



数据库系统概论

An Introduction to Database System

第六章 关系数据理论

中国人民大学信息学院

第六章 关系数据理论



6.1 问题的提出

6.2 规范化

6.3 数据依赖的公理系统

***6.4 模式的分解**

6.5 小结

6.2 规范化



规范化理论正是用来改造关系模式，通过分解关系模式来消除其中不合适的数据依赖，以解决插入异常、删除异常、更新异常和数据冗余问题。

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

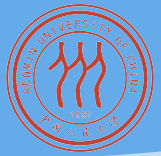
6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.1 函数依赖



- ❖ 函数依赖
- ❖ 平凡函数依赖与非平凡函数依赖
- ❖ 完全函数依赖与部分函数依赖
- ❖ 传递函数依赖

一、函数依赖



定义 6.1 设 $R(U)$ 是一个属性集 U 上的关系模式， X 和 Y 是 U 的子集。

若对于 $R(U)$ 的任意一个可能的关系 r ， r 中不可能存在两个元组在 X 上的属性值相等，而在 Y 上的属性值不等，则称 “ X 函数确定 Y ” 或 “ Y 函数依赖于 X ”，记作 $X \rightarrow Y$ 。

说明



1. 所有关系实例均要满足
2. 语义范畴的概念
3. 数据库设计者可以对现实世界作强制的规定

二、平凡函数依赖与非平凡函数依赖



在关系模式 $R(U)$ 中，对于 U 的子集 X 和 Y ，

如果 $X \rightarrow Y$ ，但 $Y \not\subseteq X$ ，则称 $X \rightarrow Y$ 是非平凡的函数依赖

若 $X \rightarrow Y$ ，但 $Y \subseteq X$ ，则称 $X \rightarrow Y$ 是平凡的函数依赖

❖ 例：在关系 $SC(Sno, Cno, Grade)$ 中，

非平凡函数依赖： $(Sno, Cno) \rightarrow Grade$

平凡函数依赖： $(Sno, Cno) \rightarrow Sno$

$(Sno, Cno) \rightarrow Cno$

平凡函数依赖与非平凡函数依赖（续）



- 若 $X \rightarrow Y$ ，则 X 称为这个函数依赖的决定属性组，也称为决定因素（Determinant）。
- 若 $X \rightarrow Y$ ， $Y \rightarrow X$ ，则记作 $X \leftrightarrow Y$ 。
- 若 Y 不函数依赖于 X ，则记作 $\nexists X \rightarrow Y$ 。

三、完全函数依赖与部分函数依赖



定义 6.2 在 $R(U)$ 中，如果 $X \rightarrow Y$ ，并且对于 X 的任何一个真子集 X' ，都有 $X' \not\rightarrow Y$ ，则称 Y 对 X **完全函数依赖**，记作 $X \xrightarrow{F} Y$ 。

若 $X \rightarrow Y$ ，但 Y ~~不~~完全函数依赖于 X ，则称 Y 对 X **部分函数依赖**，记作 $X \xrightarrow{P} Y$ 。

完全函数依赖与部分函数依赖（续）



[例 1] 中 $(Sno, Cno) \xrightarrow{F} Grade$ 是完全函数依赖，
 $(Sno, Cno) \xrightarrow{P} Sdept$ 是部分函数依赖
因为 $Sno \rightarrow Sdept$ 成立，且 Sno 是
 (Sno, Cno) 的真子集

四、传递函数依赖



定义 6.3 在 $R(U)$ 中，如果 $X \rightarrow Y$ ， $\setminus (Y \subseteq X)$ ， $Y \rightarrow X$ $Y \rightarrow Z$ ，则称 Z 对 X **传递函数依赖**。
记为： $X \xrightarrow{\text{传递}} Z$

注：如果 $Y \rightarrow X$ ，即 $X \longleftrightarrow Y$ ，则 **Z 直接依赖于 X** 。

例：在关系 $\text{Std}(\text{Sno}, \text{Sdept}, \text{Mname})$ 中，有：
 $\text{Sno} \rightarrow \text{Sdept}$ ， $\text{Sdept} \rightarrow \text{Mname}$
 Mname 传递函数依赖于 Sno

