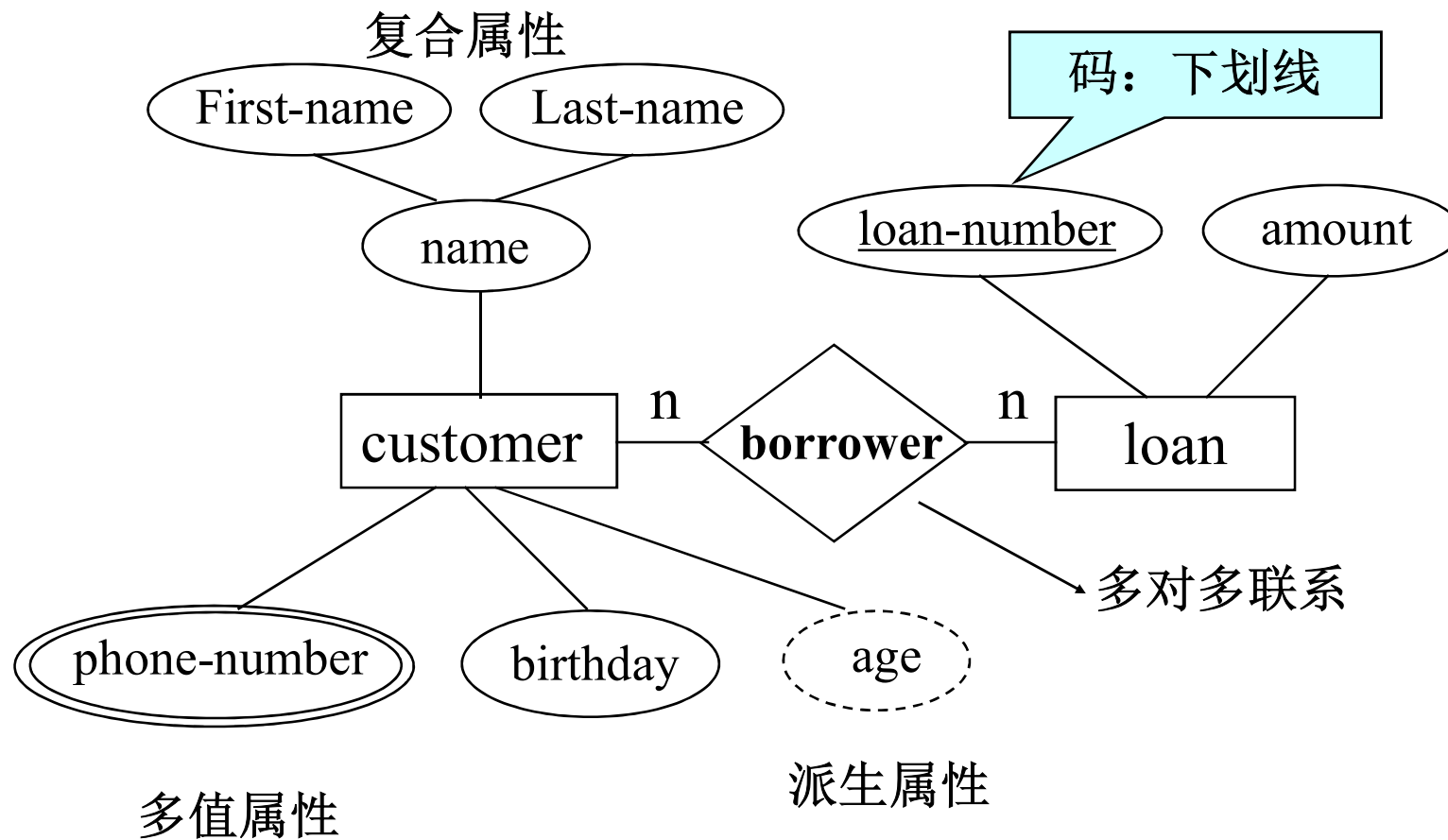
- **E-R图**目前还没有统一标准，常见的有两种类型：
 - **陈氏E-R图**：Peter Chen，课程例子
 - **IDEF1X**：IDEF是ICAM DEFinition method 的缩写，是美国空军在70年代末80年代初ICAM (Integrated Computer Aided Manufacturing)工程在结构化分析和设计方法基础上发展的一套系统分析和设计方法。是比较经典的系统分析理论与方法。**IDEF1X**是IDEF系列方法中**IDEF1**的扩展版本，是在**E-R (实体联系)**法的原则基础上，增加了一些规则，使语义更为丰富的一种方法。用于建立系统信息模型。

E-R图组件

- 数据库的全局逻辑结构（概念层次）可以通过**E-R图**做图形化表示
- 由于**E-R图**简单清晰使**E-R模型**得到广泛应用
- **E-R图**的几个主要组件：
 - 矩形:表示实体集
 - 椭圆:表示属性
 - 菱形:表示联系集
 - 线段:将属性连接到实体集或将实体集连接到联系集
 - 双椭圆:表示多值属性
 - 虚椭圆:表示派生属性
 - 双线:表示一个实体全部参与到联系集中
 - 双矩形:表示弱实体集

Summary of Symbols Used in E-R Notation





客户和贷款的E-R图

目录

- E-R模型
- 约束
- E-R图
- E-R设计问题
- 弱实体集
- 扩展E-R特性
- E-R模型与关系模式

数据库E-R模式的设计

- 数据库E-R模式设计的步骤
 - 数据需求，产生用于需求规格说明书
 - 概念设计
 - 实体集设计
 - 联系集设计
 - E-R图（先画实体联系，再画属性）
 - 完善的概念模式应指明企业的功能需求，即功能需求说明书也作为概念设计的一个产品

E-R模式设计问题

- 实体集与联系集的概念不精确，定义一组实体及相互联系可以有多种不同的方式。这就涉及一些设计问题：
 - 用实体集还是用属性
 - 用实体集还是用联系集
 - 二元联系集与n元联系集
 - 联系属性的布局

用实体集还是用属性

- 例子：一个实体集**employee**具有**employee_name**和 **telephone_number**等属性，显然电话可以作为一个单独的实体，具有属性**telephong_number**和**location**(办公室)。那么，必须重新定义为：
 - 实体集**employee**，具有属性**employee_name**
 - 实体集**telephone**，具有属性**telephone_number**和**location**
 - 联系集**emp_telephone**，表示员工及其电话间的联系

