

国家“十三五”重点出版规划项目
上海高校优秀教材奖获得者主编

上海市高校精品课程
特色教材(立体化新形态)

数据库原理及应用

——基于SQL Server 2016

主编 贾铁军 徐方勤

副主编 邓红霞 戴春妮 王佃来 曹锐

参编 陶维天 张野 降爱莲



机械工业出版社
CHINA MACHINE PRESS

第9章 关系数据库的规范化



9.1 数据库规范化问题

9.2 函数依赖概述

9.3 关系模式的分解

9.4 关系模式的范式及规范化



- 了解数据库关系模式存在的异常问题及形成原因
- 理解函数依赖的相关概念
- 掌握关系模式的分解
- 掌握关系模式范式的概念及规范化过程

重点





9.1 数据库规范化问题

上海市高校精品课程
国家十三五规划项目

■ 9.1.1 规范化理论研究的内容

关系数据库设计理论即关系的规范化理论，是根据现实世界存在的数据依赖关系进行关系模式的规范化处理。在实际应用中，针对给定的业务应用环境和用户需求，需要建立一个合适的数据库模式，使数据库系统能够高效地存储和管理数据。为使数据库设计合理可靠，通过长期的数据库应用，形成了关系数据库的规范化理论。对规范化理论的研究，主要涉及的是关系模式中各属性之间的依赖关系及其对关系模式的影响，讨论良好的关系模式应具备的特性，以及达到良好关系模式的方法。

■ 9.1.1 规范化理论研究的内容

关系数据库设计理论具有很好的应用价值。关系模式的规范化理论主要包括各属性之间的依赖关系，以及对关系模式性能的影响，提供判断关系模式优劣的理论标准，预测可能出现的问题，提供自动产生各种模式的算法。其中，关系数据库设计理论的核心是数据间的函数依赖；衡量的标准是关系规范化的程度及分解的无损连接和保持函数依赖性；模式设计方法是自动化设计的基础。

9.1 规范化的主要问题

上海市高校精品课程
国家十三五规划项目

■ 9.1.2 关系模式的异常问题

【案例9-2】设有一个关系模式：读者（读者号，姓名，单位，地址，图书号，借书时间，还书时间），主键为(读者号，ISBN)。设同一个单位的读者具有同样的地址。该关系模式部分数据下表所示。

读者号	姓名	单位	地址	图书号	借书时间	还书时间
2010121036	尹梁栋	软件学院	2号楼	SC10801	2016-04-15 13:23:43	
2010121036	尹梁栋	软件学院	2号楼	WC13401	2016-04-21 08:43:12	
2008521432	杜晓晓	外语学院	4号楼	WC17401	2015-12-15 10:42:53	2016-01-15 18:32:02
2008521432	杜晓晓	外语学院	4号楼	WC17401	2014-12-11 18:43:23	2015-01-05 15:54:23
2008521432	杜晓晓	外语学院	4号楼	TG54324	2011-10-23 14:27:52	2012-11-11 09:54:24
2008521432	杜晓晓	外语学院	4号楼	TG42563	2012-08-14 11:24:14	2012-10-06 12:15:54
2009121221	刘小雪	汽车学院	5号楼	WC18401	2015-09-24 13:43:19	2015-09-30 12:05:50
2009121221	刘小雪	汽车学院	5号楼	NH87532	2015-12-16 14:57:14	2016-01-15 15:21:03
2012060907	徐冬	外语学院	4号楼	NH87533	2012-10-28 15:35:39	2012-11-13 13:57:28
2011031234	王大山	外语学院	4号楼	SC12501	2013-10-28 20:24:13	2013-11-13 14:53:37
2012180508	曹秉慧	商学院	3号楼	WC17402	2015-11-30 18:38:43	2015-12-30 13:56:34

9.1.2 关系模式的异常问题

一个关系模式若设计不当，将会出现数据冗余、异常、不一致等异常问题：

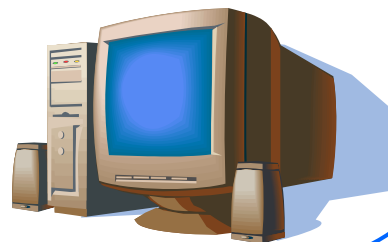
1. **数据冗余**：数据冗余不但急剧增加数据库的数据量，耗费大量的存储空间和运行时间，而且容易造成数据的不一致或其他异常问题，增加数据查询和统计的复杂度。
2. **更新异常**：对于数据冗余多的数据库，当执行数据修改操作时，容易导致部分信息被修改，部分信息没有修改的问题，从而造成数据库数据不一致，影响数据的完整性。
3. **插入异常**：插入的数据不能正常地被插入到数据库中。
4. **删除异常**：在删除某种数据的同时将其他数据也删除了。

9.1.2 关系模式的异常问题

一个好的关系模式应具备以下四个方面的具体条件要求：

- (1) 数据冗余尽可能的少。
- (2) 更新数据时没有更新异常。
- (3) 插入数据时没有插入异常。
- (4) 删除数据时没有删除异常。

解决方法：通过**模式分解**的方法进行规范化



9.1 规范化的主要问题

上海市高校精品课程
国家十三五规划项目

讨论思考:

- (1) 规范化理论的主要内容是什么?
- (2) 不良的关系模式可能存在什么问题?
- (3) 良好的关系模式的设计要求是什么?



9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.1 函数依赖的概念

定义9-1 设 X, Y 是关系 R 的两个属性集合，当任何时刻 R 中的任意两个元组中的 X 属性值相同时，则它们的 Y 属性值也相同，则称 X 函数决定 Y ，或 Y 函数依赖于 X ，记为 $X \rightarrow Y$ 。

特别理解：

此定义说明：若 $R(U)$ 为关系模式， U 是 R 的属性集合， X 和 Y 是 U 的子集。对于 $R(U)$ 的任意一个可能的关系 r ，如果 r 中不存在两个元组，且在 X 上的属性值相同，而在 Y 上的属性值不同，则称“ X 函数决定 Y ”或“ Y 函数依赖于 X ”，记做 $X \rightarrow Y$ 。

定义9-1 设 X, Y 是关系 R 的两个属性集合， X' 是 X 的真子集，存在 $X \rightarrow Y$ ，但对每一个 X' 都有 $X' \not\rightarrow Y$ ，则称 Y 完全函数依赖于 X ，记为 $X \xrightarrow{f} Y$ 。反之，若存在 $X' \not\rightarrow Y$ ，则称 Y 部分函数依赖于 X ，记为 $X \xrightarrow{p} Y$ 。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.1 函数依赖的概念

【案例9-3】 设有一个关系模式：读者（读者号，姓名，性别，单位，地址，书号，借书日期，还书日期），主键为(读者号，ISBN)，则函数依赖关系有：

读者号 \rightarrow 姓名

姓名函数依赖于读者号

(读者号, 书号) \rightarrow 姓名

姓名部分函数依赖于读者号和书号

(读者号, 书号) \rightarrow 借书日期

借书日期完全函数依赖于读者号和书号

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.2 函数依赖的逻辑蕴含

通常，函数依赖是以命题形式定义的，可将两个函数依赖集之间存在着一些互为因果的关系称为逻辑蕴含，即一个函数依赖集逻辑地蕴含另一个函数依赖集。如函数依赖集 $F=\{A\rightarrow B, B\rightarrow C\}$ 和 $\{A\rightarrow B, B\rightarrow C, A\rightarrow C\}$ 相互逻辑蕴含。