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.2 码



定义 6.4 设 K 为 $R\langle U, F \rangle$ 中的属性或属性组合。若 $K \xrightarrow{F} U$ ，则 K 称为 R 的**候选码**（Candidate Key）。

若候选码多于一个，则选定其中的一个做为**主码**（Primary Key）。

码（续）



❖ 主属性与非主属性

- 包含在任何一个候选码中的属性，称为主属性（ Prime attribute ）
- 不包含在任何码中的属性称为非主属性（ Nonprime attribute ）或非码属性（ Non-key attribute ）

❖ 全码

- 整个属性组是码，称为全码（ All-key ）

码（续）



[例 2]

关系模式 $S(\underline{Sno}, Sdept, Sage)$ ，单个属性 Sno 是码，
 $SC(\underline{Sno}, \underline{Cno}, Grade)$ 中， (Sno, Cno) 是码

[例 3]

关系模式 $R(P, W, A)$

P ：演奏者 W ：作品 A ：听众

一个演奏者可以演奏多个作品

某一作品可被多个演奏者演奏

听众可以欣赏不同演奏者的不同作品

码为 (P, W, A) ，即 All-Key

外部码



定义 6.5 关系模式 R 中属性或属性组 X 并非 R 的码，但 X 是另一个关系模式的码，则称 X 是 R 的**外部码**（**Foreign key**）也称外码

- ❖ 如在 SC （ Sno ， Cno ， $Grade$ ）中， Sno 不是码，但 Sno 是关系模式 S （ Sno ， $Sdept$ ， $Sage$ ）的码，则 Sno 是关系模式 SC 的外部码
- ❖ 主码与外部码一起提供了表示关系间联系的手段

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

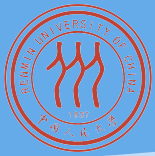
6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.3 范式



- ❖ 范式是符合某一种级别的关系模式的集合
- ❖ 关系数据库中的关系必须满足一定的要求。满足不同程度要求的为不同范式
- ❖ 范式的种类：

第一范式 (1NF)

第二范式 (2NF)

第三范式 (3NF)

BC 范式 (BCNF)

第四范式 (4NF)

第五范式 (5NF)

6.2.3 范式



❖ 各种范式之间存在联系：

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

- ❖ 某一关系模式 R 为第 n 范式，可简记为 $R \in nNF$ 。
- ❖ 一个低一级范式的关系模式，通过模式分解可以转换为若干个高一级范式的关系模式的集合，这种过程就叫规范化。

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.4 2NF



❖ 1NF 的定义

如果一个关系模式 R 的所有属性都是不可分的基本数据项，则 $R \in 1NF$

❖ 第一范式是对关系模式的最起码的要求。不满足第一范式的数据库模式不能称为关系数据库

❖ 但是满足第一范式的关系模式并不一定是一个好的关系模式

2NF (续)



[例 4] 关系模式 S-L-C(Sno, Sdept, Sloc, Cno, Grade)

