



数据库技术与应用

关系数据理论

第六章 关系数据理论

6.1 问题的提出

6.2 规范化

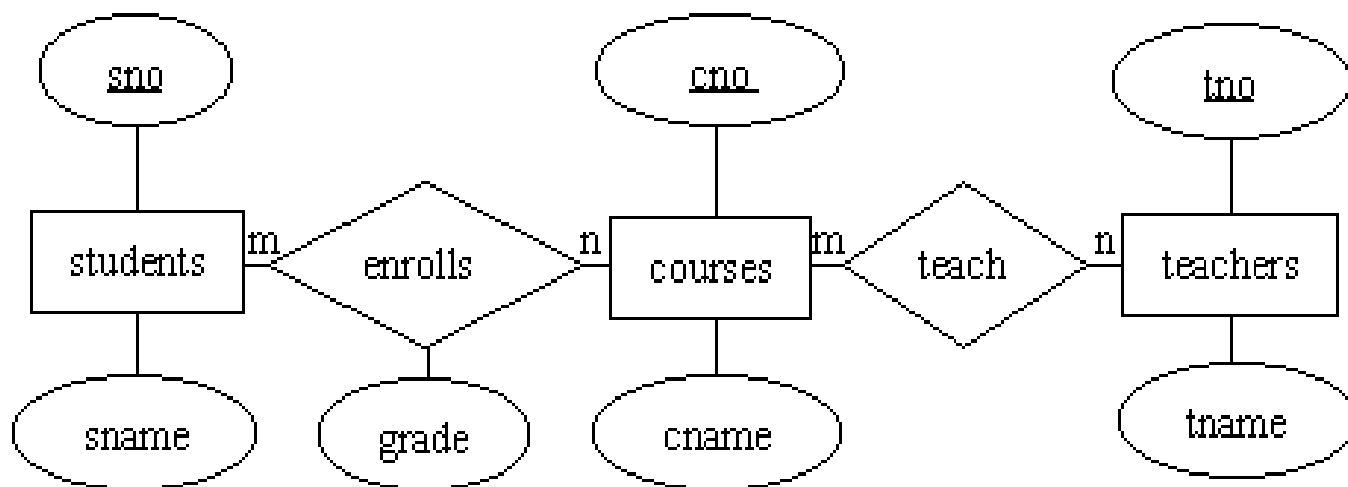
6.3 数据依赖的公理系统

6.4 模式的分解

6.5 小结

关系模式的评价

- ❖ 关系数据库设计的核心：关系模式设计
- ❖ 关系模式的设计
 - 按照一定的原则从数量众多而又相互关联的数据中，构造出一组既能较好地反映现实世界，而又有良好的操作性能的关系模式。
- ❖ 关系模式优劣，如何评价，如何改进？



简单教学管理的实体联系模型 ER 图

解一 sct (sno, cno, tno, sname, grade, cname, tname)

Sno	Cno	Tno	Sname	Grade	Cname	Tname
S1	C1	T1	赵民	90	OS	彭
S1	C2	T2	赵民	90	DS	杨
S1	C3	T3	赵民	85	C++	刘
S1	C4	T4	赵民	87	DB	张
S2	C1	T4	李军	90	OS	张
S3	C1	T4	陈江	75	OS	张
S3	C2	T2	陈江	70	DS	杨
S3	C4	T4	陈江	56	DB	张
S4	C1	T1	魏致	90	OS	彭
S4	C2	T2	魏致	85	DS	杨
S5	C1	T1	乔远	95	OS	彭
S5	C4	T4	乔远	80	DB	张

解法一问题分析

- 冗余度高
- 修改困难
- 插入问题
- 删除问题

★产生问题的原因

属性间约束关系（即数据间的依赖关系）太强

★解决问题的方法： 分解关系

解二：

students (sno, sname)

courses (cno, tno, cname)

teachers (tno, cno, tname)

enrolls (sno, cno, grade)

解三：

students (sno, sname)

courses (cno, cname)

teachers (tno, tname)

enrolls (sno, cno, grade)

teaching (tno, cno)

