

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

6.1 问题的提出

关系数据库逻辑设计

- 针对具体问题，如何构造一个适合于它的数据模式
- 数据库逻辑设计的工具——关系数据库的规范化理论

6.1 问题的提出

一、概念回顾

二、关系模式的形式化定义

三、什么是数据依赖

四、关系模式的简化定义

五、数据依赖对关系模式影响

一、概念回顾

- **关系**：描述实体、属性、实体间的联系。
 - 从形式上看，它是一张二维表，是所涉及属性的笛卡尔积的一个子集。
- **关系模式**：用来定义关系。
- **关系数据库**：基于关系模型的数据库，利用关系来描述现实世界。
 - 从形式上看，它由一组关系组成。
- **关系数据库的模式**：定义这组关系的关系模式的全体。

二、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

三、什么是数据依赖

- 数据依赖概念：

是通过一个关系中属性间值的相等与否体现出来的数据间的相互关系。是现实世界属性间相互联系的抽象，是数据内在的性质，是语义的体现。

所讨论的是一个关系中，各个属性间数值的关系。比如：学生的姓名与学号、国家的名称与国旗等等。

三、什么是数据依赖

● 数据依赖的类型

- 函数依赖 (Functional Dependency, FD)
- 多值依赖 (Multivalued Dependency, MVD) (自学)

四、关系模式的简化表示

- 关系模式 $R(U, D, DOM, F)$

简化为一个三元组：

$$R(U, F)$$

- 当且仅当 U 上的一个关系 r 满足 F 时， r 称为关系模式 $R(U, F)$ 的一个关系

五、数据依赖对关系模式的影响

例：描述学校的数据库：

学生的学号（Sno）、所在系（Sdept）

系主任姓名（Mname）、课程名（Cname）

成绩（Grade）

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

$U = \{ \text{Sno, Sdept, Mname, Cname, Grade} \}$

F 数据依赖关系应如何分析？

五、数据依赖对关系模式的影响

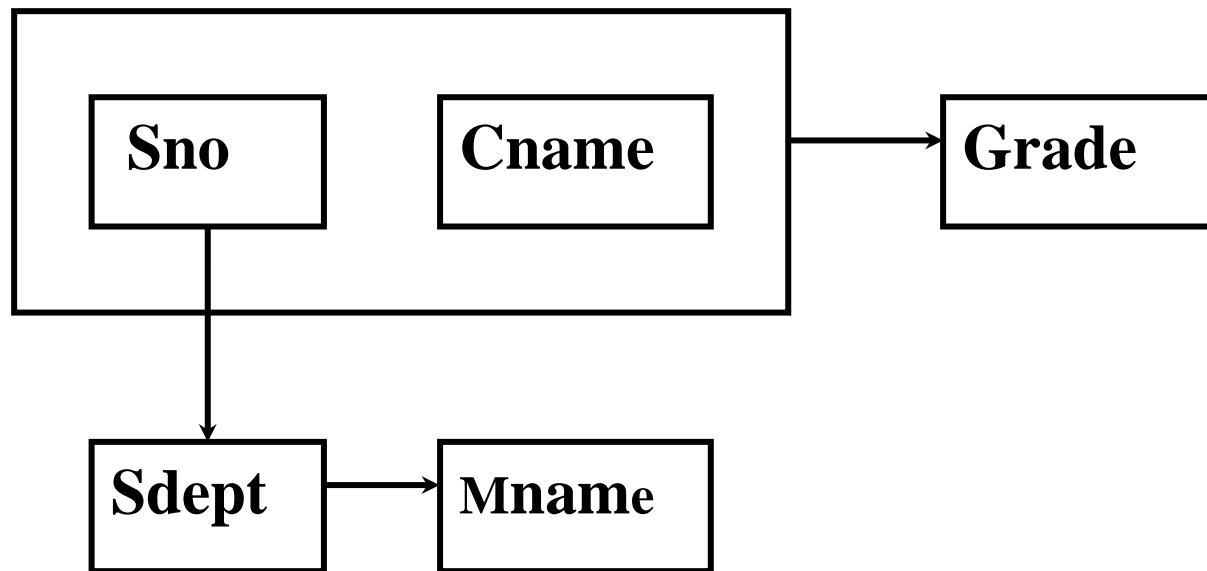
学校数据库的语义：

- 1.一个系有若干学生，一个学生只属于一个系；
- 2.一个系只有一名主任；
- 3.一个学生可以选修多门课程， 每门课程有若干学生选修；
4. 每个学生所学的每门课程都有一个成绩。