Sloc 为学生住处，假设每个系的学生住在同一个地方

❖ 函数依赖包括：

$(Sno, Cno) \xrightarrow{E} Grade$

$Sno \rightarrow Sdept$

$(Sno, Cno) \xrightarrow{P} Sdept$

$Sno \rightarrow Sloc$

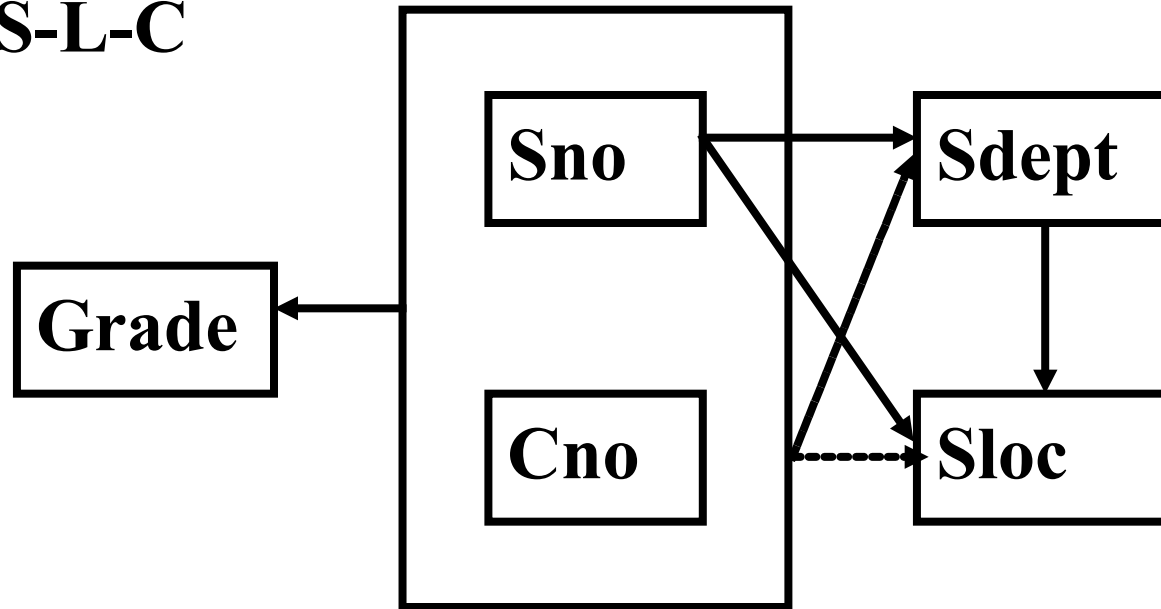
$(Sno, Cno) \xrightarrow{P} Sloc$

$Sdept \rightarrow Sloc$

2NF (续)



S-L-C



- ❖ S-L-C 的码为 (Sno, Cno)
- ❖ S-L-C 满足第一范式。
- ❖ 非主属性 Sdept 和 Sloc 部分函数依赖于码 (Sno, Cno)

S-L-C 不是一个好的关系模式（续）



- (1) 插入异常
- (2) 删除异常
- (3) 数据冗余度大
- (4) 修改复杂

S-L-C 不是一个好的关系模式（续）



❖ 原因

Sdept、Sloc 部分函数依赖于码。

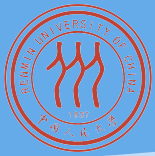
❖ 解决方法

S-L-C 分解为两个关系模式，以消除这些部分函数依赖

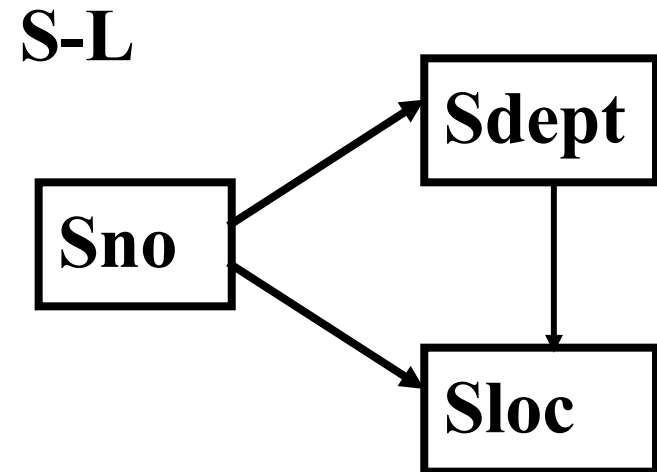
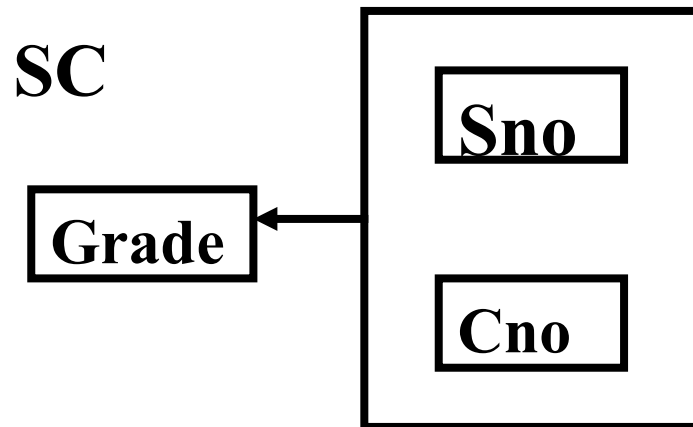
→→SC (Sno, Cno, Grade)