Students

Sno	Sname
S1	赵民
S2	李军
S3	陈江
S4	魏致
S5	乔远

Teachers

Cno	Cname
T1	彭
T2	杨
T3	刘
T4	张

Enrolls

Sno	Cno	Grade
S1	C1	90
S1	C2	90
S1	C3	85
S1	C4	87
S2	C1	90
S3	C1	75
S3	C2	70
S3	C4	56
S4	C1	90
S4	C2	85
S5	C1	95
S5	C4	80

Courses

Cno	Cname
C1	OS
C2	DS
C3	C++
C4	DB

Teaching

Cno	Tno
C1	T1
C1	T4
C2	T2
C3	T3
C4	T4

冗余度高、修改困难、插入问题、删除问题：
已不存在

关系数据库设计中出现的问题

- ❖ 优秀的数据库设计是应用成功的基石
- ❖ 关系数据库设计需要找到一个“好”的数据模式，“不好”的数据模式会导致：
 - 信息重复.
 - 无法表达确定的信息.
- ❖ 设计目标：
 - 避免数据冗余
 - 避免数据更新时的各种异常

目标：如何设计一套理论来

- ❖ 判断一个关系模式 R 是否是“好”模式
- ❖ 如果一个关系模式 R 不是“好”模式的话, 就将它分解成多个模式的集合 $\{R_1, R_2, \dots, R_n\}$, 从而保证每个模式是“好”模式, 并且分解是无损连接的
- ❖ 理论依据
 - 数据依赖：函数依赖、多值依赖
 - 范式（NF）

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。

关系模式的形式化定义

关系模式由五部分组成，即它是一个五元组：

$R(U, D, DOM, F)$

R: 关系名

U: 组成该关系的属性名集合

D: 属性组U中属性所来自的域

DOM: 属性向域的映象集合

F: 属性间数据的依赖关系集合

简化为一个三元组：

$R(U, F)$

什么是数据依赖

- ❖ 客观世界中事物间的联系：
 - 实体与实体的联系——数据模型
 - 实体内部属性间的联系——数据依赖
- ❖ 属性间的联系分为三类：
 - ◆ 一对一 ◆ 一对多 ◆ 多对多
- ❖ 数据依赖：关系中属性值之间相互依赖相互制约的联系。
- ❖ 属性间的数据依赖类型主要有两种：
 - 函数依赖
 - 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

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

例：R(Sno, Sname, Sdept, Sage)

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 $Y \subsetneq X$ ，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(trivial)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 $X \nrightarrow Y$ 。

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 $X \xrightarrow{P} Y$ 。

例： $R(\underline{Sno}, Sname, Sdept, Sage)$



思考题

❖ 平凡函数依赖与部分函数依赖有什么区别？

四、传递函数依赖

定义6.3 在 $R(U)$ 中, 如果 $X \rightarrow Y$, $(Y \not\subseteq X)$, $Y \not\rightarrow X$, $Y \rightarrow Z$,
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 。

确定函数依赖的方法

❖ 属性间的联系决定函数依赖关系

- X、Y有1:1关系，则 $X \rightarrow Y$ ， $Y \rightarrow X$ 。可表示成： $X \leftrightarrow Y$ 。
- X、Y有1:m关系，则 $Y \rightarrow X$ ，但 $X \not\rightarrow Y$ 。（如班主任：学生，则学生 \rightarrow 班主任，但班主任 \rightarrow 学生）
- X、Y有n:m关系，则X与Y不存在任何函数依赖。