定义9-3 设 F 是在关系模式 R 上成立的函数依赖的集合， $X\rightarrow Y$ 是一个函数依赖。若对于 R 的每个满足 F 的关系 r 也满足 $X\rightarrow Y$ ，则称 F 逻辑蕴含 $X\rightarrow Y$ ，记为 $F \models X\rightarrow Y$ 。

定义9-4 设 F 是函数依赖集，被 F 逻辑蕴含的函数依赖全体构成的集合，称为函数依赖集 F 的闭包（closure），记为 F^+ 。即 $F^+ = \{ X\rightarrow Y \mid F \models X\rightarrow Y \}$ 。

定义9-5 对于函数依赖 $X\rightarrow Y$ ，若 $Y\subseteq X$ ，则称 $X\rightarrow Y$ 是一个平凡的函数依赖，否则称为非平凡的函数依赖。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.3 函数依赖的推理规则

设 U 是关系模式 R 的属性集， F 是 R 上成立的只涉及到 U 中属性的函数依赖集。函数依赖的推理规则（基本公理）有以下三条：

(1) A1（自反律，Reflexivity）：若 $X \rightarrow Y$ 在 R 上成立。

根据这条规则，可以推导出一些平凡函数依赖。由于(Φ 为空属性集, U 为全集)，所以 $X \rightarrow \Phi$ 和 $U \rightarrow X$ 都是平凡函数依赖。

(2) A2（增广律，Augmentation）：若 $X \rightarrow Y$ 在 R 上成立，且，则 $XZ \rightarrow YZ$ 在 R 上成立。

注意：有一些特殊情形。例如，当时，若 $X \rightarrow Y$ ，则对于 U 的任何子集 W 有 $XW \rightarrow Y$ 。在 $W=Z$ 时，若 $X \rightarrow Y$ ，则 $XW \rightarrow YW$ 。若 $X \rightarrow Y$ ，则 $X \rightarrow XY$ 。

(3) A3（传递律，Transitivity）：若 $X \rightarrow Y$ 和 $Y \rightarrow Z$ 在 R 上成立，则 $X \rightarrow Z$ 在 R 上成立。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.3 函数依赖的推理规则

【案例9-4】 设有一个关系模式：读者（读者号，姓名，性别，单位，地址），主键为(读者号)，假设一个单位只有一个地址，则函数依赖关系有：

读者号→姓名

读者号→单位

单位→地址

姓名函数依赖于读者号

单位函数依赖于读者号

地址函数依赖于单位

因此有

读者号 $\xrightarrow{\text{传递}}$ 地址

地址传递函数依赖于读者号

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.3 函数依赖的推理规则

定理9-1 函数依赖推理规则A1、A2和A3是正确的。即，若 $X \rightarrow Y$ 是从 F 用推理规则导出，则 $X \rightarrow Y$ 在 F^+ 中。

若给定关系模式 $R(U, F)$ ， X 、 Y 为 U 的子集， $F = \{X \rightarrow Y\}$ ，则
 $F^+ = \{X \rightarrow \Phi, X \rightarrow X, X \rightarrow Y, X \rightarrow XY, Y \rightarrow \Phi, Y \rightarrow Y, XY \rightarrow \Phi, XY \rightarrow X, XY \rightarrow Y, XY \rightarrow XY\}$

定理9-2 函数依赖的其他五条推理规则。

- (1) A4（合并性规则）： $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$ 。
- (2) A5（分解性规则）： $\{X \rightarrow Y, ZY\} \models X \rightarrow Z$ 。
- (3) A6（伪传递性规则）： $\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$ 。
- (4) A7（复合性规则）： $\{X \rightarrow Y, W \rightarrow Z\} \models XW \rightarrow YZ$ 。
- (5) A8（通用一致性规则）： $\{X \rightarrow Y, W \rightarrow Z\} \models X \cup (W - Y) \rightarrow YZ$ 。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.3 函数依赖的推理规则

【案例9-5】 设有关系模式 $R(A, B, C, D, E)$ 及其上的函数依赖集 $F=\{AB \rightarrow CD, A \rightarrow B, D \rightarrow E\}$, 求证 F 必蕴含 $A \rightarrow E$ 。

证明：由于 $A \rightarrow B$ （已知）
所以 $A \rightarrow AB$ （增广率）
因为 $AB \rightarrow CD$ （已知）
所以 $A \rightarrow CD$ （传递率）
所以 $A \rightarrow C, A \rightarrow D$ （分解规则）
因为 $D \rightarrow E$ （已知）
所以 $A \rightarrow E$ （传递率）

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.3 函数依赖的推理规则

【案例9-6】 已知关系模式 $R(ABC)$, $F = \{A \rightarrow B, B \rightarrow C\}$, 求 F^+ 。

根据FD的推理规则, 可以推出 F 的 F^+ 有43个FD。

例如, 据规则A1可推出 $A \rightarrow \varphi$, $A \rightarrow A$, ...。

根据已知的 $A \rightarrow B$ 及规则A2, 可以推出 $AC \rightarrow BC$, $AB \rightarrow B$, $A \rightarrow AB$, ...。

并根据已知条件及规则A3可推出 $A \rightarrow C$ 等。

给定关系模式 $R(U, F)$, A 、 B 、 C 为 U 的子集, $F = \{A \rightarrow B, B \rightarrow C\}$, 则依据上述关于函数依赖集闭包计算公式, 可以得到 F^+ 由如下43个函数依赖组成。

9.2 函数依赖概述

9.2.4 属性集的闭包

定义9-6 设 F 是属性集 U 上的函数依赖集， X 是 U 的子集，则（相对于 F ）属性集 X 的闭包用 X^+ 表示，为一个从 F 集使用函数依赖推理规则推出的所有满足 $X \rightarrow A$ 的属性 A 的集合： $X^+ = \{ \text{属性} A \mid X \rightarrow A \text{ 在 } F^+ \text{ 中} \}$ 。

定理9-4 $X \rightarrow Y$ 可用函数依赖推理规则推出的充分必要条件是 $Y \subseteq X^+$ 。

算法9-1 求属性集 X 相对于FD集 F 的属性集闭包 X^+ 。

```
result=X
```

```
do
```

```
{
```

```
  If  $F$ 中有某个函数依赖 $Y \rightarrow Z$ 满足 $Y \subseteq \text{result}$ 
```

```
    then  $\text{result} = \text{result} \cup Z$ 
```

```
} while (result有所改变);
```

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.4 属性集的闭包

【案例9-7】 已知中 $U=\{A, B, C, D, E\}$, $F=\{AB\rightarrow C, B\rightarrow D, C\rightarrow E, EC\rightarrow B, AC\rightarrow B\}$, 求 $(AB)^+$ 。

解：

设 $X=AB$

因为 $X(0)=AB$

$X(1)=ABCD$

$X(2)=ABCDE$

$X(3)=X(2)=ABCDE$

所以 $(AB)^+=ABCDE=\{A, B, C, D, E\}$

【案例9-8】 属性集 U 为 $ABCD$, FD 集为 $\{A\rightarrow B, B\rightarrow C, D\rightarrow B\}$ 。