用实体集还是用联系集

- ❑ 一个对象最好表述为实体集还是联系集并不总是非常清楚的
- ❑ 一个基本原则是，当描述发生在实体间行为时采用联系集
- ❑ 由于描述性属性可能会导致联系集的数据冗余及数据不一致性可通过规范化方法进行处理。
- ❑ 实际设计中一个经常犯的错误是忽略了联系集。

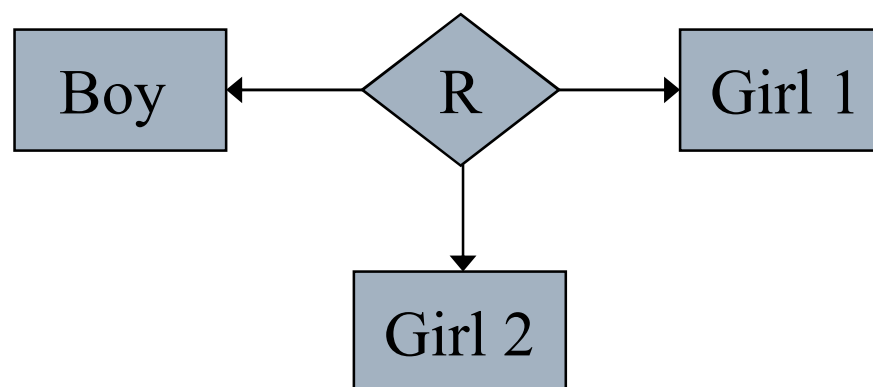
关于Ternary Relationship

- 至多只有一个箭头从三重联系出来，作为 约束（防歧义）
- E.g. **works-on** → **job** 表示一人至多一个工作
- 如有多个箭头出来，有多种解释
 - 其中的一个联系另外两个，或
 - 其中的两个的组合联系另外一个
- 产生歧义
- 学习方法, 书上抽象符号严格，不易理解

例子：多个箭头的三元联系

□ 如右图，至少有以下两种解释：

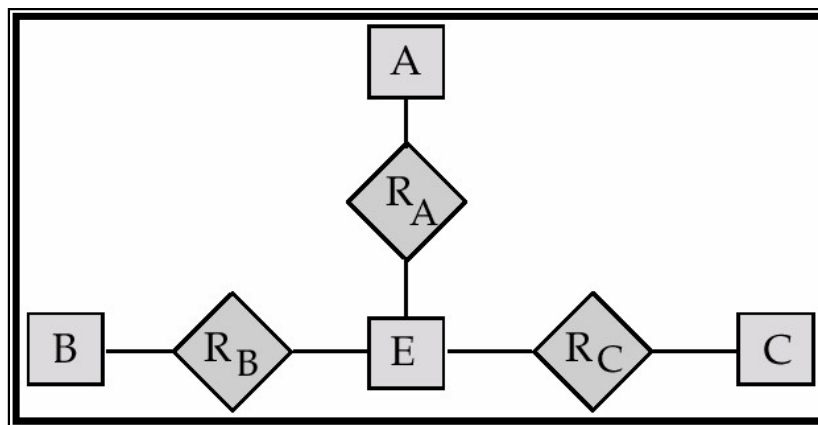
- 每个 **boy** 有两个女朋友 **girl1** 和 **girl2**（或每个 **girl** 有一个男朋友和一个女朋友，...）
- 每对情侣 (**boy_n**, **girl1_n**) 有一个女朋友 **girl2_n**



- 若只有一个箭头（假设在 **boy** 边）则只有一种等价解释：
 - 每个 **boy** 有多个 **girl1** 女朋友和多个 **girl2** 女朋友
- 没有 1:1 的组合情况，消除了歧义

二元联系集与n元联系集

- 数据库中的联系通常是二元的，一些非二元的联系实际上可以用多个二元关系更好地表示
- 如一个三元联系集合 R ，将实体集 A 、 B 和 C 联系起来，可以用实体集 E （人造实体）替代联系集 R ，并重新建立三个联系集
 - R_A ，联系 E 和 A
 - R_B ，联系 E 和 B
 - R_C ，联系 E 和 C



- 概念上，我们可以限制 $E-R$ 模型中只包含二元联系集，但实际中也存在增加设计复杂性和数据冗余、表述不够清晰以及约束转移困难等问题

目录

- E-R模型
- 约束
- E-R图
- E-R设计问题
- 弱实体集
- 扩展E-R特性
- E-R模型与关系模式

弱实体集

- ❑ 一个实体集的属性可能不足以形成主码，这样的实体集就称为弱实体集
- ❑ 存在主码的实体集称为强实体集
- ❑ 弱实体集必须与另一个称为标识实体集（强实体集）或属主实体集关联才有意义
- ❑ 弱实体集用分辨符区分实体，如{**payment-number**}