五、数据依赖对关系模式的影响

因此，属性组U上的一组函数依赖F：

$$F = \{ \text{Sno} \rightarrow \text{Sdept}, \text{Sdept} \rightarrow \text{Mname}, \\ (\text{Sno}, \text{Cname}) \rightarrow \text{Grade} \}$$



关系模式Student<U, F>中存在的问题

1. 数据冗余太大

➤ 浪费大量的存储空间

例：每一个系主任的姓名重复出现

2. 更新异常（Update Anomalies）

➤ 数据冗余，更新数据时，维护数据完整性代价大。

例：某系更换系主任后，系统必须修改与该系学生有关的每一个元组。

关系模式Student<U, F>中存在的问题

3. 插入异常 (Insertion Anomalies)

➤ 该插的数据插不进去

例：如果一个系刚成立，尚无学生，我们就无法把这个系及其系主任的信息存入数据库。

4. 删除异常 (Deletion Anomalies)

➤ 不该删除的数据不得不删

例：如果某个系的学生全部毕业了，我们在删除该系学生信息的同时，把这个系及其系主任的信息也丢掉了。

五、数据依赖对关系模式的影响

结论：

- Student关系模式不是一个好的模式。

“好”的模式：

不会发生插入异常、删除异常、更新异常，
数据冗余应尽可能少。

原因：由存在于模式中的某些数据依赖引起的

解决方法：需要进行规范化，进行关系模式的分解来消除其中不合适的数据依赖。

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

6.2 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。

6.2.1 函数依赖

一、函数依赖

二、平凡函数依赖与非平凡函数依赖

三、完全函数依赖与部分函数依赖

四、传递函数依赖

一、函数依赖定义

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 $X \rightarrow Y$ 。

X 称为这个函数依赖的**决定因素**。

$Y=f(x)$

说明:

1. 函数依赖不是指关系模式R的某个或某些关系实例满足的约束条件，而是指R的**所有关系实例**均要满足的约束条件。
2. 函数依赖是**语义范畴**的概念。只能根据数据的语义来确定函数依赖。

例如“**姓名→年龄**”这个函数依赖只有在不允许有同名人的条件下成立

说明:

3. 数据库设计者可以对现实世界作强制的规定。

例如:规定不允许同名人出现, 函数依赖“姓名 \rightarrow 年龄”成立。

所插入的元组必须满足规定的函数依赖, 若发现有同名人存在, 则拒绝装入该元组。

一、函数依赖定义

例: Student(Sno, Sname, Ssex, Sage, Sdept)

假设不允许重名, 则有:

$Sno \rightarrow Ssex,$

$Sno \rightarrow Sage,$

$Sno \rightarrow Sdept,$

$Sno \rightarrow Sname$

$Sname \rightarrow Ssex,$

$Sname \rightarrow Sage,$

$Sname \rightarrow Sdept,$

$Sname \rightarrow Sno$

$Sno \longleftrightarrow Sname$

➤ 若Y不函数依赖于X, 则记为 $X \nrightarrow Y$ 。 $Sdept \nrightarrow Sage$

➤ 若 $X \rightarrow Y$, 并且 $Y \rightarrow X$, 则记为 $X \longleftrightarrow Y$ 。 (一对一)

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： $(Sno, Cno) \rightarrow Sno$

$(Sno, Cno) \rightarrow Cno$

二、平凡函数依赖与非平凡函数依赖

- 对于任一关系模式，平凡函数依赖都是必然成立的，它不反映新的语义，因此若不特别声明，我们总是讨论非平凡函数依赖。

三、完全函数依赖与部分函数依赖

定义6.2 在关系模式 $R(U)$ 中，如果 $X \rightarrow Y$ ，并且对于 X 的任何一个真子集 X' ，有 $X' \not\rightarrow Y$ ，则称 Y 完全函数依赖于 X ，记作 $X \xrightarrow{F} Y$ 。

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

三、完全函数依赖与部分函数依赖

例：在关系SC(Sno, Cno, Grade)中，

由于：Sno \nrightarrow Grade, Cno \nrightarrow Grade,

因此：(Sno, Cno) \xrightarrow{F} Grade

请考查函数依赖的类型：

- 如果每个人每个月的薪水不同

F

- 如果每个月的薪水不变化

P

工号	月份	薪水
----	----	----

X \longrightarrow **Y**

四、传递函数依赖

定义6.3 在关系模式R(U)中, 如果 $X \rightarrow Y$, $Y \rightarrow Z$, 且 $Y \not\subseteq X$, $Y \not\rightarrow X$, 则称Z传递函数依赖于X。