S-L (Sno, Sdept, Sloc)

2NF (续)



函数依赖图：



- ❖ 关系模式 SC 的码为 (Sno , Cno)
- ❖ 关系模式 S-L 的码为 Sno
- ❖ 这样非主属性对码都是完全函数依赖

2NF (续)



❖ 2NF 的定义

定义 6.6 若 $R \in 1NF$ ，且每一个非主属性完全函数依赖于码，则 $R \in 2NF$ 。

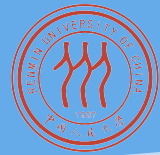
例：S-L-C(Sno, Sdept, Sloc, Cno, Grade) $\in 1NF$

S-L-C(Sno, Sdept, Sloc, Cno, Grade) $\notin 2NF$

SC (Sno , Cno , Grade) $\in 2NF$

S-L (Sno , Sdept , Sloc) $\in 2NF$

2NF (续)



- ❖ 采用投影分解法将一个 1NF 的关系分解为多个 2NF 的关系，可以在一定程度上减轻原 1NF 关系中存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。
- ❖ 将一个 1NF 关系分解为多个 2NF 的关系，并不能完全消除关系模式中的各种异常情况和数据冗余。

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

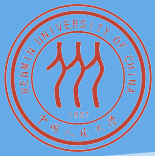
6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.5 3NF



❖ 3NF 的定义

定义 6.7 关系模式 $R<U, F>$ 中若不存在这样的码 X 、属性组 Y 及非主属性 Z ($Z \subseteq Y$)，使得 $X \rightarrow Y$ ， $Y \rightarrow Z$ 成立，

$Y \rightarrow X$ ，则称 $R<U, F> \in 3NF$ 。

- 若 $R \in 3NF$ ，则每一个非主属性既不部分依赖于码也不传递依赖于码。

3NF (续)



例：2NF 关系模式 S-L(Sno, Sdept, Sloc) 中

- 函数依赖：

$Sno \rightarrow Sdept$

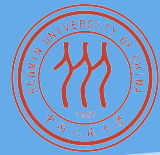
$Sdept \twoheadrightarrow Sno$

$Sdept \rightarrow Sloc$

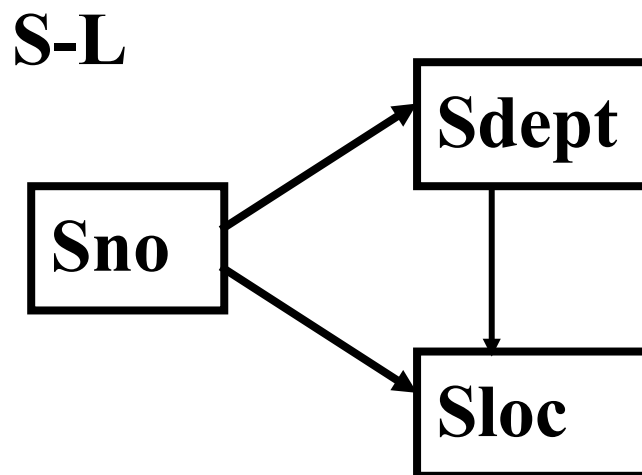
可得：

$Sno \xrightarrow{\text{传递}} Sloc$ ，即 S-L 中存在非主属性对码的传递函数依赖， $S-L \notin 3NF$

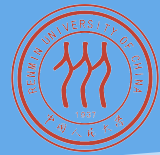
3NF (续)



函数依赖图：



3NF (续)