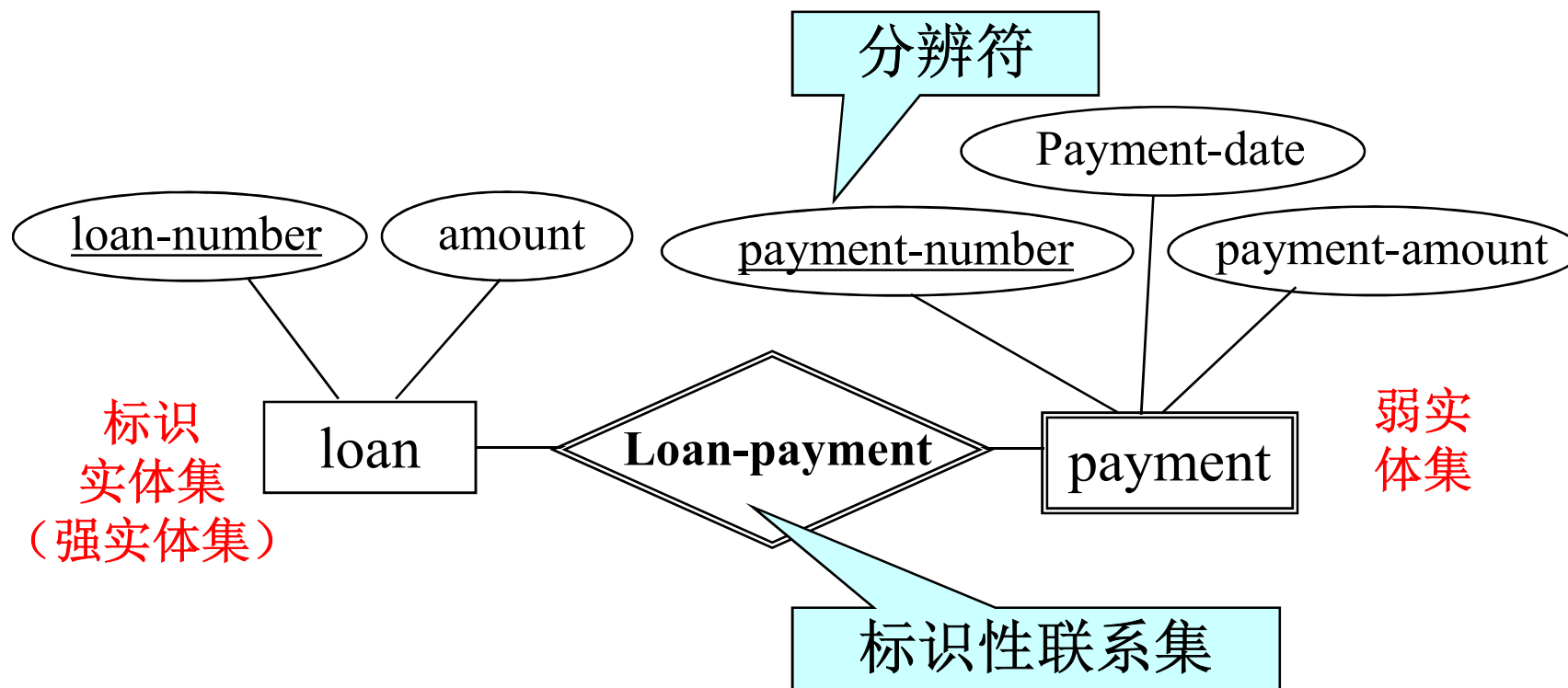
关于弱实体的直观解释

- 例1：一位董事长（强实体） 的 k 号助理（弱实体）， $k=1,2,3,\dots$
- 例2：在旧社会， 妇女弱势群体（弱实体）， **K-th wife of Some One.....**
- 例3：教学楼附楼（强实体）， 教室（弱实体， 阶2-3， 弱实体区分码）

标识性联系

- ❑ 将弱实体集与其标识实体集相联的实体集称为标识性联系。
- ❑ 标识性联系没有描述性属性，弱实体描述。
- ❑ 标识性联系是多对一联系：弱实体 到 强实体 为 **m:1**
- ❑ 标识性联系集合的属性为强实体集的主码和弱实体集的分辩符。（弱实体集属性包含了）
- ❑ 弱实体集还可以参与到标识性联系以外的其他联系。

弱实体集例子



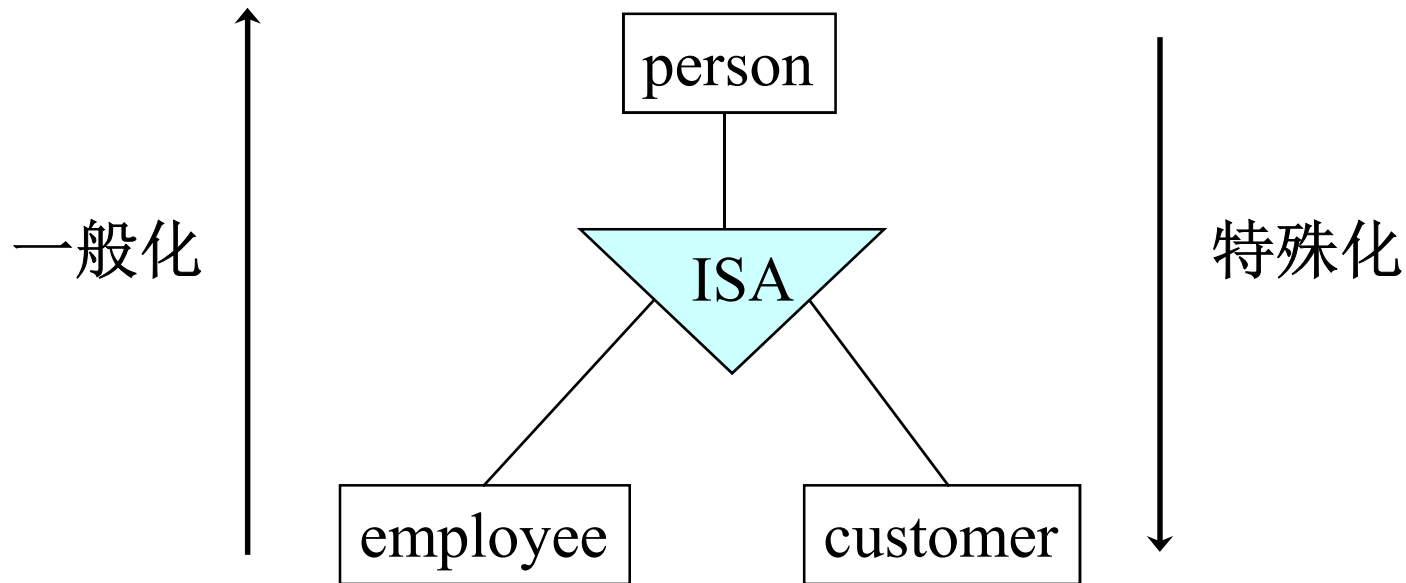
什么情况下采用弱实体集？

- 如果弱实体集只参与标识性联系，而且其属性不多，那么在建模时将其描述为一个属性更为恰当。
- 相反，如果弱实体集参与到标识性联系以外的联系中，或者属性比较多时，则建模时将其表述为弱实体集更恰当。
- 语义角度理解，**E-R**模型是语义模型、概念模型