则用上述算法, 可求出 $A^+=ABC$, $(AD)^+=ABCD$, $(BD)^+=BCD$ 等。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.5 候选键的求解和算法

定义9-7 设关系模式 R 的属性集是 U ， X 是 U 的一个子集。
若 $X \rightarrow U$ 在 R 上成立，则称 X 是 R 的一个超键。若 $X \rightarrow U$ 在 R 上成立，但对于 X 的任一真子集 X_1 都有 $X_1 \rightarrow U$ 不成立，则称 X 是 R 上的一个候选键。

注意：在本章提到的键都是指候选键（不含有多余属性的超键）。



9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.5 候选键的求解和算法

【案例9-9】对案例9-2的关系模式：读者（读者号，姓名，单位，地址，图书号，借书时间，还书时间），如何确定其候选键？

解：

由于每位读者有一个对应的姓名、单位和地址，一个读者对一本图书可以进行多次借阅，每次借阅有相应的借书时间和还书时间，可知（读者号，图书号）可以函数决定该关系模式的其他属性。在（读者号，图书号）集上增加其他属性，如（读者号，图书号，姓名）、（读者号，图书号，借书时间，还书时间）等属性集，虽然也能唯一标识关系中的元组，但包含了多余的属性，只能称为超键，而不能成为候选键。根据定义，不含多余属性（多个属性的最小集合）的超键为候选键，可知该关系模式的候选键即为（读者号，图书号）。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.5 候选键的求解和算法

快速求解候选键的一个充分条件：

对给定的关系模式 $R(A_1, \dots, A_n)$ 和FD集 F ，可将其属性分为四类：

- (1) L类：仅出现在函数依赖集 F 左部的属性。
- (2) R类：仅出现在函数依赖集 F 右部的属性。
- (3) N类：在函数依赖集 F 左右都未出现的属性。
- (4) LR类：在函数依赖集 F 左右都出现的属性。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.5 候选键的求解和算法

定理9-5 对于给定的关系模式R及其FD集F

- (1) 若 $X (X \in R)$ 为L类属性, 则X必为R的任一候选键的成员。
- (2) 若 $X (X \in R)$ 为L类属性, 且 X^+ 包含R的全部属性, 则X必为R的惟一候选键。
- (3) 若 $X (X \in R)$ 为R类属性, 则X不在任何候选键中。
- (4) 若 $X (X \in R)$ 为N类属性, 则X包含在R的任一候选键中。
- (5) 若 $X (X \in R)$ 为R的N类和L类属性组成的属性集, 且 X^+ 包含了R的全部属性, 则X为R的惟一候选键。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

【案例9-10】 设关系模式R的属性集 $U = (A, B, C, D)$ ，函数依赖集 $F = \{A \rightarrow C, C \rightarrow B, AD \rightarrow B\}$ 。求R的候选键。

解：

(1) 检查F发现，A、D只出现在函数依赖的左部，所以为L类属性，而F包含了全属性，即不存在N类的属性。

(2) 根据求属性闭包的算法，F中 $A \rightarrow C$ ， $AD \rightarrow B$ 可以求得 $(AD)^+ = ABCD = U$ ，而在AD中不存在一个真子集能决定全属性，故AD为R的候选键。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

【案例9-11】 设关系模式R的属性集 $U=(A, B, C, D, E, F)$,
 $F=\{ A \rightarrow B, D \rightarrow B, EF \rightarrow D, B \rightarrow D, DA \rightarrow F \}$ 。求R的候选键。

解：

(1) 检查F发现，A、E只出现在函数依赖的左部，所以为L类属性。C为N类的属性。

(2) 根据求属性闭包的算法，F中 $A \rightarrow B$, $B \rightarrow D$, $DA \rightarrow F$ ，可求得 $(ACE)^+ = ABCDEF = U$ ，而在ACE中不存在一个真子集能决定全部属性，故ACE为R的候选键。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.6 函数依赖推理规则的完备性

推理规则的**正确性**是指从函数依赖集 F ，利用推理规则集推出的函数依赖必定在 F^+ 中，**完备性**是指 F^+ 中的函数依赖都能从 F 集使用推理规则集导出。即正确性保证推出的所有函数依赖都正确，完备性则可保证推出所有被蕴含的函数依赖，以保证推导的有效性和可靠性。



9.2 函数依赖概述

9.2.6 函数依赖推理规则的完备性

定理9-6 函数依赖推理规则{A1, A2, A3}是完备的。

完备性的证明，即证明不能从F使用推理规则过程推出的函数依赖不在F+中成立。

设F是属性集U上的一个函数依赖集，有一个函数依赖 $X \rightarrow Y$ 不能从F中使用推理规则推出。现在要证明 $X \rightarrow Y$ 不在F+中，即 $X \rightarrow Y$ 在模式R(U)的某个关系r上不成立。因此可以采用构造r的方法来证明。

(1) 证明F中每个FD $V \rightarrow W$ 在r上成立。

由于V有两种情况： VX^+ ，或 VX^+ 。若 VX^+ ，根据定理9-4有 $X \rightarrow V$ 成立。根据已知的 $V \rightarrow W$ 和规则A3，可知 $X \rightarrow W$ 成立。再根据定理9-4，有 WX^+ ，所以 VX^+ 和 WX^+ ，同时成立，则 $V \rightarrow W$ 在r是成立的。若 VX^+ ，即V中含有 X^+ 以外的属性。此时关系r的元组在V值上不相等，因此 $V \rightarrow W$ 也在r上成立。

(2) 证明 $X \rightarrow Y$ 在关系r上不成立。

因为 $X \rightarrow Y$ 不能从F使用推理规则推出，根据定理9-4，可知 YX^+ 。在关系r中，可知两元组在X上值相等，在Y上值不相等，因而 $X \rightarrow Y$ 在r上不成立。

综合(1)和(2)可知，只要 $X \rightarrow Y$ 不能用推理规则推出，则F就不逻辑蕴含 $X \rightarrow Y$ ，也就是推理规则是完备的。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.7 最小函数依赖集

定义9-8 若关系模式 $R(U)$ 上的两个函数依赖集 F 和 G , 有 $F^+=G^+$, 则称 F 和 G 是等价的函数依赖集。

定义9-9 设 F 是属性集 U 上的函数依赖集, $X \rightarrow Y$ 是 F 中的函数依赖。

函数依赖中**无关属性**:

- (1) 若 $A \in X$, 且 F 逻辑蕴含 $(F - \{X \rightarrow Y\}) \cup \{(X - A) \rightarrow Y\}$, 则称属性 A 是 $X \rightarrow Y$ 左部的无关属性。
- (2) 若 $A \in X$, 且 $(F - \{X \rightarrow Y\}) \cup \{X \rightarrow (Y - A)\}$ 逻辑蕴含 F , 则称属性 A 是 $X \rightarrow Y$ 右部的无关属性。
- (3) 若 $X \rightarrow Y$ 的左右两边的属性都是无关属性, 则函数依赖 $X \rightarrow Y$ 称为无关函数依赖。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.7 最小函数依赖集