举例

[例1]建立一个描述学生的数据库：

学生的学号（Sno）、所在系（Sdept）

学生住处（Sloc）、课程号（Cno）

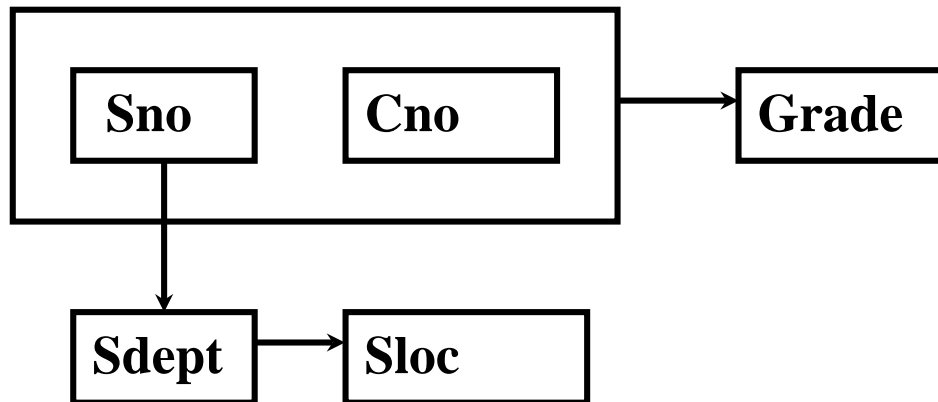
成绩（Grade）

单一的关系模式： $\text{Student} \langle U, F \rangle$

$U = \{ Sno, Sdept, Sloc, Cno, Grade \}$

举例

$$F = \{ \text{Sno} \rightarrow \text{Sdept}, \text{Sdept} \rightarrow \text{Sloc}, \\ (\text{Sno}, \text{Cno}) \rightarrow \text{Grade} \}$$



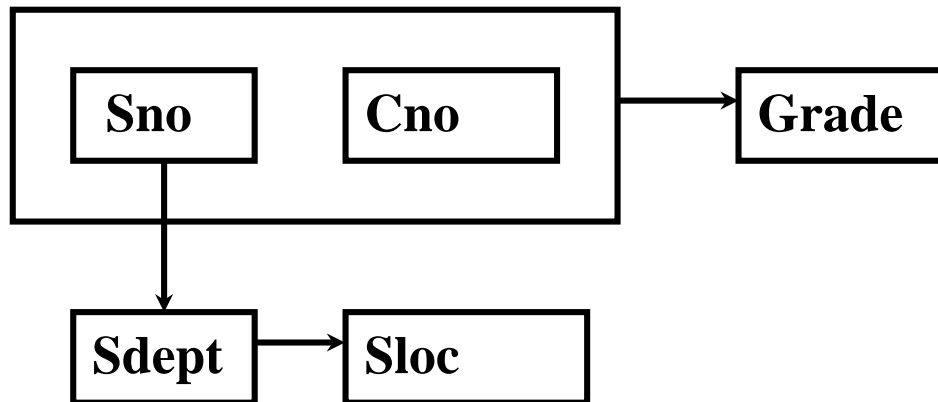
$(\text{Sno}, \text{Cno}) \xrightarrow{F} \text{Grade}$ 是完全函数依赖,

$(\text{Sno}, \text{Cno}) \twoheadrightarrow \text{Sdept}$ 是部分函数依赖

因为 $\text{Sno} \rightarrow \text{Sdept}$ 成立, 且 Sno 是 (Sno, Cno) 的真子集

举例

$F = \{ \text{Sno} \rightarrow \text{Sdept}, \text{Sdept} \rightarrow \text{Sloc},$
 $(\text{Sno}, \text{Cno}) \rightarrow \text{Grade} \}$



$\text{Sno} \rightarrow \text{Sdept}, \text{Sdept} \rightarrow \text{Sloc}$
Sloc传递函数依赖于Sno