目录

- E-R模型
- 约束
- E-R图
- E-R设计问题
- 弱实体集
- 扩展E-R特性
- E-R模型与关系模式

特殊化(一般化)的图示



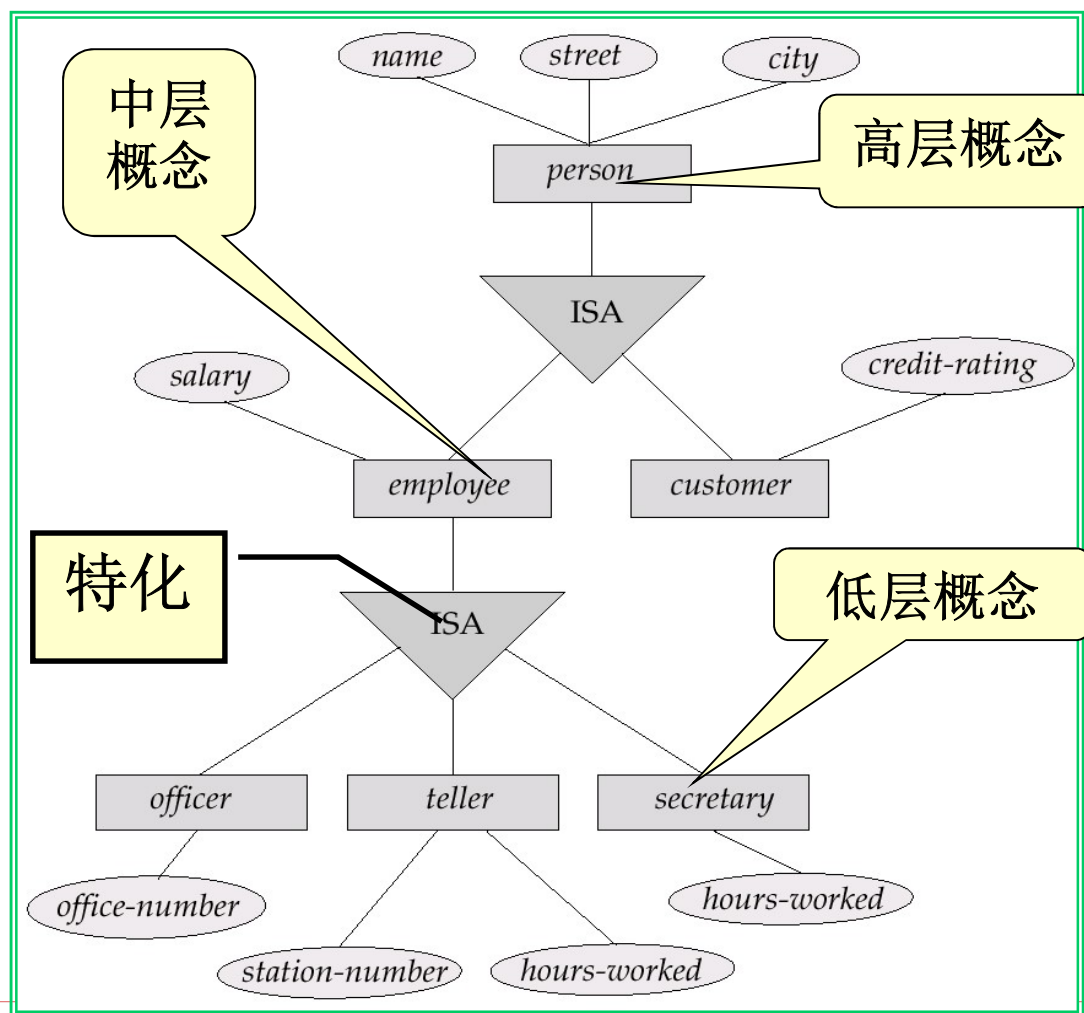
- 三角符号表示父类-子类联系，表示**employee**“is a”**person**，员工“是一个”人

属性继承

- 除了父类的码外其他属性均被子类继承，子类还可以有自己的附加属性
- 在层次结构中，如果底层的子类只参与了一个**ISA**联系，则仅继承单一的父类属性，称为单继承，若参与多个**ISA**联系，则继承了多个父类属性，称为多继承

例子：特殊化、一般化、属性继承

- 层次结构
- **Employee** 继承了 **person** 的 **name**、**street**、**city** 属性，同时还有自己的附加属性 **salary**