定义9-10 设 F 是属性集 U 上的函数依赖集。若 F_{min} 是 F 的一个最小依赖集，则 F_{min} 应满足下列四个条件：

- (1) $F_{min}^+ = F^+$;
- (2) 每个函数依赖的右边都是单属性;
- (3) F_{min} 中没有冗余的函数依赖（即 F 中不存在这样的函数依赖 $X \rightarrow Y$ ，使得 F 与 $F - \{X \rightarrow Y\}$ 等价）；
- (4) 每个函数依赖的左边没有冗余的属性（即 F 中不存在这样的函数依赖 $X \rightarrow Y$ ， X 有真子集 W 使得 $F - \{X \rightarrow Y\} \cup \{W \rightarrow Y\}$ 与 F 等价）。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.7 最小函数依赖集

算法9-2 计算函数依赖集 F 的最小函数依赖集 F_{min} 。

- (1) 对 F 中的任一函数依赖 $X \rightarrow Y$ ，若 $Y = Y_1, Y_2, \dots, Y_k$ ($k \geq 2$) 多于一个属性，就用分解律，分解为 $X \rightarrow Y_1, X \rightarrow Y_2, \dots, X \rightarrow Y_k$ ，替换 $X \rightarrow Y$ ，得到一个与 F 等价的函数依赖集 F_{min} ， F_{min} 中每个函数依赖的右边均为单属性。
- (2) 去掉 F_{min} 中各函数依赖左部多余的属性。
- (3) 在 F_{min} 中消除冗余的函数依赖。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

9.2.7 DF最小函数依赖集

【案例9-13】 设 F 是关系模式 $R(ABC)$ 的FD集， $F = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$ ，试求 F_{min} 。
解：

(1) 先将 F 中的FD写成右边是单属性形式：

$$F = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$$

其中，多了一个 $A \rightarrow B$ ，可删去。得 $F = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, AB \rightarrow C\}$

(2) F 中 $A \rightarrow C$ 可从 $A \rightarrow B$ 和 $B \rightarrow C$ 推出，因此 $A \rightarrow C$ 是冗余的，可删去。得 $F = \{A \rightarrow B, B \rightarrow C, AB \rightarrow C\}$

(3) F 中 $AB \rightarrow C$ 可从 $A \rightarrow B$ 和 $B \rightarrow C$ 推出，因此 $AB \rightarrow C$ 也可删去。最后得 $F = \{A \rightarrow B, B \rightarrow C\}$ ，即所求的 F_{min} 。

9.2 函数依赖概述

上海市高校精品课程
国家十三五规划项目

讨论思考:

- (1) 设计一个客户购货的关系数据库，给出关系模式，分析存在哪些函数依赖，讨论哪些是完全函数依赖，哪些是部分函数依赖？
- (2) 如何求关系模式的候选键？
- (3) 如何求函数依赖集的最小函数依赖集？



9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.1 模式分解问题

定义9-11 设有关系模式 $R(U)$ ，属性集为 U ， R_1 、...、 R_k 都是 U 的子集，并且有 $R_1 \cup R_2 \cup \dots \cup R_k = U$ 。关系模式 R_1 、...、 R_k 的集合用 ρ 表示， $\rho = \{R_1, \dots, R_k\}$ 。用 ρ 代替 R 的过程称为关系模式的分解。其中， ρ 称为 R 的一个分解，也称为数据库模式。

通常将上述的 R 称为泛关系模式， R 对应的当前值称为泛关系。数据库模式 ρ 对应的当前值称为数据库实例，由数据库模式中的每个关系模式的当前值组成，用 $\sigma = \langle r_1, r_2, \dots, r_k \rangle$ 表示。

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.1 模式分解问题

分解后的模式是否与原有的模式等价，存在3种情况：

- (1) 分解具有无损连接性。
- (2) 分解要保持函数依赖。
- (3) 分解既要保持无损连接，又要保持函数依赖。

泛关系模式

R



数据库模式

$\rho = \{R_1, \dots, R_k\}$

r



$\sigma = \langle r_1, \dots, r_k \rangle$

泛关系

数据库实例 (数据库)

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.2 无损分解

定义9-12 设 R 是一个关系模式， F 是 R 上的一个 FD 集。 R 分解成数据库模式 $\rho=\{R_1, R_2, \dots, R_k\}$ 。若对 R 中满足 F 的每一个关系 r ，有 $r=\pi_{R_1}(r) \bowtie \pi_{R_2}(r) \bowtie \dots \bowtie \pi_{R_k}(r)$

则就称分解 ρ 相对于 F 是**无损连接分解**；否则称为**损失分解**。

定理9-7 设 $\rho=\{R_1, R_2, \dots, R_k\}$ 是关系模式 R 的一个分解， r 是 R 的任一关系， $r_i=\pi_{R_i}(r)$ ($1 \leq i \leq k$)，则有下列性质：

- (1) $r \subseteq \pi_{\rho}(r)$;
- (2) 若 $s=\pi_{\rho}(r)$ ，则 $\pi_{R_i}(s)=r_i$;
- (3) $\pi_{\rho}(\pi_{\rho}(r))=\pi_{\rho}(r)$ ，这个性质称为幂等性。

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.3 无损分解的测试方法

定理9-8 R 的一个分解 $\rho=\{R_1, R_2\}$ 具有无损连接性的充分必要条件是：

$$R_1 \cap R_2 \rightarrow R_1 - R_2 \in F_+$$

或 $R_1 \cap R_2 \rightarrow R_2 - R_1 \in F_+$

当模式 R 分解成两个模式 R_1 和 R_2 时，若两个模式的公共属性（ \emptyset 除外）能够函数决定 R_1 （或 R_2 ）中的其他属性，则此分解具有无损连接性。

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.3 无损分解的测试方法

算法9-3 判别一个分解的无损连接性。

设 $\rho = \{ R_1 \langle U_1, F_1 \rangle, \dots, R_k \langle U_k, F_k \rangle \}$ 是 $R \langle U, F \rangle$ 的一个分解,
 $U = \{ A_1, \dots, A_n \}$, $F = \{ FD_1, FD_2, \dots, FD_p \}$, 且 F 是一极小依赖集,
记为 FD_i 为 $X_i \rightarrow A_{1i}$ 。

(1) 构造一个 k 行 n 列的表格 R_p , 表中每一列对应一个属性 A_j ($1 \leq j \leq n$), 每一行对应一个模式 R_i ($1 \leq i \leq k$)。若 A_j 在 R_i 中, 则在表中的第 i 行第 j 列处填上符号 a_j , 否则填上 b_{ij} 。

(2) 将表格看成模式 R 的一个关系, 根据 F 中的每个函数依赖, 在表中寻找 X 分量上相等的行, 分别对 Y 分量上的每列做修改:

- ①若列中有一个是 a_j , 则这一列上 (X 相同的行) 的元素都改成 a_j ;
- ②若列中没有 a_j , 则这一列上 (X 相同的行) 的元素都改成 b_{ij} (下标 ij 取 i 最小的那个)。
- ③对 F 中所有的函数依赖, 反复地执行上述的修改操作, 一直到表格不能再修改为止 (这个过程称为“追踪”过程)。

(3) 若修改到最后, 表中有一行全为 a , 即 $a_1 a_2 \dots a_n$, 则称 ρ 相对于 F 是无损连接分解, 否则为有损分解。

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.3 无损分解的测试方法

【案例9-13】 设有关系模式 $R(A, B, C, D, E)$, $F = \{AC \rightarrow E, E \rightarrow D, A \rightarrow B, B \rightarrow D\}$, 请判断如下两个分解是否无损分解。