举例

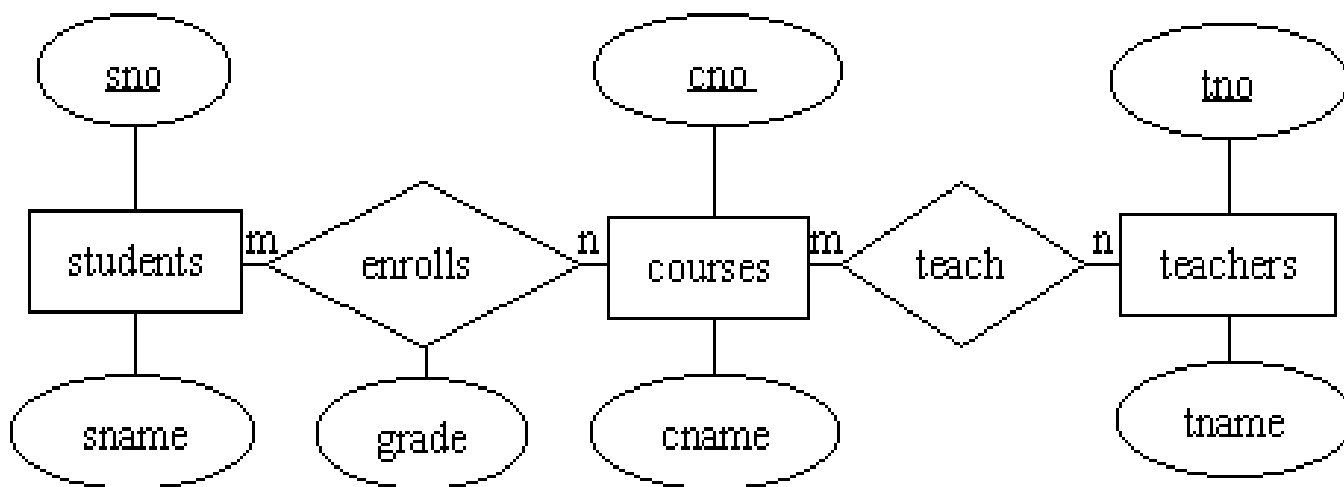
❖ 把这个单一模式分成3个关系模式：

S (Sno, Sdept, Sno \rightarrow Sdept) ;

SC (Sno, Cno, Grade, (Sno, Cno) \rightarrow Grade) ;

DEPT (Sdept, Sloc, Sdept \rightarrow Sloc) ;

举例



简单教学管理的实体联系模型 ER 图

❖ sct (sno, cno, tno, sname, grade, cname, tname)
关系模式sct的函数依赖和FD图？

$sno \rightarrow sname, cno \rightarrow cname, tno \rightarrow tname$

$(sno, cno) \rightarrow grade$

$(sno, cno) \rightarrow tno$

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<U,F>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）。

❖ 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

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
- ❖ 范式概念的发展
 - ❖ 1971~1972年，Codd系统地提出了**1NF**、**2NF**、**3NF**的概念
 - ❖ 1974年，Codd和Boyce共同提出了**BCNF**
 - ❖ 1976年，Fagin提出了**4NF**
 - ❖ 后来又有人提出了**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, Cno, Sloc, Sdept, Grade)

❖ 函数依赖包括:

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

$Sno \rightarrow Sdept$

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

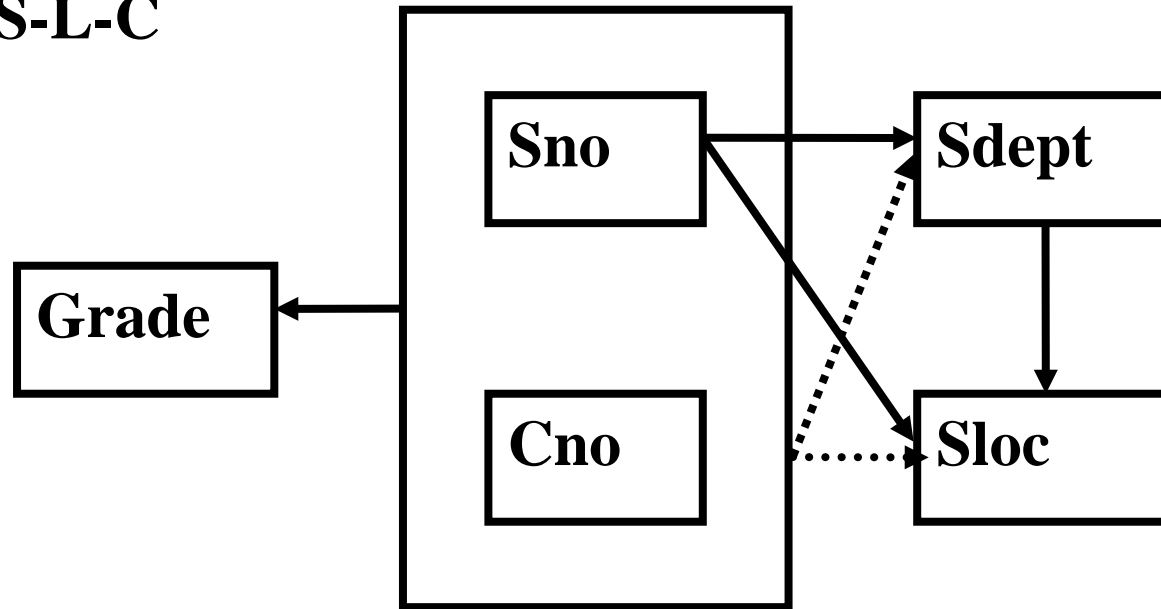
$Sno \rightarrow Sloc$

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

$Sdept \rightarrow Sloc$

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）

S-L-C(Sno, Cno, Sdept, Sloc, Grade)

插入异常：插入一个学生，还未选课

删除异常：删除一个学生选的唯一课程

数据冗余度大：一个学生选修k门课，Sdept、Sloc重复存储k次

❖ 原因

Sdept、Sloc部分函数依赖于码。

❖ 解决方法

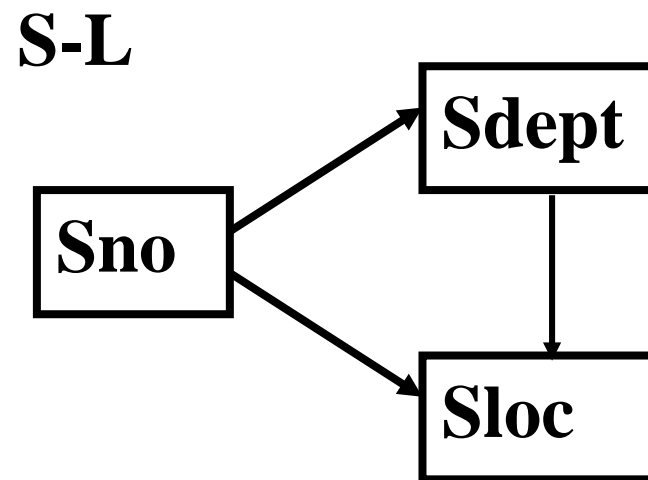
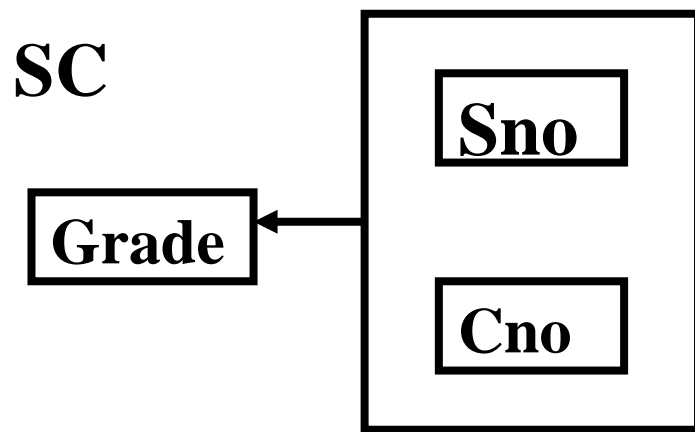
S-L-C分解为两个关系模式，以消除这些部分函数依赖