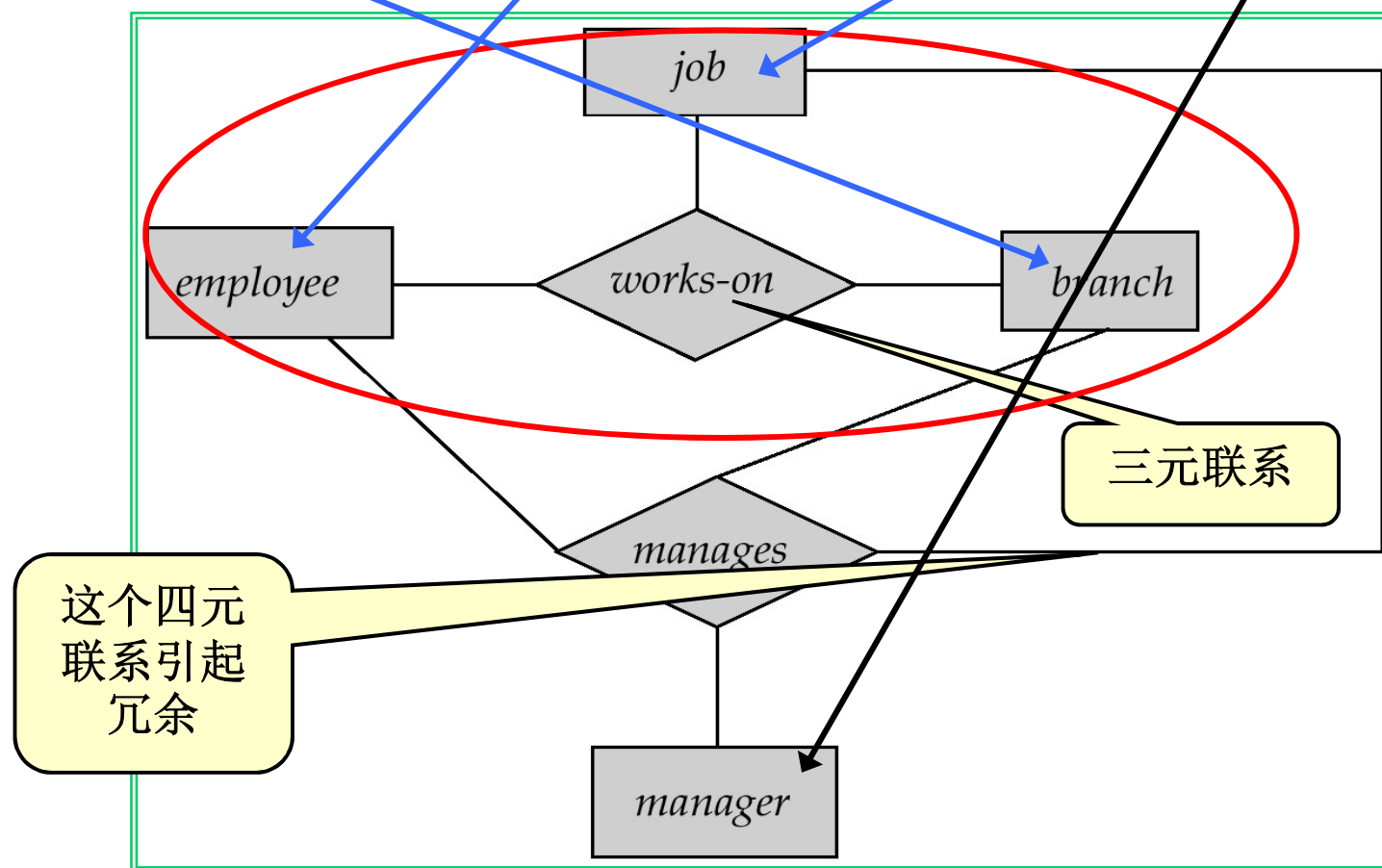


聚集

- ❑ **E-R**模型的一个局限：不能表达联系间的聚集
- ❑ **E-R**模型的聚集是图的聚集，聚集是一种抽象，通过这种抽象，联系被当作高层的实体来看待。
- ❑ 注意：**E-R**模型的聚集不是关系代数中的聚集，不是一个概念。

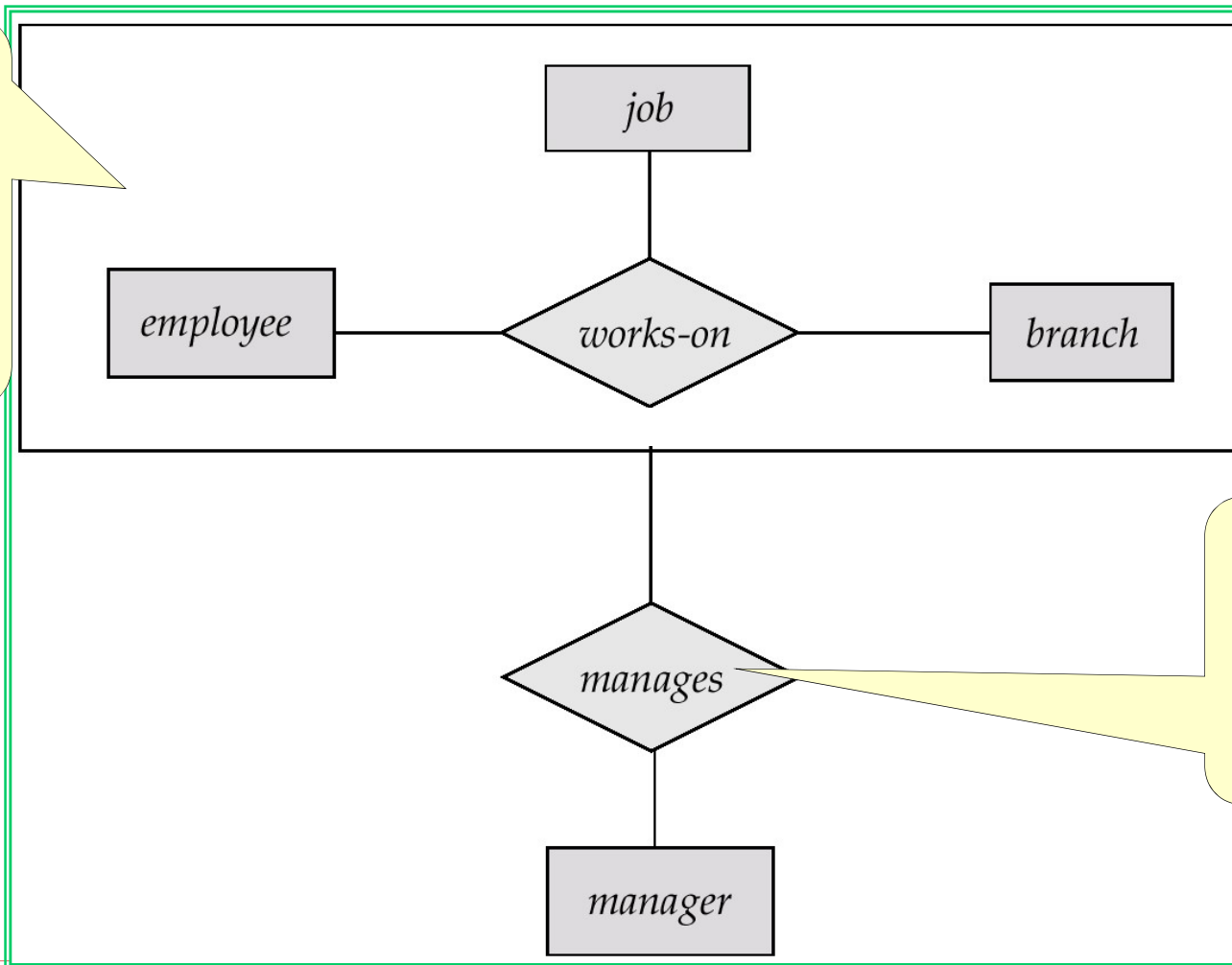
例子：聚集

□ 在一个支行 的职工 所作的工作的 管理员



例子：聚集

把框内对象看作一个实体
给一个简洁的名称
职工信息



从四元
联系变
为二元
联系

目录

- E-R模型
- 约束
- E-R图
- E-R设计问题
- 弱实体集
- 扩展E-R特性
- E-R模型与关系模式

将E-R图转换为表

- ☐ 用表表示强实体集
- ☐ 用表表示弱实体集
- ☐ 用表表示联系集（主码、模式冗余、模式合并）
- ☐ 复合属性
- ☐ 多值属性
- ☐ 一般化
- ☐ 聚集