❖ 解决方法

采用投影分解法，把 S-L 分解为两个关系模式，以消除传递函数依赖：

$\rightarrow\rightarrow S-D \text{ (Sno , Sdept)}$

$D-L \text{ (Sdept , Sloc)}$

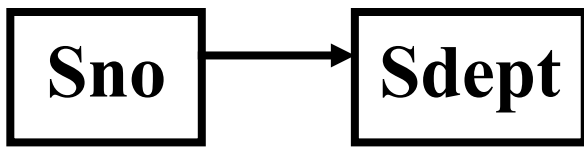
S-D 的码为 Sno， D-L 的码为 Sdept。

- 分解后的关系模式 S-D 与 D-L 中不再存在传递依赖

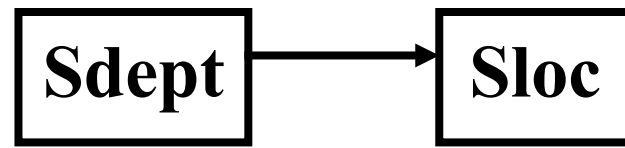
3NF (续)



S-D 的码为 Sno , D-L 的码为 Sdept



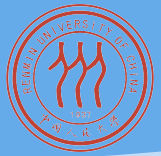
S-D



D-L

- ❖ $S-L(Sno, Sdept, Sloc) \in 2NF$
 $S-L(Sno, Sdept, Sloc) \notin 3NF$
 $S-D(Sno, Sdept) \in 3NF$
 $D-L(Sdept, Sloc) \in 3NF$

3NF (续)



- ❖ 采用投影分解法将一个 2NF 的关系分解为多个 3NF 的关系，可以在一定程度上解决原 2NF 关系中存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。
- ❖ 将一个 2NF 关系分解为多个 3NF 的关系后，仍然不能完全消除关系模式中的各种异常情况和数据冗余。

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

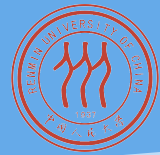
6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.6 BC 范式 (BCNF)



❖ **定义 6.8** 关系模式 $R<U, F> \in 1NF$ ，若 $\nexists X \rightarrow Y$ 且 $Y \subseteq X$ 时 X 必含有码，则 $R<U, F> \in BCNF$ 。

❖ 等价于：每一个决定属性因素都包含码

BCNF (续)



❖ 若 $R \in \text{BCNF}$

- 所有非主属性对每一个码都是完全函数依赖
- 所有的主属性对每一个不包含它的码，也是完全函数依赖
- 没有任何属性完全函数依赖于非码的任何一组属性

❖ $R \in \text{BCNF} \xrightleftharpoons[\text{不必要}]{\text{充分}} R \in 3\text{NF}$

BCNF (续)



[例 5] 关系模式 C (Cno , Cname , Peno)

- $C \in 3NF$
- $C \in BCNF$

[例 6] 关系模式 S (Sno , Sname , Sdept , Sage)

- 假定 S 有两个码 Sno , Sname
- $S \in 3NF$ 。
- $S \in BCNF$

BCNF (续)



[例 7] 关系模式 SJP (S , J , P)

- 函数依赖: $(S, J) \rightarrow P$; $(J, P) \rightarrow S$
- (S, J) 与 (J, P) 都可以作为候选码, 属性相交
- $SJP \in 3NF$,
- $SJP \in BCNF$

BCNF (续)



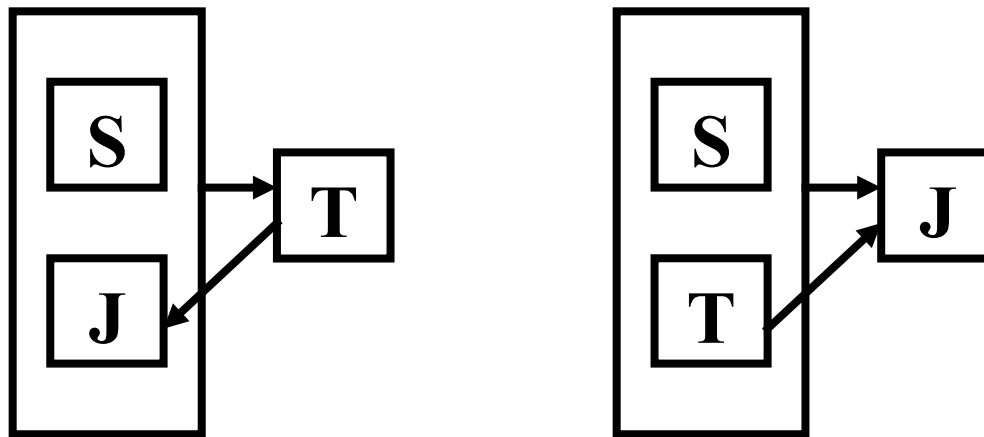
[例 8] 在关系模式 STJ (S , T , J) 中, S 表示学生, T 表示教师, J 表示课程。