非平凡函数依赖

注: 如果 $Y \rightarrow X$, 即 $X \leftrightarrow Y$, 则Z直接依赖于X, 不是传递函数依赖了。

说明X与Y是多对一映射

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

6.2.2 码

定义6.4 设K为关系模式R<U,F>中的属性或属性组合。若 $K \xrightarrow{F} U$ ，则K称为R的一个**候选码**。

若关系模式R有多个候选码，则选定其中的一个做为**主码**

新的定义方式(决定性、最小性)

①决定性: $K \rightarrow U$

②最小性: $\neg \exists K' \subset K$, 使得
 $K' \rightarrow U$

- 主属性与非主属性
- 外码

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 $R \in nNF$ 。
- **规范化**过程：一个低一级范示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。

6.2.4 2NF

非规范化关系

SUPERVISOR	SPECIALITY	POSTGRADUATE	
		PG1	PG2
张清玫 刘逸	信息专业 信息专业	李勇 王敏	刘晨

□ 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。

仅满足1NF的关系模式

例: 关系模式 SLC(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:

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

$\text{Sno} \rightarrow \text{Sdept}$

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

$\text{Sno} \rightarrow \text{Sloc}$

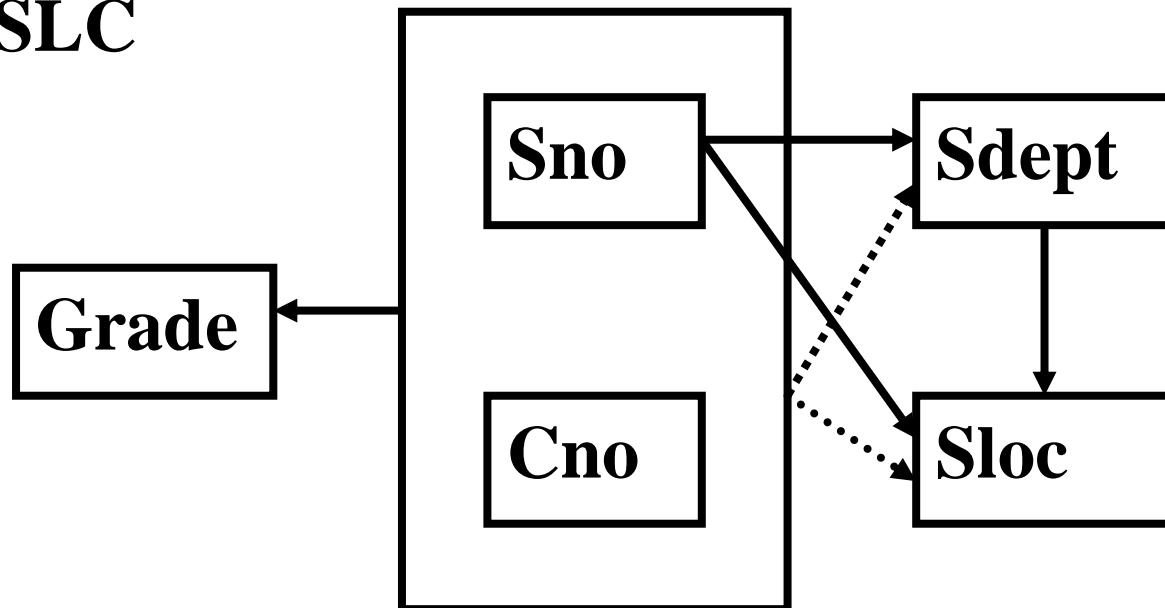
$(\text{Sno}, \text{Cno}) \xrightarrow{\text{P}} \text{Sloc}$