用表表示强实体集

用n个不同列的表E来表示这个实体集，如(L11,1500)表示贷款号为L11的贷款金额为1500元. 通过在表中增加一行,表示增加一个新实体,也可以进行删除和修改

Loan-number	amount
L-11	900
L-14	1500
L-15	1500
L-16	1300
L-17	1000
L-23	2000
L-93	500

Loan表

用表表示弱实体集

设A是具有属性a1, a2, ..., am的弱实体集, 设B是A所依赖的强实体集, 其主码包含属性b1, b2, ..., bn. 用表A表示实体集A, 表中各列对应以下属性集合中的各个属性:

$\{a1, a2, \dots, am\} \cup \{b1, b2, \dots, bn\}$

弱实体集的分辩符

强实体集
loan的属性
(主码)

loan-number	payment-number	payment-date	payment-amount
L-11	53	2001.7.7	125
L-14	69	2001.5.28	500
L-15	22	2001.5.23	300
L-16	58	2001.6.18	135
L-17	5	2001.5.10	50
L-17	6	2001.6.7	50
L-17	7	2001.6.17	100
L-23	11	2001.5.17	75
L-93	103	2001.7.3	900
L-93	104	2001.6.13	200

用表表示联系集

设**R**是联系集,而所有参与**R**的实体集的主码属性集合为**a1, a2, ..., am**,如果**R**有描述性属性**b1,b2,...,bn**,用表**R**表示该联系集,表中各列对应于以下属性集合中的各个属性:

$$\{a1, a2, ..., am\} \cup \{b1, b2, ..., bn\}$$

customer-id	Loan-number
019-28-3746	L-11
019-28-3746	L-23
244-66-8800	L-93
321-12-3123	L-17
335-57-7991	L-16

borrower表

复合属性

- ❑ E-R图中的属性通常可以直接映射到相应表中的列,但对复合属性和多值属性则需要进行处理
- ❑ 通过为每个子属性创建一个独立的属性来处理复合属性,每个独立属性对应一列。拉平
- ❑ 但复合属性本身则不作为单独的列存在. 如复合属性**address**由**street**和**city**组成,则创建两个独立属性,则表中将包括**address-street**和**address-city**列,而没有单独的一列**address**.
- ❑ 处理后的复合属性的子属性在一个表中

多值属性

❑ 多值属性需要通过创建新表来解决。

踢出

❑ 对于一个多值属性 **M**, 为其创建新表 **T**, **T** 中由一列 **C** 与 **M** 对应, 而 **T** 中其余列则对应于以 **M** 为多值属性的实体集和联系集的主码.

...	Phone-number
...	001
...	002
...	...

customer表

Pnumber	number
001	88256088
001	88256068
001	88256064
002	62310268
002	62310266
...	...

phonenumbers表

一般化的表示方式：法1

□ 方法1:

- 为高层实体集创建一个模式
- 低层实体建表，含高层实体主关键字和局部属性

□ 例子：根据E-R图为每个实体建模式（这里只建一部分，派生属性，继承属性略）

- $person = (\underline{person_id}, name, street, city)$
- $customer = (\underline{person_id}, credit_rating)$
- $employee = (\underline{person_id}, salary)$

□ 若想得到customer的完整结果，连接操作

□ 缺点：了解职工信息，查两表

一般化的表示方式：法2

□ 方法2:

- 为高层实体集创建一个模式
- 低层实体建表，含高层实体主关键字和全部属性（包括继承来的）

□ 例子：根据E-R图为每个实体建模式

- $person = (\underline{person_id}, name, street, city)$
- $customer = (\underline{person_id}, name, street, city, credit_rating)$
- $employee = (\underline{person_id}, name, street, city, salary)$

□ 如果特化是全参与的，无必要建立高层表

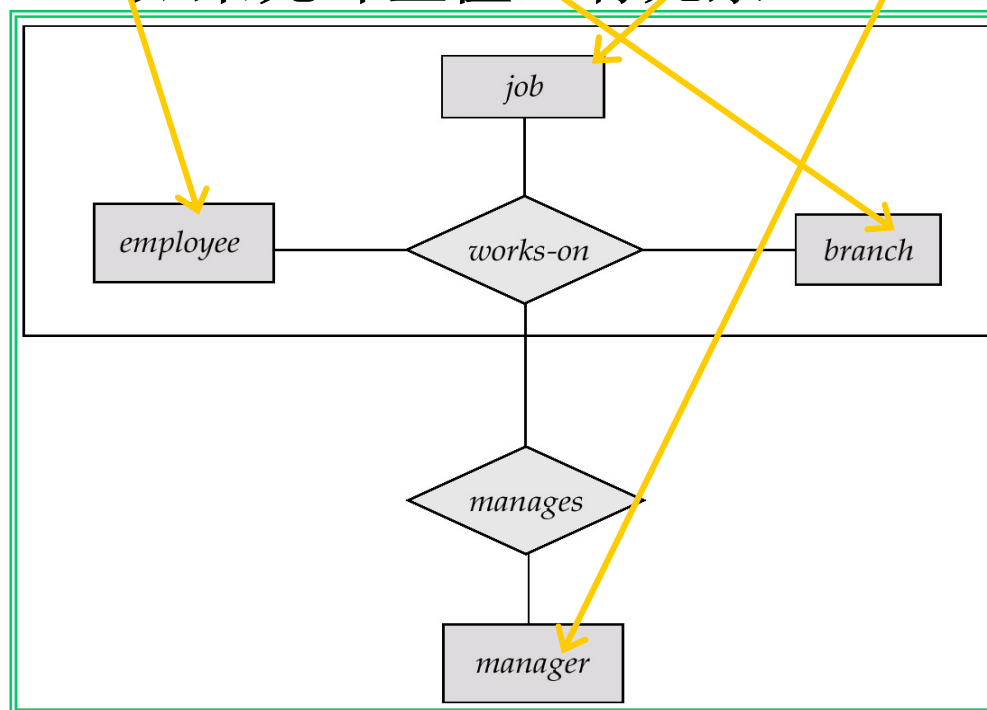
□ 缺点，冗余信息

聚集表示

- ❑ To represent aggregation, create a table containing, 建表, 包含
 - primary key of the aggregated relationship, 联系的主属性
 - the primary key of the associated entity set, 联系实体主关键字
 - Any descriptive attributes 描述性属性
- ❑ 见下页