(1) $\rho_1 = \{AC, ED, AB\}$

(2) $\rho_2 = \{ABC, ED, ACE\}$

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.4 保持函数依赖的分解

定义9-13 设 F 是属性集 U 上的FD集， Z 是 U 的子集， F 在 Z 上的投影用 $\Pi_Z(F)$ 表示，定义为 $\Pi_Z(F) = \{X \rightarrow Y \mid X \rightarrow Y \in F^+, \text{ 且 } XY \subseteq Z\}$

定义9-14 设 $\rho = \{R_1, \dots, R_k\}$ 是 R 的一个分解， F 是 R 上的FD集，若有 $\cup \Pi_{R_i}(F) \models F$ ，则称分解 ρ 保持函数依赖集 F 。

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

9.3.4 保持函数依赖的分解

【案例9-14】 关系模式 $R=\{CITY, ST, ZIP\}$ ，其中 $CITY$ 为城市， ST 为街道， ZIP 为邮政编键， $F=\{(CITY, ST) \rightarrow ZIP, ZIP \rightarrow CITY\}$ 。若将 R 分解成 R_1 和 R_2 ， $R_1=\{ST, ZIP\}$ ， $R_2=\{CITY, ZIP\}$ ，检查分解是否具有无损连接和保持函数依赖。

解： 1) 检查无损连接性。

求得： $R_1 \cap R_2 = \{ZIP\}$ ； $R_2 - R_1 = \{CITY\}$ 。

因为 $(ZIP \rightarrow CITY) \in F^+$ 。

所以 分解具有无损连接性

2) 检查分解是否保持函数依赖

求得： $R_1(F) = \Phi$ ； $R_2(F) = \{ZIP \rightarrow CITY\}$ 。

因为 $R_1(F) \cup R_2(F) = \{ZIP \rightarrow CITY\} \neq F^+$ 。

所以 该分解不保持函数依赖。

注意： 一个无损连接分解不一定是保持函数依赖的，一个保持函数依赖的分解也不一定是无损连接的。

9.3 关系模式的分解

上海市高校精品课程
国家十三五规划项目

讨论思考:

- (1) 为什么要进行关系模式的分解?
- (2) 什么是无损分解? 如何测试?
- (3) 进行模式分解时如何能做到既保持无损连接, 又保持函数依赖分解?



9.4 关系模式的范式及规范化

衡量关系模式的好坏的标准是关系模式的范式，范式的种类与数据依赖有着直接的联系，满足不同程度要求的称为不同的范式等级。其中，满足最低要求的关系称为第一范式，简称**1NF**，以此类推，还有第二范式（**2NF**）、第三范式（**3NF**）、**BC**范式（**BCNF**）、第四范式（**4NF**）、第五范式（**5NF**）等多种。不同的范式表示关系模式遵守的不同规则。

“第几范式”用于表示关系的某个级别，称某一关系模式**R**为第几范式，记做 **$R \in xNF$** （ **$x=1,2, \dots, 5, N$** ）。各种范式之间是一种包含关系，具体为：

$$1NF \supset 2NF \supset 3NF \supset BCNF \supset 4NF \supset 5NF$$

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

1. 第一范式 (1NF)

定义9-15 若关系模式R的每个关系r的属性值都是不可分的原子值，则称R是**第一范式1NF** (First normal form) 的模式。

满足1NF的关系称为**规范化的关系**，否则称为非规范化的关系。关系数据库研究的关系都是规范化的关系，1NF是关系模式应具备的最基本的条件。如关系模式：系（系名，研究生人数），假设研究生分为博士生和硕士生，对这两类学生的人数要进行分别统计，在这种情况下，学生人数不再是基本属性，而是一个由分属性“博士”和“硕士”组成的复合属性，因此，该关系模式是非规范化的关系模式，并不满足1NF的范式要求。1NF仍可能出现数据冗余和异常操作问题，还需要去除局部函数依赖。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

1. 第一范式 (1NF)

【案例9-15】 关系模式R存放的是各系的研究生人数，如表9-9所示。请判断R是否符合1NF？若不符合，能否规范化至1NF？

表9-9 关系模式R

系名	研究生人数	
	博士	硕士
电子系	25	60
计算机系	31	82
通信系	20	57

9.4 关系模式的范式及规范化

表9-9的关系模式R中，“研究生人数”是由两个其他属性“博士”和“硕士”组成的一个复合属性，由于它不是基本属性，可判断R并不符合1NF。对R的规范化有两种方法：

(1) 方法1是将“研究生人数”这个复合属性分解成2个属性“博士人数”和“硕士人数”：

系名	研究生人数
信息系	60
计算机系	82
电子系	57

系名	本科生人数
信息系	302
计算机系	421
电子系	286

(2) 方法2是将关系模式R分解成2个不含复合属性的关系模式：

系名	本科生人数	研究生人数
信息系	302	60
计算机系	421	82
电子系	286	53

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

2. 第二范式 (2NF)

定义9-16 对于FD $W \rightarrow A$ ，若存在 XW 有 $X \rightarrow A$ 成立，则称 $W \rightarrow A$ 是局部依赖（ A 局部依赖于 W ）；否则称 $W \rightarrow A$ 是完全依赖。完全依赖也称为“左部不可约依赖”。

定义9-17 若 A 是关系模式 R 中候选键属性，则称 A 是 R 的主属性；否则称 A 是 R 的非主属性。

定义9-18 若关系模式 R 是1NF，且每个非主属性完全函数依赖于候选键，则称 R 是第二范式（2NF）的模式。若数据库模式中每个关系模式都是2NF，则称数据库模式为2NF的数据库模式。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

将一个1NF的关系模式变为2NF的方法是，通过模式分解，使任一非主属性都完全函数依赖于它的任一候选键，目的是消除非主属性对键的部分函数依赖。下面分析一个不是2NF的案例。

【案例9-6】关系模式 $R(C\#, TYPE, ADDR, P\#, Credit)$

其中， R 的属性分别表示客户编号、购物类别、地址、购货编号、积分等含义，且各购物类别的客户放在同一类。

候选键为 $(C\#, P\#)$ 。函数依赖为：

$(C\#, P\#) \rightarrow Credit$

$C\# \rightarrow TYPE, (C\#, P\#) \rightarrow TYPE$

$C\# \rightarrow ADDR, (C\#, P\#) \rightarrow ADDR,$

$TYPE \rightarrow ADDR$ （因为选购同一类商品的客户只存放一个地方）

可知此关系模式存在部分函数依赖，不符合2NF。

9.4 关系模式的范式及规范化

2. 第二范式 (2NF)

分析上述案例，可以看出主要问题在于关系模式中的非主属性有两种：一种如***Credit***对键是完全函数依赖；另一种如**TYPE**和**ADDR**，对键不是完全函数依赖。解决的办法是将关系模式**R**分解为两个关系模式：

$R_1 (C\#, P\#, Credit)$

$R_2 (C\#, TYPE, ADDR)$

关系模式 **R_1** 的键为 **$(C\#, P\#)$** ，关系模式 **R_2** 的键为 **$C\#$** ，因此，就使得非主属性对键都是完全依赖。



9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

【案例9-17】 设关系模式R (C#, P#, AMOUNT, ENAME, ADDR) 的属性分别表示客户编号、购买商品的编号、数量、生产企业的名称和企业地址等含义。(C#, P#) 是R的候选键。请对其进行模式分解。