$\text{Sdept} \rightarrow \text{Sloc}$

仅满足1NF的关系模式

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

SLC

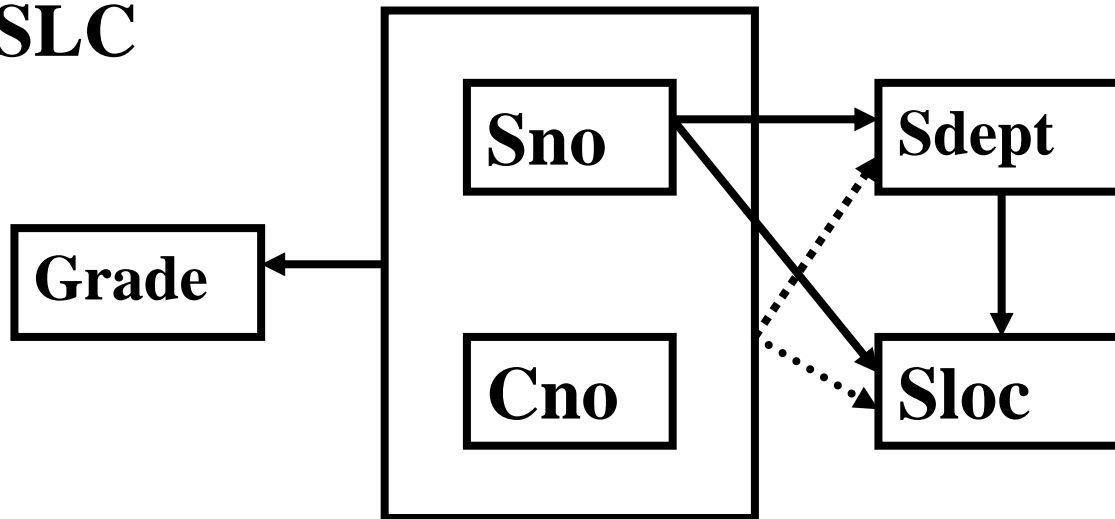


SLC关系模式存在的问题

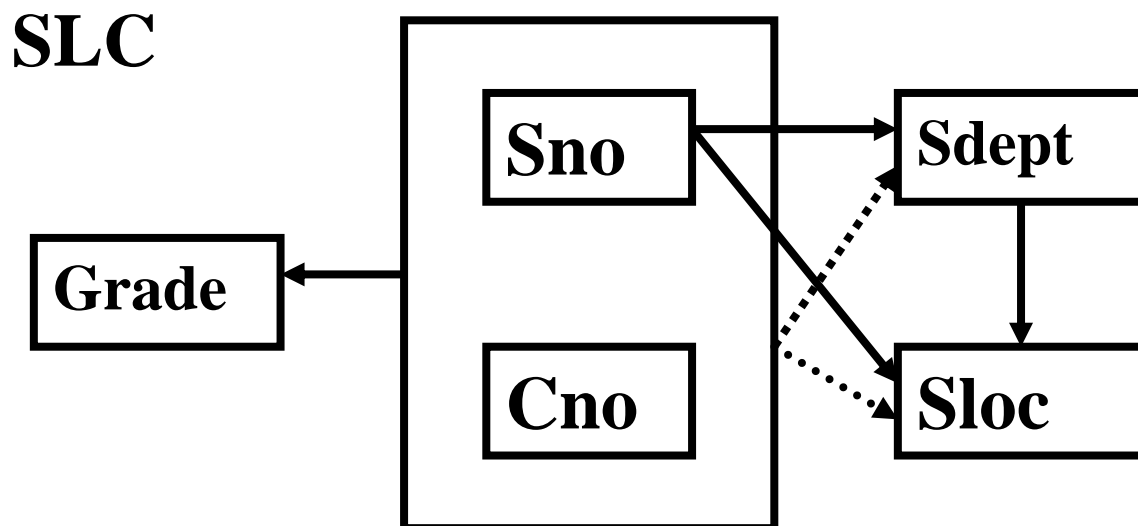
(1) 插入异常

假设 $Sno=95102$, $Sdept=IS$, $Sloc=N$ 的学生还未选课, 因**课程号是主属性**, 因此该学生的信息无法插入SLC。

SLC



SLC关系模式存在的问题



(2) 删除异常

假定某个学生本来只选修了3号课程这一门课。现在他连3号课程也不选修了。因**课程号是主属性**，此操作将导致该学生信息的整个元组都要删除。

SLC关系模式存在的问题

(3) 数据冗余度大

关系模式 SLC(Sno, Sdept, Sloc, Cno, Grade) 的
Sdept和Sloc值就要重复存储了多次。

SLC关系模式存在的问题

关系模式 SLC(Sno, Sdept, Sloc, Cno, Grade)