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

3NF

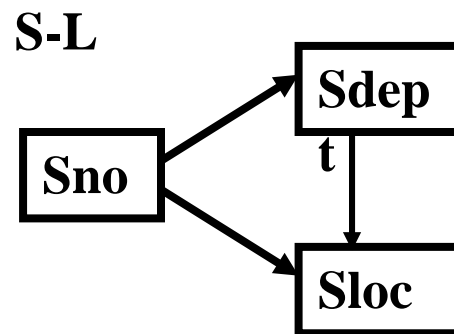
例：2NF关系模式S-L(Sno, Sdept, Sloc)

修改复杂：一个系换学生宿舍楼

■ 函数依赖：

⑩ Sno → Sdept

⑩ Sdept → Sloc



Sno ^{传递} → Sloc, 即S-L中存在非主属性对码的传递函数依赖

3NF (续)

解决方法

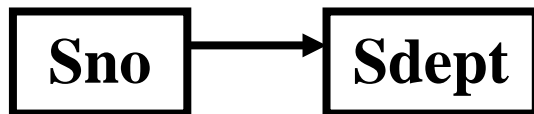
采用投影分解法，把S-L分解为两个关系模式，以消除传递函数依赖：

S-D (Sno, Sdept)

D-L (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



S-D



D-L

3NF (续)

❖ 3NF的定义

定义6.7 关系模式 $R\langle U, F \rangle$ 中若不存在这样的码 X 、属性组 Y 及非主属性 Z ($Z \not\subseteq Y$), 使得 $X \rightarrow Y$, $Y \rightarrow Z$ 成立, 且 $Y \not\rightarrow X$, 则称 $R\langle U, F \rangle \in 3NF$ 。

■ 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。

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

如何判断R为第几范式？

- ❖ 已知一个关系模式的属性之间的语义，也就是相互依赖的关系，如何判断该模式满足第几范式？
 - 1、首先要通过语义把属性之间的函数依赖关系列出来，
 - 2、然后确定哪些属性组合可以候选码，从而找出非主属性和主属性。
 - 3、然后判断是否存在非主属性与码之间的部分函数依赖关系，如果存在，则不满足**2NF**，如不存在部分函数依赖，则属于**2NF**，
 - 4、继续进行下一步判断；判断非主属性与码之间存在传递依赖关系，不存在，则为**3NF**；
 - 5、决定因素是否包含码，满足条件则为**BCNF**

练习

❖ 分析下面关系的函数依赖，是否为3NF？并分解。

❖ 关系W（工号，姓名，工种，定额）

每个工种有一个定额

❖ 关系R（材料号、材料名、生产厂）

材料号	材料名	生产厂
M1	线材	武汉
M2	型材	武汉
M3	板材	广东
M4	型材	武汉

练习

- ❖ 关系W（工号，姓名，工种，定额）
工号→姓名，工号→工种，工种→定额，
候选码为工号，存在传递依赖：不是3NF
（工号，姓名，工种），
（工种，定额）
- ❖ 关系R（材料号、材料名、生产厂）
材料号→材料名，材料名→生产厂，
候选码为材料号，存在传递依赖，不是3NF
（材料号，材料名）
（材料名，生产厂）

练习

- ❖ 假设某商业集团数据库中有一关系模式R如下
 - R(商店编号, 商品编号, 商品库存数量, 部门编号, 负责人)
- ❖ 如果规定
 - (1)每个商店的每种商品只在该商店的一个部门销售;
 - (2)每个商店的每个部门只有一个负责人;
 - (3)每个商店的每种商品只有一个库存数量。
- ❖ 试回答下列问题
 - (1)写出关系模式R的函数依赖关系和主码。
 - (2)该关系模式最高满足第几范式? 举例可能的异常
 - (3)将该关系模式分解为3NF

练习

❖ (1)

- (商店编号, 商品编号) --> 部门编号
- (商店编号, 部门编号) --> 负责人
- (商店编号, 商品编号) --> 库存量
- 主码: 商店编号, 商品编号

❖ (2)

- 不存在非主属性对码的部分依赖, 属于2NF
- (商店编号, 商品编号) --> 负责人
- 存在非主属性“负责人”对码的传递依赖, 不属于3NF

❖ (3)