解：

在R上有两个函数依赖： $(C\#, P\#) \rightarrow (ENAME, ADDR)$ 和 $P\# \rightarrow (ENAME, ADDR)$ ，所以，前一个FD是局部依赖，R不是2NF模式。此时R的关系就会出现冗余和异常现象。例如某一类商品有200个客户选购，则在关系中就会存在200个元组，因而企业的客户名称和地址就会重复200次。

若将R分解成R1 (P#, ENAME, ADDR) 和R2 (C#, P#, AMOUNT) 后，局部依赖 $(C\#, P\#) \rightarrow (ENAME, ADDR)$ 就消失了。R1和R2都是2NF模式。

9.4 关系模式的范式及规范化

2. 第二范式 (2NF)

算法9-4 分解成2NF模式集的算法

设关系模式 $R(U)$ ，主键是 W ， R 上还存在FD $X \rightarrow Z$ ，并且 Z 是非主属性和 $X \not\rightarrow W$ ，则 $W \rightarrow Z$ 就是一个局部依赖。此时应将 R 分解成两个模式：

$R_1(XZ)$ ，主键是 X ；

$R_2(Y)$ ，其中 $Y=U-Z$ ，主键仍是 W ，外键是 X （参数， R_1 ）。

利用外键和主键的联接可以从 R_1 和 R_2 重新得到 R 。

若 R_1 和 R_2 还不是2NF，则重复上述过程，一直到数据库模式中每一个关系模式都成为2NF为止。

9.4 关系模式的范式及规范化

【案例9-18】请判断案例9-2的关系模式：读者（读者号，姓名，单位，地址，图书号，借书时间，还书时间）是否属于2NF？若不是，请将其规范化为2NF。

解：

该关系模式中存在主属性对主键的部分函数依赖，如：

$(\text{读者号}, \text{图书号}) \xrightarrow{p} \text{姓名}$

所以R不是2NF的。将R转换为2NF关系模式的过程如下：

(1) 将主键属性集合的3个子集作为主键构建关系模式 R1 (读者号, ...)

R2 (图书号, ...)

R3 (读者号, 图书号, ...)

(2) 将非主属性依次放入到相应的关系模式

R1 (读者号, 姓名, 单位, 地址)

R2 (图书号)

R3 (读者号, 图书号, 借书时间, 还书时间)

去掉仅由主属性构成的关系模式R2，剩下的R1、R3即为所求。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

3. 第三范式 (3NF)

定义9-19 若 $X \rightarrow Y$, $Y \rightarrow A$, 且 $Y \not\rightarrow X$ 和 $A \not\rightarrow Y$, 则称 $X \rightarrow A$ 是传递依赖 (A传递依赖于X)。

3NF的目的是消除非主属性对键的传递函数依赖。

定义9-20 若关系模式R是1NF, 且每个非主属性都不传递依赖于R的候选键, 则称R是第三范式 (3NF) 的模式。若数据库模式中每个关系模式都是3NF, 则称其为3NF的数据库模式。



9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

【案例9-19】在上述案例中，R2是2NF模式，而且也已是3NF模式。但R1（P#，ENAME，ADDR）是2NF模式，却不一定是3NF模式。若R1中存在函数依赖 $P\# \rightarrow ENAME$ 和 $ENAME \rightarrow ADDR$ ，则 $P\# \rightarrow ADDR$ 就是一个传递依赖，即R1不是3NF模式。此时R1的关系中也会出现冗余和异常操作。例如一个企业生产五种产品，则关系中就会出现五个元组，企业的地址就会重复五次。若将R₁分解成R₁₁(ENAME，ADDR)和R₁₂(P#，ENAME)后， $C\# \rightarrow ADDR$ 就不会出现在R₁₁和R₁₂中。这样R₁₁和R₂₂都是3NF模式。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

3. 第三范式 (3NF)

算法9-5 分解成3NF模式集的算法。

设关系模式 $R(U)$ ，主键是 W ， R 上还存在FD $X \rightarrow Z$ 。并且 Z 是非主属性， ZX ， X 不是候选键，这样 $W \rightarrow Z$ 就是一个传递依赖。此时应将 R 分解成两个模式：

$R1(XZ)$

$R2(Y)$

$R1(XZ)$ 主键是 X ；

$R2(Y)$ 其中 $Y=U-Z$ ，主键仍是 W ，外键是 X （参数， $R1$ ）。

利用外键和主键相匹配机制， $R1$ 和 $R2$ 通过联接可以重新得到 R 。若 $R1$ 和 $R2$ 还不是3NF，则重复上述过程，一直到数据库模式中每一个关系模式都是3NF为止。

9.4 关系模式的范式及规范化

【案例9-20】请判断案例9-20的关系模式：读者（读者号，姓名，单位，地址），借阅（读者号，图书号，借书时间，还书时间）是否属于3NF？若不是，请将其规范化为3NF。

解：

对于读者这个关系模式，有函数依赖： $\text{读者号} \rightarrow \text{单位}$ ，同时，由题意“每个单位只有一个地址”，有函数依赖： $\text{单位} \rightarrow \text{地址}$ ，可知存在传递函数依赖，故该关系模式不符合3NF。

对于借阅这个关系模式，两个非主属性之间并没有函数依赖关系，故不存在传递函数依赖，所以该关系模式符合3NF。

对读者关系的分解过程为：

(1) 删除非主属性“单位”的依赖因子“地址”，得到

读者1（读者号，姓名，单位）

(2) 新建关系模式，放入从步骤(1)删除掉的依赖因子“地址”及其决定因子“单位”，得到

单位（单位，地址）

设置主键和外键，得到

读者1（读者号，姓名，单位），单位为引用单位关系模式的外键

单位（单位，地址）

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

3. 第三范式 (3NF)

定理9-9 若R是3NF模式，则R也是2NF模式。证明（略）

局部依赖和传递依赖是模式产生冗余和异常的两个重要原因。由于3NF模式中不存在非主属性对候选键的局部依赖和传递依赖，因此消除了很大一部分存储异常，具有较好的性能。而对于非3NF的1NF和2NF，甚至非1NF的关系模式，由于其性能上的弱点，一般不宜作为数据库模式，通常需要将它们变换成更高级的范式，这种变换过程，称为“关系的规范化”。

定理9-10 设关系模式R，当R上每一个函数依赖 $X \rightarrow A$ 满足下列三个条件之一时：

1. $A \subseteq X$ （即 $X \rightarrow A$ 是一个平凡的函数依赖）；
2. X是R的超键；
3. A是主属性。

关系模式R就是3NF模式。

9.4 关系模式的范式及规范化

3. 第三范式 (3NF)

算法9-6 将一个关系模式分解为**3NF**，使它既具有无损连接性又具有保持函数依赖性。

- (1) 根据算法9-5求出保持函数依赖的分解： $\rho = \{R_1, R_2, \dots, R_k\}$ 。
- (2) 判定 ρ 是否具有无损连接性，若是，转(4)。
- (3) 令 $\rho = \rho \cup \{X\} = \{R_1, R_2, \dots, R_k, X\}$ ，其中 X 是 R 的候选键。
- (4) 输出 ρ 。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

3. 第三范式 (3NF)

【案例9-21】 将SD(C#, CNAME, SAge, Dept, Manager)规范到3NF。

(1) 根据算法9-6求出保持函数依赖的分解: $\rho = \{S(C\#, CNAME, SAge, Dept), D(Dept, Manager)\}$ 。