(4) 修改复杂

例如学生转系，在修改此学生元组的Sdept值的同时，还可能需要修改住处（Sloc）。

如果这个学生选修了K门课，则必须无遗漏地修改K个元组中全部Sdept、Sloc信息。

6.2.4 2NF

- SLC关系模式存在问题的原因是什么？

Sdept、Sloc部分函数依赖于码。

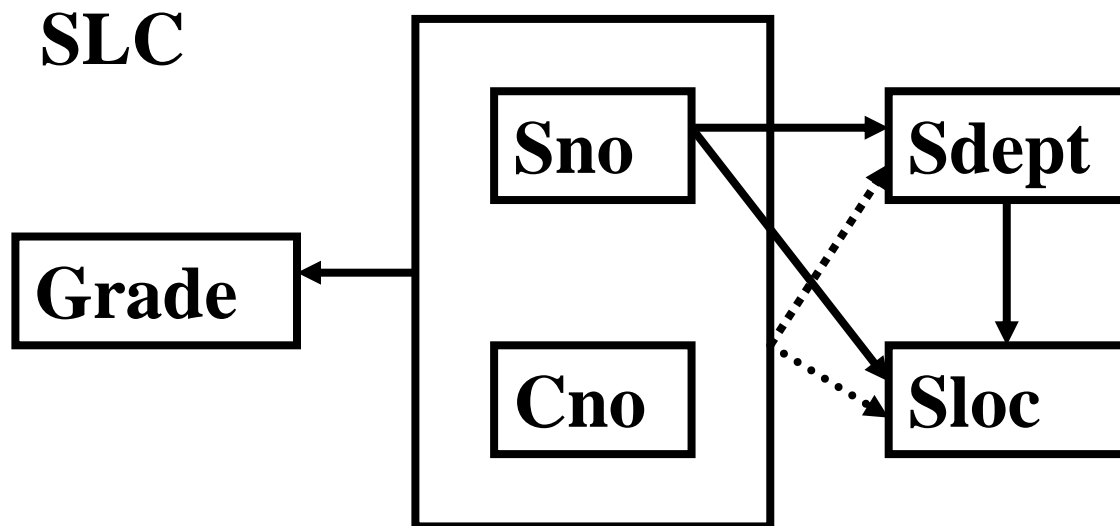
- 解决方法

SLC**分解**为两个关系模式，以消除这些部分函数依赖

SC (**Sno**, **Cno**, Grade)

SL (**Sno**, Sdept, Sloc)

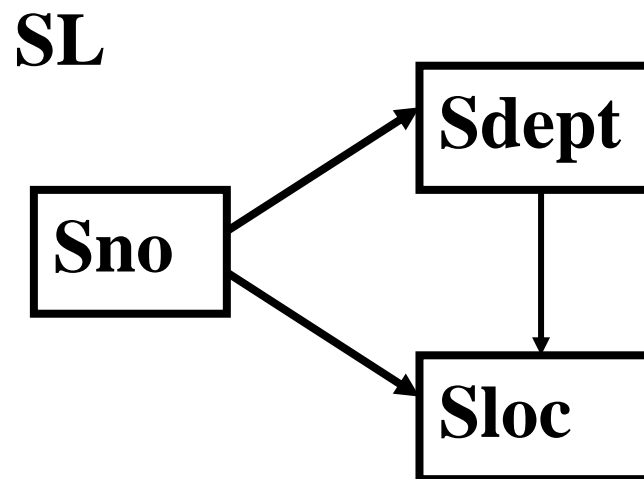
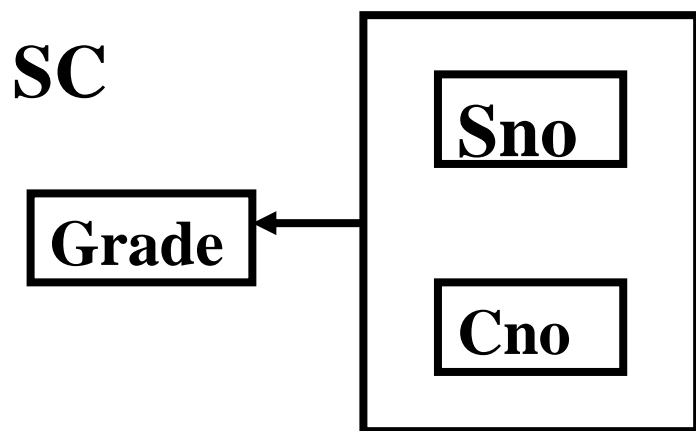
6.2.4 2NF



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

6.2.4 2NF

分解后的函数依赖图:



6.2.4 2NF

● 2NF的定义

定义6.6 若关系模式 $R \in 1NF$ ，并且每一个非主属性都完全函数依赖于任何一个候选码，则 $R \in 2NF$ 。

例：SLC(Sno, Sdept, Sloc, Cno, Grade) $\in 1NF$

SLC(Sno, Sdept, Sloc, Cno, Grade) $\notin 2NF$

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

6.2.4 2NF

- 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. 5 3NF

考查：2NF关系模式SL(Sno, Sdept, Sloc)中
函数依赖：

$Sno \rightarrow Sdept$

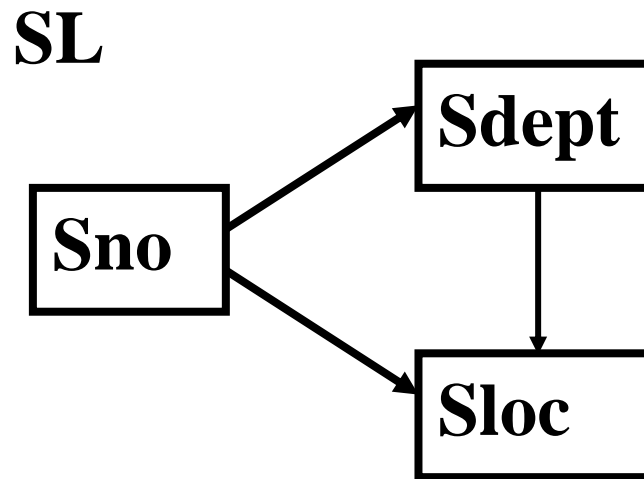
$Sdept \rightarrow Sloc$

$Sno \rightarrow Sloc$

Sloc传递函数依赖于Sno，即SL中存在非主属性
对码的传递函数依赖。

6.2.5 3NF

函数依赖图:



6.2.5 3NF

●解决方法

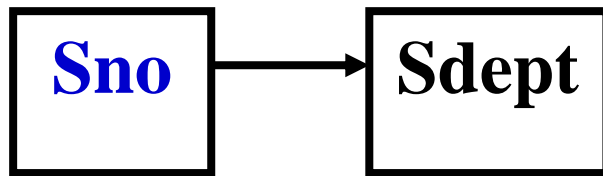
采用投影分解法，把SL分解为两个关系模式，以
消除传递函数依赖：

SD (Sno, Sdept)

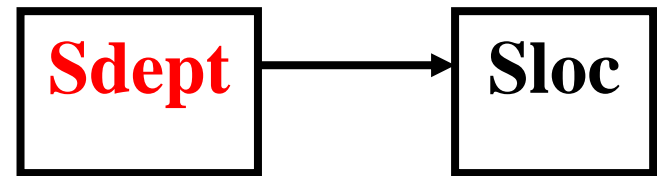
DL (Sdept, Sloc)

6.2.5 3NF

SD的码为Sno, DL的码为Sdept。



SD



DL

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 ($Z \not\subseteq Y$), 使得 $X \rightarrow Y$, ($Y \not\rightarrow X$), $Y \rightarrow Z$, 成立, 则称 $R<U, F> \in 3NF$ 。

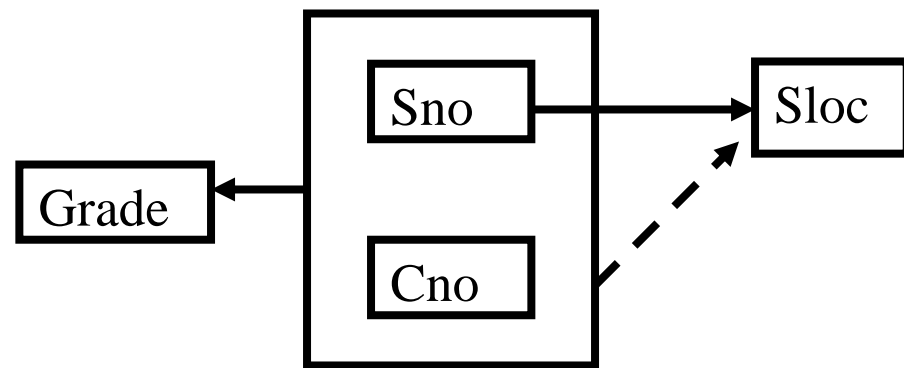
- 说明:
- ① 若 $Z \subseteq Y$, 则 $X \rightarrow Y$ 时必然有 $X \rightarrow Z$
 - ② 若 $Y \rightarrow X$, 则 $X \leftrightarrow Y$, $Y \rightarrow Z$ 为自身固有的依赖
 - ③ **3NF** 不含非主属性对码的传递依赖的关系模式

例: $SL(Sno, Sdept, Sloc) \in 2NF$
 $SL(Sno, Sdept, Sloc) \notin 3NF$
 $SD(Sno, Sdept) \in 3NF$
 $DL(Sdept, Sloc) \in 3NF$

6.2.5 3NF

- 若 $R \in 3NF$ ，则 R 的每一个非主属性既不部分函数依赖于候选码也不传递函数依赖于候选码。
- 如果 $R \in 3NF$ ，则 R 也是2NF。