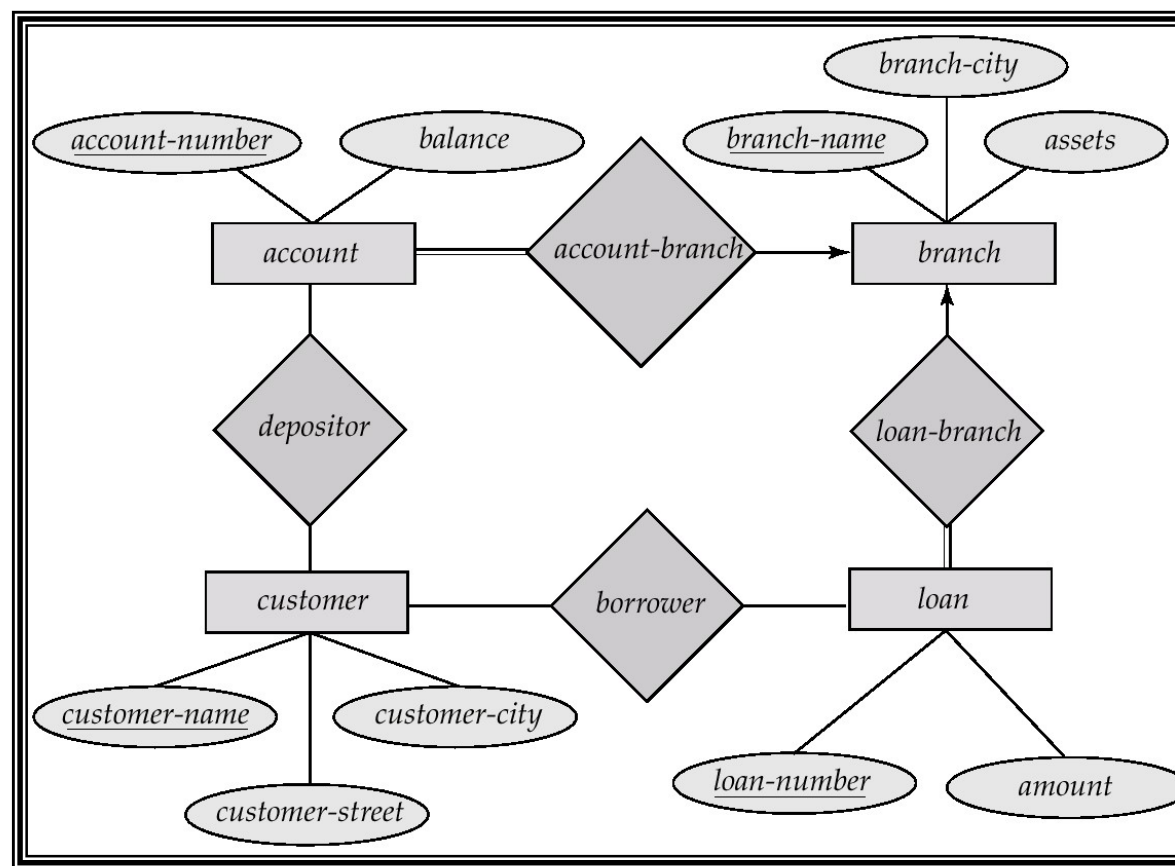
例子：聚集表示

- ❑ 建表表示聚集
- ❑ *manages(employee-id, branch-name, title, manager-name)*
- ❑ Table *works-on* 如果允许空值，有冗余