(2) 判定 ρ 是否具有无损连接性

SD分解为 $\rho = \{S(C\#, CNAME, SAge, Dept), D(Dept, Manager)\}$ 时, S、D都属于3NF, 且既具有无损连接性又具有保持函数依赖性。

9.4 关系模式的范式及规范化

4. BCNF (Boyce–Codd NF)

定义9-21 若关系模式R是1NF，且每个属性都不传递依赖于R的候选键，则称R是BCNF的模式。若数据库模式中每个关系模式都是BCNF，则称为BCNF的数据库模式。

讨论BCNF的目的是消除主属性对键的部分函数依赖和传递依赖，具有如下性质：

- (1) 若 $R \in \text{BCNF}$ ，则R也是3NF。
- (2) 若 $R \in 3\text{NF}$ ，则R不一定是BCNF。

属于3NF的关系模式有的属于BCNF，有的不属于BCNF。

9.4 关系模式的范式及规范化

4. BCNF (Boyce–Codd NF)

BCNF和3NF的**区别**:

- (1) BCNF不仅强调其他属性对键的完全的直接的依赖,而且强调主属性对键的完全的直接的依赖,它包括3NF,即 $R \in \text{BCNF}$, 则 R 一定属于3NF。
- (2) 3NF只强调非主属性对键的完全直接依赖,这样就可能出现主属性对键的部分依赖和传递依赖。



9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

【案例9-22】先考察关系模式C (P#, PNAME, PP#)，其中属性分别表示商品编号、商品名称和先行商品编号键，它只有一个键P#，这里没有任何属性对P#部分或传递依赖，所以C属于3NF。同时C中P#是唯一的决定因素，所以C属于BCNF。

【案例9-23】关系模式CFM (C, F, M) 中，C表示客户，F表示企业，M表示商品。每一个企业只生产一种商品。生产每种商品有若干企业，某一个客户选购某种商品，则对于一个固定的企业，具有的函数依赖为：

$$(C, M) \rightarrow F \quad (C, F) \rightarrow M \quad F \rightarrow M$$

其中，(C, M)和(C, F)都是候选键。

由于没有任何非主属性对键传递依赖或部分依赖，所以CFM属于3NF。但CFM不是BCNF关系，因为F是决定因素，但F不包含键。

9.4 关系模式的范式及规范化

4. BCNF (Boyce–Codd NF)

对于不是BCNF的关系模式，仍然存在不合适的地方。非BCNF的关系模式也可以通过分解成为BCNF。例如CFM可以分解为CF(C, F)与FM(F, M)，都是BCNF。

【案例9-24】 设有关系模式SNC (SNo, SN, CNo, Score)， $Sno \leftrightarrow SN$ 。

存在着主属性对键的部分函数依赖：SN部分依赖于SNo和CNo，SNo部分依赖于SN和CNo，所以SNC不是BCNF。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

4. BCNF (Boyce–Codd NF)

算法9-7 无损分解成BCNF模式集。

- (1) 令 $\rho = \{R\}$ 。
- (2) 若 ρ 中所有模式都是BCNF，则转 (4)。
- (3) 若 ρ 中有一个关系模式 S 不是BCNF，则 S 中必能找到一个函数依赖 $X \rightarrow A$ 且 X 不是 S 的候选键，且 A 不属于 X ，设 $S_1 = XA$ ， $S_2 = S - A$ ，用分解 $\{S_1, S_2\}$ 代替 S ，转 (2)。
- (4) 分解结束，输出 ρ 。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

4. BCNF (Boyce–Codd NF)

【案例9-25】 将SNC(SNo, SN, CNo, Score)规范到BCNF。

解：

候选键：(SNo, CNo) 和 (SN, CNo)

函数依赖：

$F = \{SNo \rightarrow SN, SN \rightarrow SNo, (SNo, CNo) \rightarrow Score, (SN, CNo) \rightarrow Score\}$

(1) 令 $\rho = \{SNC(SNo, SN, CNo, Score)\}$ 。

(2) 经过前面分析可知， ρ 中关系模式不属于BCNF。

(3) 用分解 $\{S1(SNo, SN), S2(SNo, CNo, Score)\}$ 代替SNC。

(4) 分解结果为：S1(SNo, SN) 描述客户实体；S2(SNo, CNo, Score) 描述客户与商品的联系。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

4. BCNF (Boyce–Codd NF)

【案例9-26】 将关系模式TCS(T, C, S) 规范到BCNF, 其中候选键为 (S, C) 和 (S, T), 函数依赖是 : $F = \{ (S, C) \rightarrow T, (S, T) \rightarrow C, T \rightarrow C \}$

解:

分解 $\{TC(T, C), ST(S, T)\}$ 代替TCS,

消除函数依赖 $(S, T) \rightarrow C$, $ST \in BCNF$, $TC \in BCNF$

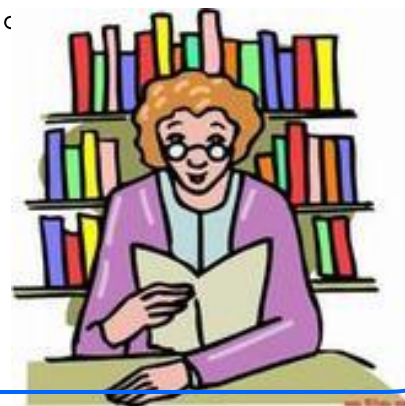
9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

5. 第四范式 (4NF)

定义9-22 设有一关系模式 $R(U)$ ， U 是其属性全集， X 、 Y 是 U 的子集， D 是 R 上的数据依赖集。若对于任一多值依赖 $X \twoheadrightarrow Y$ ，此多值依赖是平凡的，或 X 包含了 R 的一个候选键，则称 R 是**第四范式的关系模式**，记为 $R \in 4NF$ 。

注意：介绍BCNF目的是消除非平凡且非DF的多值依赖；BCNF的关系模式不一定是4NF；4NF的关系模式必定是BCNF的关系模式；4NF是BCNF的推广。



9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

5. 第四范式 (4NF)

算法9-8 第四范式 (4NF) 的分解:

- (1) 令 $\rho=\{R\}$ 。
- (2) 若 ρ 中所有模式 R_i 都是4NF, 则转 (4)。
- (3) 若 ρ 中有一个关系模式 S 不是4NF, 则 S 中必能找到一个多值依赖 $X \twoheadrightarrow Y$ 且 X 不包含 S 的候选键, $Y-X \neq \emptyset$, $XY \neq S$, 令 $Z=Y-X$, 设 $S_1=XZ$, $S_2=S-Z$, 用分解 $\{S_1, S_2\}$ 代替 S , 由于 $S_1 \cap S_2 = X$, $S_1 - S_2 = Z$, 所以有 $(S_1 \cap S_2) \twoheadrightarrow (S_1 - S_2)$, 分解具有无损连接性, 转 (2)。
- (4) 分解结束, 输出 ρ 。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

9.4.2 关系模式的规范化

在关系数据库中，低级范式的关系模式通常存在数据冗余和操作异常现象，为此需要寻求解决这个问题方法，从而引出了关系模式的规范化问题。一个低一级范式的关系模式，通过模式分解转化为若干个高一级范式的关系模式的集合，这种分解过程称为**关系模式的规范化**。