思考：模式SCGL (Sno Cno Grade Sloc)是否属于2NF或3NF



2NF:不可以有非主属性对码部分函数依赖
SCGL不属于2NF

3NF:不可以有非主属性对码传递函数依赖
及部分函数依赖

SCGL不属于3NF

6.2.5 3NF

- 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，并不能完全消除关系模式中的各种异常情况和数据冗余。

6.2.5 3NF

定理： 如果 $R \in 3NF$ ，则 $R \in 2NF$ 。

证： $R \in 3NF \Rightarrow R \in 2NF \Leftrightarrow R \notin 2NF \Rightarrow R \notin 3NF$

设 R 不属于 $2NF$ ，则存在非主属性 A 部分依赖于码 K ， $K \xrightarrow{p} A$ ，即存在 K 的真子集 K' ，使得 $K' \rightarrow A$ ；又 $K' \subset K$ ，则 $K \rightarrow K'$ ；即存在 $K \rightarrow K'$ ， $K' \rightarrow A$ ， $K' \not\rightarrow K$ ， $A \not\subseteq K'$ ，由 $3NF$ 定义， R 不属于 $3NF$ 。

推论： $R \in 3NF$ ，则 R 中的每一个非主属性既不部分依赖于码，也不传递依赖于码。

6.2.6 BC范式 (BCNF)

- 定义6.8 设关系模式 $R\langle U, F \rangle \in 1NF$ ，如果对于R的每个函数依赖 $X \rightarrow Y$ ($Y \subseteq X$)，**X必包含码**，则 $R \in BCNF$ (Boyce Codd Normal Form)，又称修正(或扩充)的第三范式。

若 $R \in BCNF$

- 每一个决定因素都包含（候选）码
- 则 $R \in 3NF$
- 若 $R \in 3NF$ 则R不一定 $\in BCNF$

6.2.6 BCNF

例：在关系模式STC (S, T, C) 中，S表示学生，T表示教师，C表示课程。同书上185页例[6.8]

语义分析：

每一教师只教一门课。

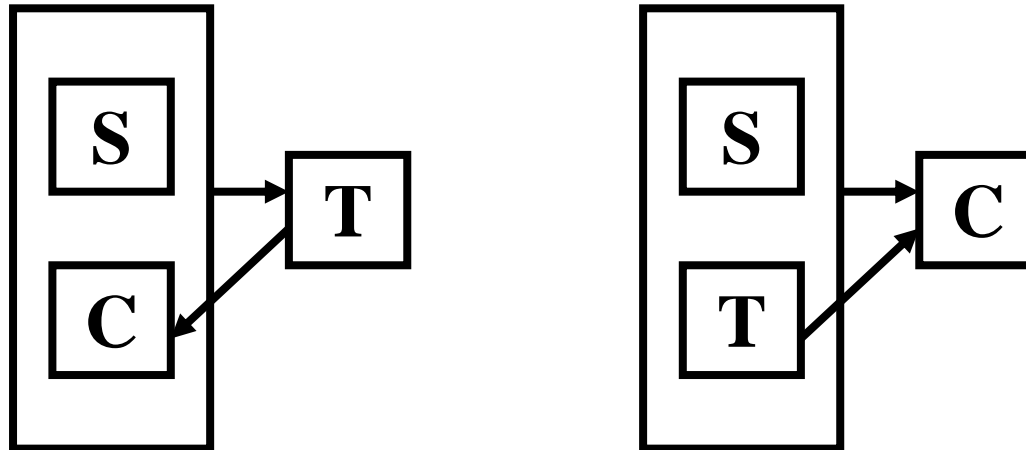
每门课由若干教师教，某一学生选定某门课，就确定了一个固定的教师。

某个学生选修某个教师的课就确定了所选课的名称。

数据依赖：

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

6. 2. 6 BCNF



STC

BCNF

$STC \in 3NF$

- (S, C) 和 (S, T) 都可以作为候选码
- S 、 T 、 C 都是主属性

$STC \notin BCNF$

- $T \rightarrow C$, T 是决定因素, T 不是候选码

BCNF

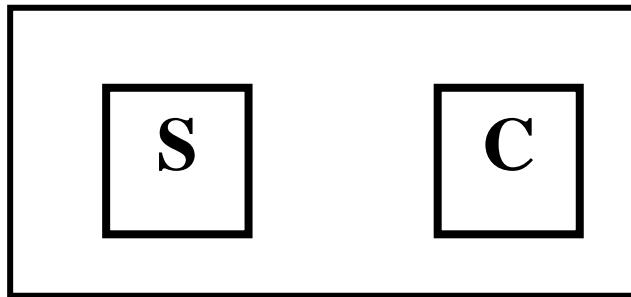
- K为候选码，有 $K \xrightarrow{F} U$ ，但是若A是U的一个属性，能推出来 $K \xrightarrow{F} A$ ？