- R1(商店编号, 商品编号, 商品库存数量, 部门编号)
- R2(商店编号, 部门编号, 负责人)

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 $R\langle U, F \rangle \in 1NF$, 若 $X \rightarrow Y$ 且 $Y \not\subseteq X$ 时 X 必含有码, 则 $R\langle U, F \rangle \in BCNF$ 。

❖ 等价于: 每一个决定属性因素都包含码

一个BCNF关系模式中,消除了任何属性(包括主属性)对码的传递依赖与部分依赖

反例: $K \rightarrow R$, K 为码; 若有 $X \rightarrow Y$, 而 X 不包含 K , 则

- 1、 $K \rightarrow X$, $X \rightarrow Y$, 而 $K \rightarrow Y$, 即 Y 传递依赖于码;
- 2、 K 包含 X , 而 $X \rightarrow Y$, $K \rightarrow Y$, 即 Y 部分依赖于码;

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

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

- 若 R 只有一个候选码，必要性成立

BCNF (续)

[例5] 关系模式C (Cno, Cname, Pcnno)

- $C \in 3NF$

- $C \in BCNF$

[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）

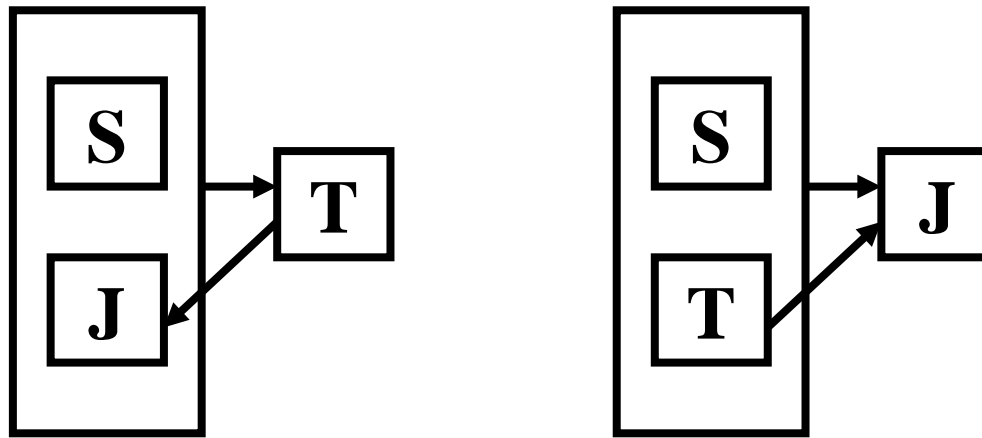
[例7] 关系模式SJP（S，J，P）中，S表示学生，J表示课程，P表示名次。

- 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
- $SJP \in 3NF$,
- $SJP \in BCNF$

[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：
 $(S, J) \rightarrow T, T \rightarrow J$
- (S, J) 和 (S, T) 都是候选码

BCNF (续)



STJ中的函数依赖

❖ $STJ \in 3NF$

- 没有任何非主属性对码传递依赖或部分依赖

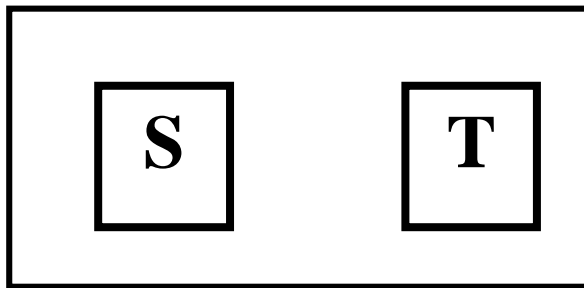
❖ $STJ \notin BCNF$

- T是决定因素，T不包含码

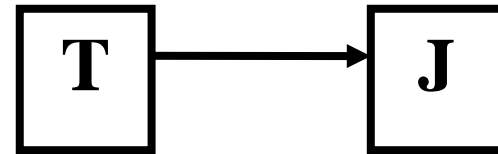
BCNF (续)

❖ 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

作业

❖ 教材第六章 习题2