- 函数依赖:

$(S, J) \rightarrow T, (S, T) \rightarrow J, T \rightarrow J$

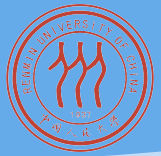
- (S, J) 和 (S, T) 都是候选码

BCNF (续)



STJ 中的函数依赖

BCNF (续)



❖ $STJ \in 3NF$

- 没有任何非主属性对码传递依赖或部分依赖

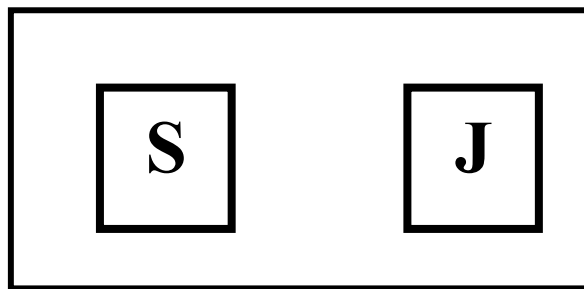
❖ $STJ \notin BCNF$

- T 是决定因素，T 不包含码

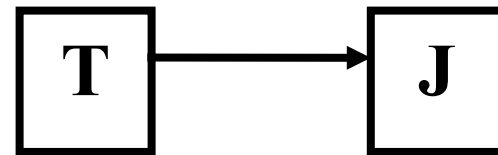
BCNF (续)



- ❖ 解决方法：将 STJ 分解为二个关系模式：
 $ST(S, T) \in BCNF$, $TJ(T, J) \in BCNF$



ST



TJ

没有任何属性对码的部分函数依赖和传递函数依赖

3NF 与 BCNF 的关系



❖ $R \in \text{BCNF} \xrightleftharpoons[\text{不必要}]{\text{充分}} R \in \text{3NF}$

❖ 如果 $R \in \text{3NF}$ ，且 R 只有一个候选码

$R \in \text{BCNF} \xrightleftharpoons[\text{必要}]{\text{充分}} R \in \text{3NF}$

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.7 多值依赖



[例 9] 学校中某一门课程由多个教师讲授，他们使用相同的一套参考书。每个教员可以讲授多门课程，每种参考书可以供多门课程使用。

多值依赖（续）



❖ 非规范化关系

课 程 C	教 员 T	参 考 书 B
物理	$\left\{ \begin{array}{l} \text{李 勇} \\ \text{王 军} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{普通物理学} \\ \text{光学原理} \\ \text{物理习题集} \end{array} \right\}$
数学	$\left\{ \begin{array}{l} \text{李 勇} \\ \text{张 平} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{数学分析} \\ \text{微分方程} \\ \text{高等代数} \end{array} \right\}$
计算数学	$\left\{ \begin{array}{l} \text{张 平} \\ \text{周 峰} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{数学分析} \\ \dots \end{array} \right\}$
⋮	⋮	⋮

多值依赖（续）



❖ 用二维表表示 Teaching

课程 C	教员 T	参考书 B
物理	李勇	普通物理学
物理	李勇	光学原理
物理	李勇	物理习题集
物理	王军	普通物理学
物理	王军	光学原理
物理	王军	物理习题集
数学	李勇	数学分析
数学	李勇	微分方程
数学	李勇	高等代数
数学	张平	数学分析
数学	张平	微分方程
数学	张平	高等代数
...