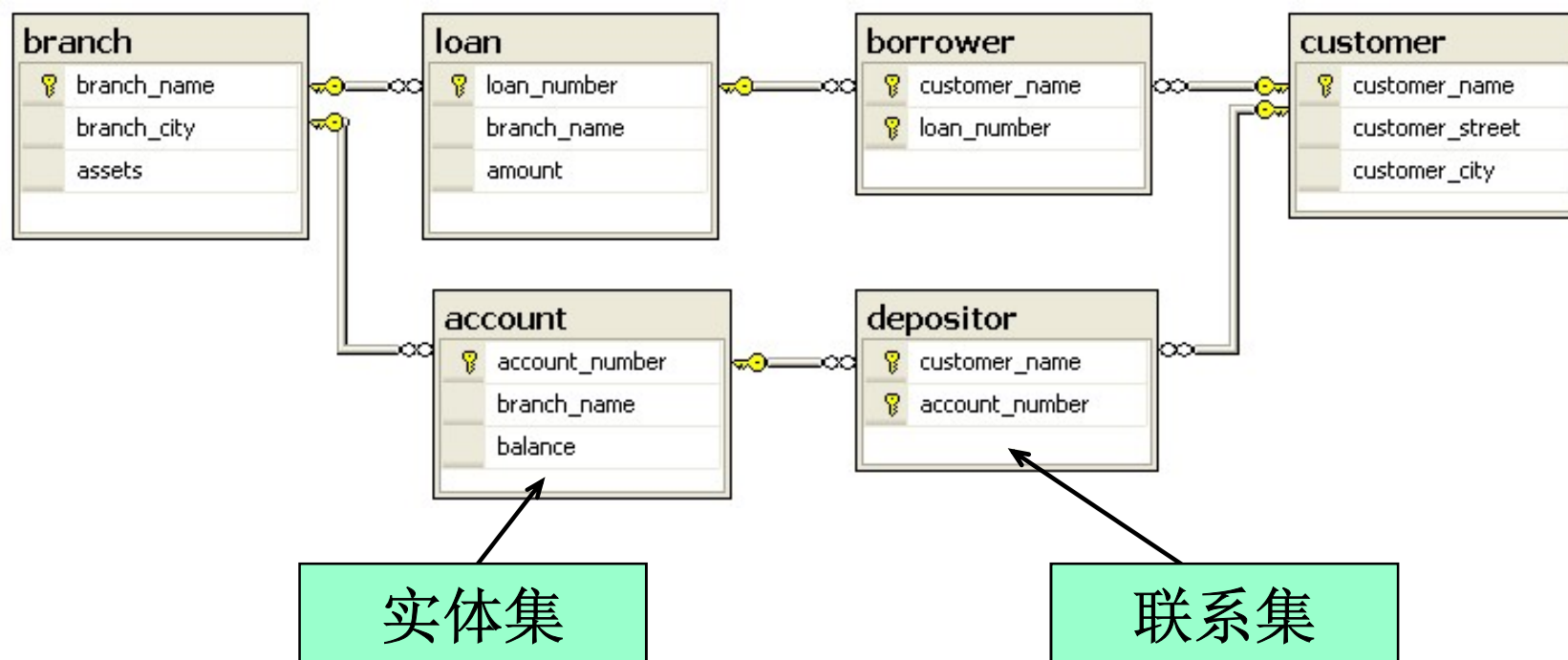
E-R模型—E-R图

- 一个银行企业例子（简化的），下图为概念设计的E-R图



E-R模型到表映射的例子

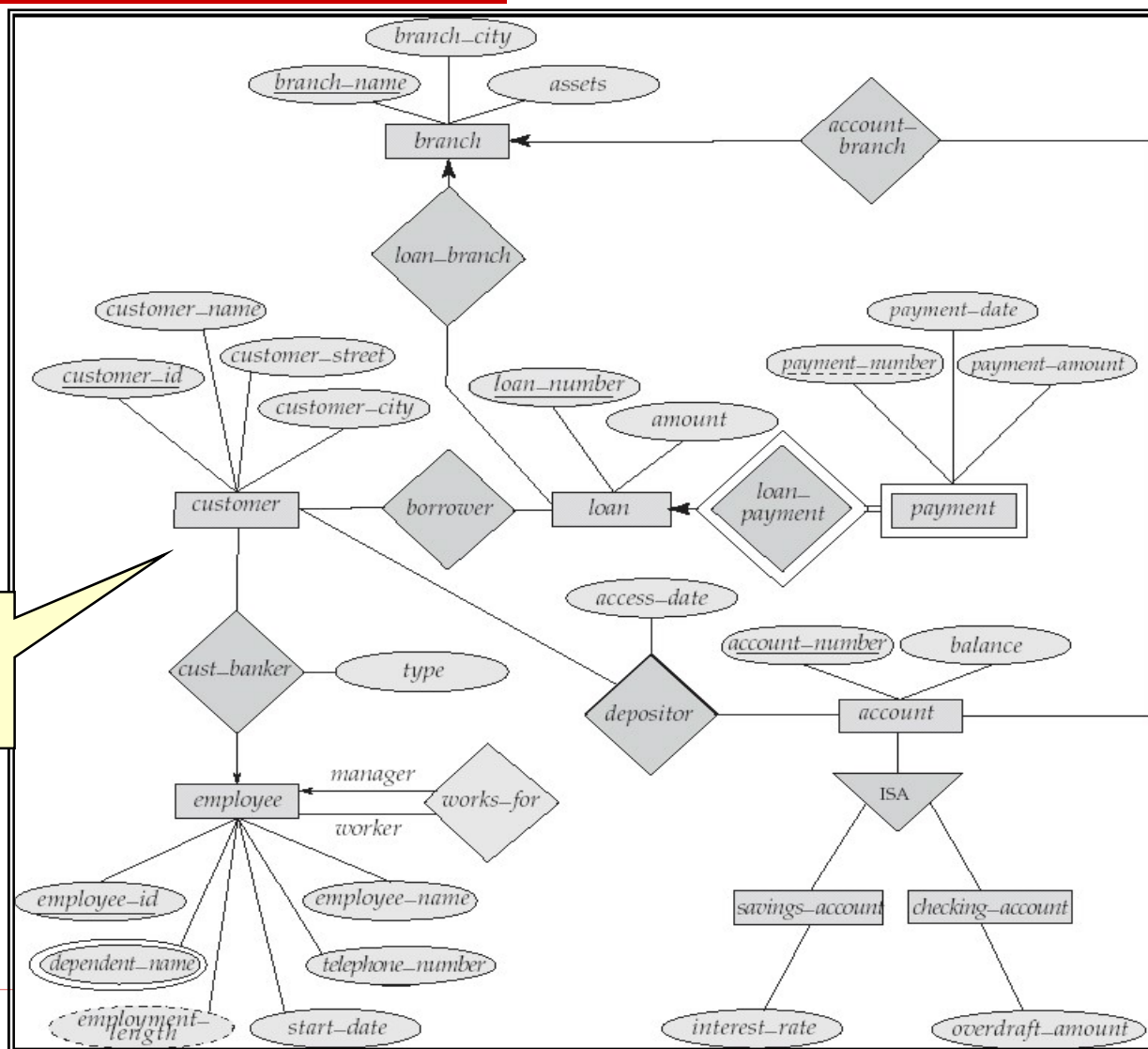
- 上述E-R图在SQL Server 2005中的具体实现（模式图）



例子：银行企业数据库设计

□ 银行数据库设计中的E-R图

试着分析每个实体及联系



小结

- **E-R模型**基于对现实世界这样一种认识：世界由一组基本对象(称为实体)及这些对象间的联系组成
- **E-R模型**目的是服务于数据库设计过程，企业模式可以用**E-R模型**描述
- **E-R模型**的三要素：实体集,联系集,属性
- **E-R图**描述分类：
 - 实体集和属性
 - 联系
 - 基数约束
 - 一般化和特殊化
- **E-R模型**本质上是语义模型或概念模型