BCNF

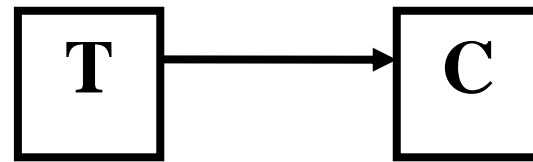
解决方法：将STC分解为二个关系模式：

$SC(\underline{S}, C) \in \text{BCNF}$, $TC(\underline{T}, C) \in \text{BCNF}$

每一个决定因素必含（候选）码



SC



TC

没有任何属性对码是部分函数依赖和传递函数依赖

3NF与BCNF的关系

推论:

- 如果关系模式 $R \in \text{BCNF}$ ，必定有 $R \in 3\text{NF}$

反证法:

设关系 $R \in \text{BCNF}$ ，但 $R \notin 3\text{NF}$ 。则关系 R 中存在候选码 X ，属性组 Y 和非主属性 Z ($Z \not\subseteq Y$)，满足 $X \rightarrow Y$ ， $Y \not\rightarrow X$ ， $Y \rightarrow Z$ 。

由于 $R \in \text{BCNF}$ ，则 Y 包含候选码或 Y 为候选码，则必有 $Y \rightarrow X$ ，与 $Y \not\rightarrow X$ 矛盾。

所以假设不成立。

- 如果关系模式 $R \in 3\text{NF}$ ，不一定有 $R \in \text{BCNF}$ P185

例6.8

- 如果 $R \in 3\text{NF}$ ，且 R 只有一个候选码，则 R 必属于 BCNF 。P184 例6.5

3NF与BCNF的关系

- 例[6.5]考察关系模式C (Cno, Cname, Cpno), 码仅有一个即Cno, 这里没有任何属性对Cno部分依赖或者传递依赖, 所以 $C \in 3NF$ 。同时C中Cno是唯一的决定因素, 所以 $C \in BCNF$ 。

对SC (Sno,Cno,Grade) 来说呢?

BCNF

●若 $R \in \text{BCNF}$ ，没有任何属性对码是部分函数依赖和传递函数依赖。

证明从两个层次来说：

1) 首先证明不存在非主属性对码具有部分函数依赖和传递函数依赖。

$R \in \text{BCNF} \Rightarrow R \in 3\text{NF} \Rightarrow R \in 2\text{NF}$ （前面已经证明），所以BCNF的R不存在非主属性对码的部分函数依赖和传递函数依赖。

2) 再证明不存在主属性对码具有部分函数依赖和传递函数依赖。

- 证明不存在主属性对码的部分函数依赖（反证法，假设某主属性A对码 K_2 有部分函数依赖，白板演示证明过程，根据BCNF的定义，这里的码 K_2 是不包含A的）
- 证明不存在主属性对码的传递函数依赖（反证法，假设某主属性对码有传递函数依赖，白板演示证明过程）

BC范式 (BCNF)

- 若 $R \in \text{BCNF}$ ，则 R 中所有非主属性对每一个码都完全函数依赖。如果关系模式 $R \in \text{BCNF}$ ，必定有 $R \in 3\text{NF}$ 。如果 $R \in 3\text{NF}$ ，则 $R \in 2\text{NF}$ 。由 2NF 定义得到，不存在非主属性对码的部分函数依赖。
- 若 $R \in \text{BCNF}$ ，则 R 中所有主属性对每个不包含它的码都完全函数依赖。（反证法）
- 若 $R \in \text{BCNF}$ ，则 R 中没有任何属性完全函数依赖于非码的任何一组属性。

（反证法）假设存在某属性 A 完全函数依赖于非码的一组属性 X 。

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。
- 一个关系只要其分量都是不可分的数据项，它就是规范化的关系，但这只是最基本的规范化。
- 规范化程度可以有多个不同的级别。

6.2.9 规范化小结

- 规范化程度过低的关系不一定能够很好地描述现实世界，可能会存在插入异常、删除异常、修改复杂、数据冗余等问题。
- 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集合，这种过程就叫关系模式的规范化。

6.2.9 规范化小结

● 关系模式规范化的基本步骤