多值依赖（续）



❖ Teaching \in BCNF


❖ Teaching 具有唯一候选码 (C, T, B), 即全码

多值依赖（续）



Teaching 模式中存在的问题

- (1) 数据冗余度大
- (2) 插入操作复杂
- (3) 删除操作复杂
- (4) 修改操作复杂



存在
多值依赖

多值依赖（续）



❖ 定义 6.9

设 $R(U)$ 是一个属性集 U 上的一个关系模式， X 、 Y 和 Z 是 U 的子集，并且 $Z = U - X - Y$ 。关系模式 $R(U)$ 中多值依赖 $X \twoheadrightarrow Y$ 成立，当且仅当对 $R(U)$ 的任一关系 r ，给定的一对 (x, z) 值，有一组 Y 的值，这组值仅仅决定于 x 值而与 z 值无关

❖ 例 Teaching (C, T, B)

多值依赖（续）



❖ 多值依赖的另一个等价的形式化的定义：

在 $R(U)$ 的任一关系 r 中，如果存在元组 t, s 使得 $t[X]=s[X]$ ，那么就必然存在元组 $w, v \in r$ ，（ w, v 可以与 s, t 相同），使得 $w[X]=v[X]=t[X]$ ，而 $w[Y]=t[Y]$ ， $w[Z]=s[Z]$ ， $v[Y]=s[Y]$ ， $v[Z]=t[Z]$ （即交换 s, t 元组的 Y 值所得的两个新元组必在 r 中），则 Y 多值依赖于 X ，记为 $X \twoheadrightarrow Y$ 。这里， X, Y 是 U 的子集， $Z=U-X-Y$ 。

多值依赖（续）



❖ 平凡多值依赖和非平凡的多值依赖

- 若 $X \twoheadrightarrow Y$ ，而 $Z = \varphi$ ，则称 $X \twoheadrightarrow Y$ 为平凡的多值依赖
- 否则称 $X \twoheadrightarrow Y$ 为非平凡的多值依赖

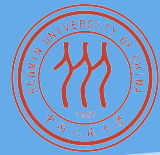
多值依赖（续）



[例 10] 关系模式 WSC (W , S , C)

- W 表示仓库， S 表示保管员， C 表示商品
- 假设每个仓库有若干个保管员，有若干种商品
- 每个保管员保管所在的仓库的所有商品
- 每种商品被所有保管员保管