1. 规范化的目的与原则

(1) 关系模式规范化的目的

关系模式规范化的**目的**是使其结构合理，消除数据中的存储异常，使数据冗余尽量小，在操作过程中便于插入、删除和更新，并保持操作数据的正确性和完整性。

(2) 关系模式规范化的原则

关系模式规范化的**原则**是：遵从概念单一化“一事一地”的原则，即一个关系模式描述一个实体或实体间的一种联系。规范的实质就是概念单一化。

2. 规范化过程

常用的关系模式规范化过程，主要包括：

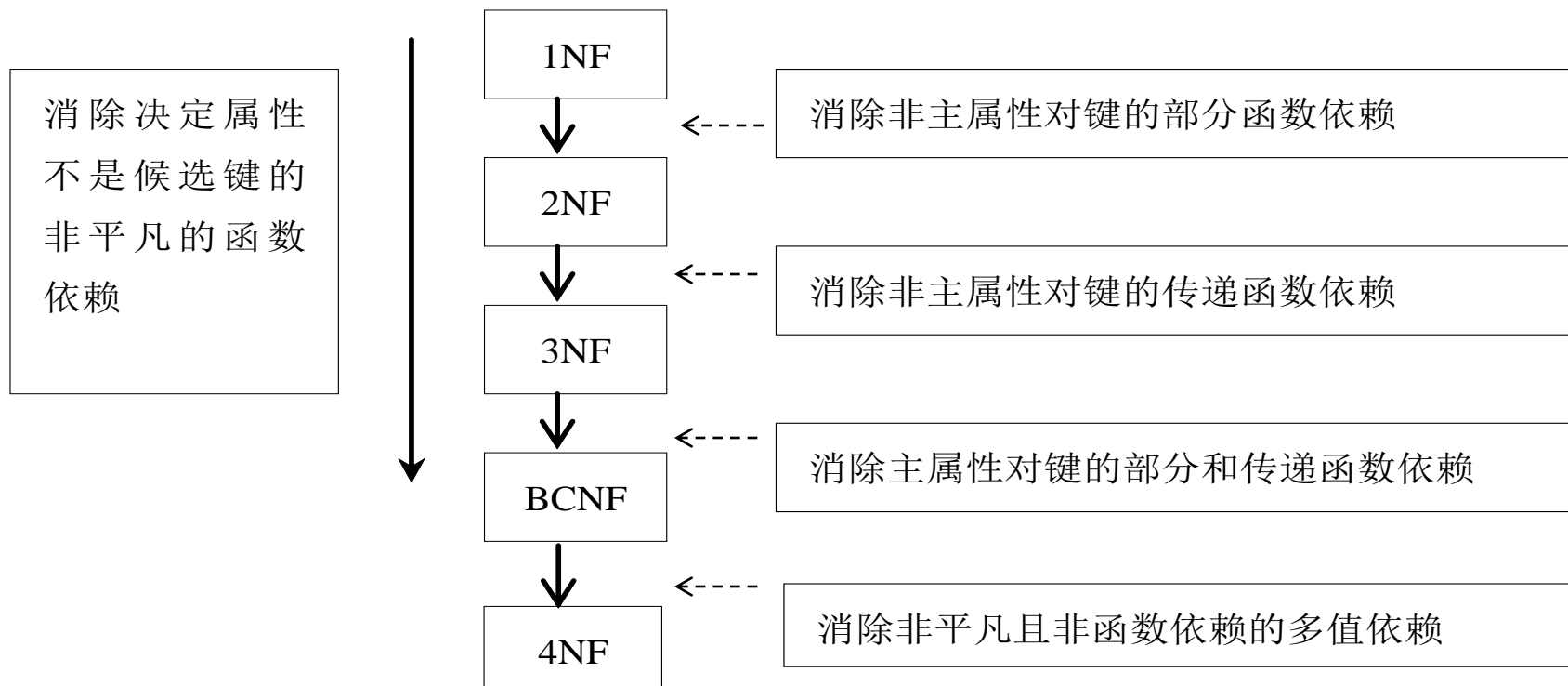
- (1) 对1NF关系进行分解，消除原关系中非主属性对键的部分函数依赖，将1NF关系转换为多个2NF。
- (2) 对2NF关系进行分解，消除原关系中非主属性对键的传递函数依赖，产生一组3NF。

在实际应用中，规范化的过程就是一个不断消除属性依赖关系中某些问题的过程，是从第一范式到第四范式的逐步递进规范的过程，实际上是通过把范式程度低的关系模式分解为若干个范式程度高的关系模式的过程。

9.4 关系模式的范式及规范化

上海市高校精品课程
国家十三五规划项目

2. 规范化过程



3. 规范化要求

数据库的规范化问题对于数据库设计很重要，直接关系到数据处理的正确性和准确性。在数据库设计中构建关系模式需要规范化原理解决数据冗余和操作异常问题。规范化的基本思想是逐步消除数据依赖中不合适的依赖关系，通过模式分解的方法使关系模式逐步消除数据冗余和操作异常。

在规范化过程中，分解后的关系模式集合应当与原关系模式“等价”，即经过自然联接可以恢复原关系而不丢失信息，并保持属性间合理的联系。

保证分解后的关系模式与原关系模式是等价的，等价的三种标准：

- 1、分解要具有无损连接性；
- 2、分解要具有函数依赖保持性；
- 3、分解既要具有无损连接性，又要具有函数依赖保持性。

9.4 关系模式的范式及规范化

讨论思考:

- (1) 为什么要进行关系模式的规范化?
- (2) 如何对关系模式进行规范化?
- (3) 关系模式的规范化的目标是追求满足更高的范式吗?



本章小结

上海市高校精品课程
国家十三五规划项目

本章重点介绍了数据库关系模式规范化设计问题。关系模式设计的正确性和完整性直接影响数据冗余度、数据一致性等问题。设计好的数据库模式必须要有模式规范化理论为基础。

数据库的不良关系模式设计会导致数据冗余，并容易引起增、删、改、查的操作异常。通过模式分解的方法，将原关系模式化成若干关系模式可以消除冗余。关系模式的规范化过程实际上是一个“分解”过程。分解是解决数据冗余的主要方法，也是规范化的一条原则：“关系模式有冗余问题就应分解”。

函数依赖 $X \rightarrow Y$ 是数据之间最基本的一种联系，在关系中有两个元组，若 X 值相等则要求 Y 值也相等。函数依赖有一个完备的推理规则集。

关系模式在分解时应保持“等价”，有数据等价和语义等价两种，分别用无损分解和保持依赖两个特征进行衡量。前者能保持泛关系在投影联接后仍能恢复，而后者能保证数据在投影或联接中其语义不会发生变化，即不会违反函数依赖的语义。但无损分解与保持依赖两者之间没有必然的联系。

范式是衡量模式优劣的标准，表达了模式中数据依赖之间应满足的关系。范式的级别越高，其数据冗余和操作异常现象就越少。分解成BCNF模式集算法能保持无损分解，但不一定能保持函数依赖集。而分解成3NF模式集的算法既能保持无损分解，又能保持函数依赖集。

国家“十三五”重点出版规划项目
上海高校优秀教材奖获得者主编

上海市高校精品课程
特色教材

诚挚谢意！



数据库原理及应用

——基于SQL Server 2016