多值依赖（续）

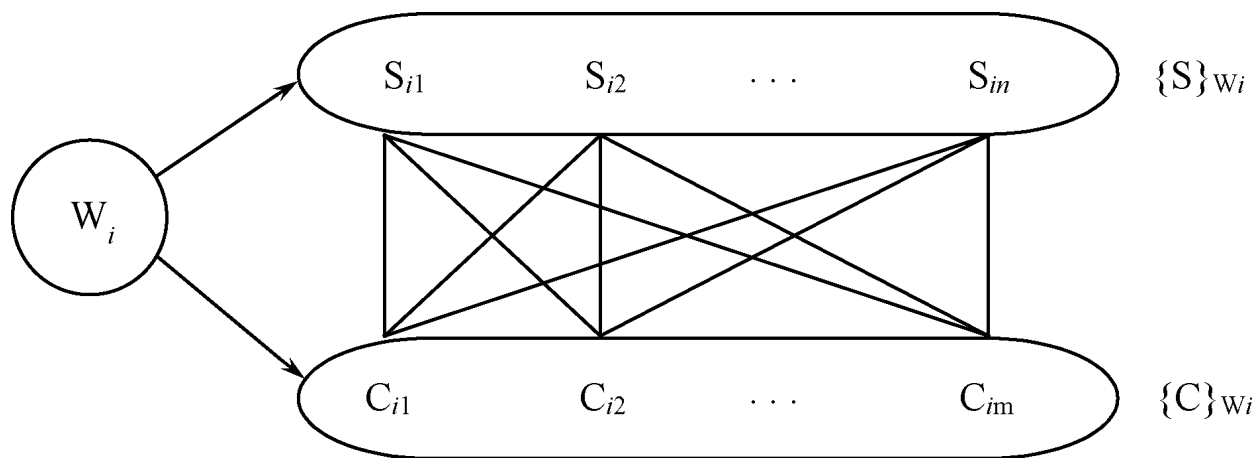


W	S	C
W1	S1	C1
W1	S1	C2
W1	S1	C3
W1	S2	C1
W1	S2	C2
W1	S2	C3
W2	S3	C4
W2	S3	C5
W2	S4	C4
W2	S4	C5

多值依赖（续）



用下图表示这种对应



$$W \twoheadrightarrow S \text{ 且 } W \twoheadrightarrow C$$

多值依赖的性质



(1) 多值依赖具有对称性

若 $X \twoheadrightarrow Y$ ，则 $X \twoheadrightarrow Z$ ，其中 $Z = U - X - Y$

(2) 多值依赖具有传递性

若 $X \twoheadrightarrow Y$ ， $Y \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Z - Y$

(3) 函数依赖是多值依赖的特殊情况。

若 $X \rightarrow Y$ ，则 $X \twoheadrightarrow Y$ 。

(4) 若 $X \twoheadrightarrow Y$ ， $X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Y \cup Z$ 。

(5) 若 $X \twoheadrightarrow Y$ ， $X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Y \cap Z$ 。

(6) 若 $X \twoheadrightarrow Y$ ， $X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Y - Z$ ， $X \twoheadrightarrow Z - Y$ 。

多值依赖与函数依赖的区别



(1) 多值依赖的有效性与属性集的范围有关

(2)

- 若函数依赖 $X \rightarrow Y$ 在 $R(U)$ 上成立，则对于任何 $Y' \subset Y$ 均有 $X \rightarrow Y'$ 成立
- 多值依赖 $X \twoheadrightarrow Y$ 若在 $R(U)$ 上成立，不能断言对于任何 $Y' \subset Y$ 有 $X \twoheadrightarrow Y'$ 成立

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

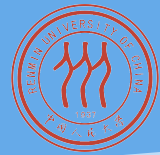
6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.8 4NF



- ❖ **定义 6.10** 关系模式 $R \langle U, F \rangle \in 1NF$ ，如果对于 R 的每个非平凡多值依赖 $X \twoheadrightarrow Y \setminus X$ ($Y \subseteq X$)， X 都含有码，则 $R \in 4NF$ 。
- ❖ 如果 $R \in 4NF$ ，则 $R \in BCNF$
 - 不允许有非平凡且非函数依赖的多值依赖
 - 允许的非平凡多值依赖是函数依赖

4NF (续)



例: Teaching(C,T,B) \notin 4NF

存在非平凡的多值依赖 $C \twoheadrightarrow T$, 且 C 不是码

- 用投影分解法把 Teaching 分解为如下两个关系模式:

$CT(C, T) \in 4NF$

$CB(C, B) \in 4NF$

$C \twoheadrightarrow T$, $C \twoheadrightarrow B$ 是平凡多值依赖

6.2 规范化



6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

6.2.9 规范化小结



- ❖ 关系数据库的规范化理论是数据库逻辑设计的工具
- ❖ 目的：尽量消除插入、删除一场，修改复杂，数据冗余
- ❖ 基本思想：逐步消除数据依赖中不合适的部分
 - 实质：概念的单一化

规范化小结（续）



❖ 关系模式规范化的基本步骤

消除决定属性
集非码的非平
凡函数依赖

1NF

↓ 消除非主属性对码的部分函数依赖

2NF

↓ 消除非主属性对码的传递函数依赖

3NF

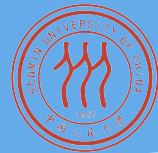
↓ 消除主属性对码的部分和传递函数依赖

BCNF

↓ 消除非平凡且非函数依赖的多值依赖

↓ 4NF

规范化小结（续）



- ❖ 不能说规范化程度越高的关系模式就越好
- ❖ 在设计数据库模式结构时，必须对现实世界的实际情况和用户应用需求作进一步分析，确定一个合适的、能够反映现实世界的模式
- ❖ 上面的规范化步骤可以在其中任何一步终止