

6.2 规范化

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 第二范式 (2NF)

6.2.5 第三范式 (3NF)

6.2.6 BC范式 (BCNF)

*6.2.7 多值依赖

*6.2.8 第四范式 (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, Boyce和Codd共同提出的范式)

第四范式(4NF)

第五范式(5NF)

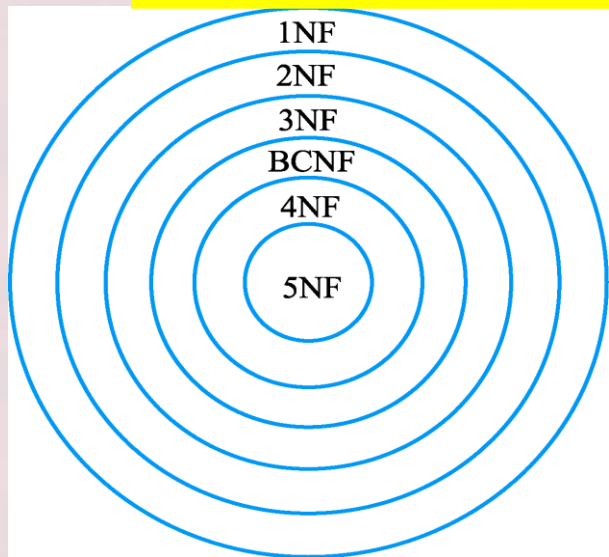


范式（续）

❖ 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（schema decomposition）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（normalization）。



第一范式

❖ 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 $R \in 1NF$ 。

科目名称	科目余额表					
	期初余额		本期发生额		期末余额	
	借方	贷方	借方	贷方	借方	贷方
现金	950		4,360	4,350	960	
银行存款	2,690		14,910	7,460	10,140	
应收帐款	16,660		1,740	18,400	0	
原材料	5,000		1,720	2,620	4,100	
预付账款	2,500		5,000	3,500	4,000	
待摊费用	500		200	100	600	
固定资产	5,400		5,000	0	10,400	
短期借款		2,000	2,000			0
应付帐款		3,700	4,400			2,000
应付票据	0	5,000	4,000		0	3,600
预提费用	0	660	0		0	1,000
实收资本		20,000			0	20,000
未分配利润		2,340	0			3,600
	33,700	33,700	43,330		33,700	30,200

职工表							
编号	职工姓名	性别	出生年月	籍贯	民族	工作时间	技术职务
1	李淑玉	女	09-01-69	北京市	汉族	07-01-90	讲师
2	王清照	女	11-01-51	山东济南	汉族	09-01-99	教授
3	辛如虎	男	08-01-54	河北无极	汉族	12-01-73	副教授
4	柳长亭	男	06-01-60	河南光山	回族	01-01-84	副教授
5	张煜	男	07-01-72	福建厦门	汉族	07-01-95	助教
6	周春花	女	01-01-63	上海市	汉族	10-01-85	副教授
7	李甫	男	02-01-68	山东青岛	汉族	06-01-92	助理研究员
8	欧阳太白	男	01-01-47	广西桂林	壮族	08-01-98	研究员

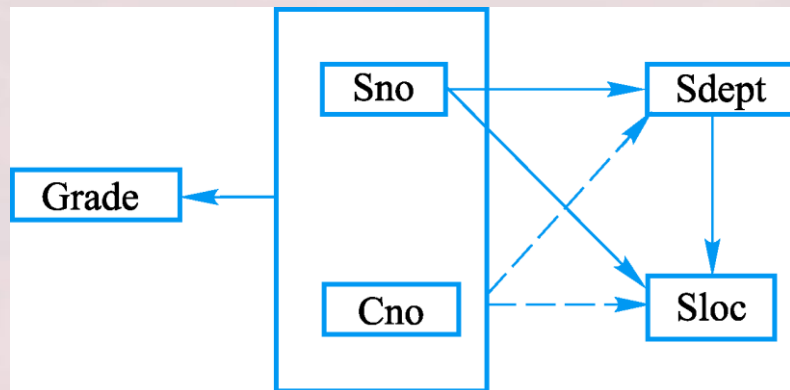
❖ 第一范式是对关系模式的最起码的要求。不满足第一范式的数据库模式不能称为关系数据模式。

6.2.4 第二范式 (2NF)

满足第一范式的关系模式并不一定是一个好的关系模式。

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

Sloc为学生住处，假设每个系的学生住在同一个楼。



$(Sno, Cno) \xrightarrow{F} Grade$
 $Sno \rightarrow Sdept$
 $(Sno, Cno) \xrightarrow{P} Sdept$
 $Sno \rightarrow Sloc$
 $(Sno, Cno) \xrightarrow{P} Sloc$
 $Sdept \rightarrow Sloc$

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



6.2.4 第二范式 (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R的码，则 $R \in 2NF$ 。

■ 例：

$S-L-C(\underline{Sno}, \underline{Cno}, Sdept, Sloc, Grade) \in 1NF$

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

非主属性 $Sdept$ 和 $Sloc$ 部分函数依赖于码 (Sno, Cno) 。



6.2.4 2NF (续)

❖ 一个关系模式R不属于2NF，就会产生问题。

❖ 例如S-L-C存在的问题：(1) 插入异常

假设Sno=2014102, 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。

Sno	Sdept	Sloc	Cno	Grade
2014101	IS	N	3	89
2014101	IS	N	2	97
2014101	IS	N	5	88
2014103	IS	N	1	86
2014103	IS	N	3	92
2014104	IS	N	3	79
2014102	IS	N	null	null

6.2.4 2NF (续)

(2) 删除异常

假定2014104学生只选修了3号课程这一门课。现在因身体不适，他连3号课程也不选修了。因课程号是主属性，此操作将导致该整个元组的删除。这样，2014104学生信息都被删除了。

Sno	Sdept	Sloc	Cno	Grade
2014101	IS	N	3	89
2014101	IS	N	2	97
2014101	IS	N	5	88
2014103	IS	N	1	86
2014103	IS	N	3	92
2014104	IS	N	3	79



6.2.4 2NF (续)

(2) 删除异常

假定2014104学生只选修了3号课程这一门课。现在因身体不适，他连3号课程也不选修了。因课程号是主属性，此操作将导致该整个元组的删除。这样，2014104学生信息都被删除了。

Sno	Sdept	Sloc	Cno	Grade
2014101	IS	N	3	89
2014101	IS	N	2	97
2014101	IS	N	5	88
2014103	IS	N	1	86
2014103	IS	N	3	92



6.2.4 2NF (续)

(3) 数据冗余度大

如果一个学生选修了8门课程，那么他的Sdept和Sloc值就要重复存储了8次。

Sno	Sdept	Sloc	Cno	Grade
2014101	IS	N	3	89
2014101	IS	N	2	97
2014101	IS	N	5	88
2014103	IS	N	1	86
2014103	IS	N	3	92
2014104	IS	N	3	79
2014101	IS	N	1	72
2014101	IS	N	4	65
2014101	IS	N	6	99
2014101	IS	N	7	83
2014101	IS	N	8	75



6.2.4 2NF (续)

(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。

Sno	Sdept	Sloc	Cno	Grade
2014101	IS	S	3	89
2014101	IS	S	2	97
2014101	IS	S	5	88
2014103	IS	N	1	86
2014103	IS	N	3	92
2014104	IS	N	3	79
2014101	IS	S	1	72
2014101	IS	S	4	65
2014101	IS	S	6	99
2014101	IS	S	7	83
2014101	IS	S	8	75



6.2.4 2NF (续)

(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。

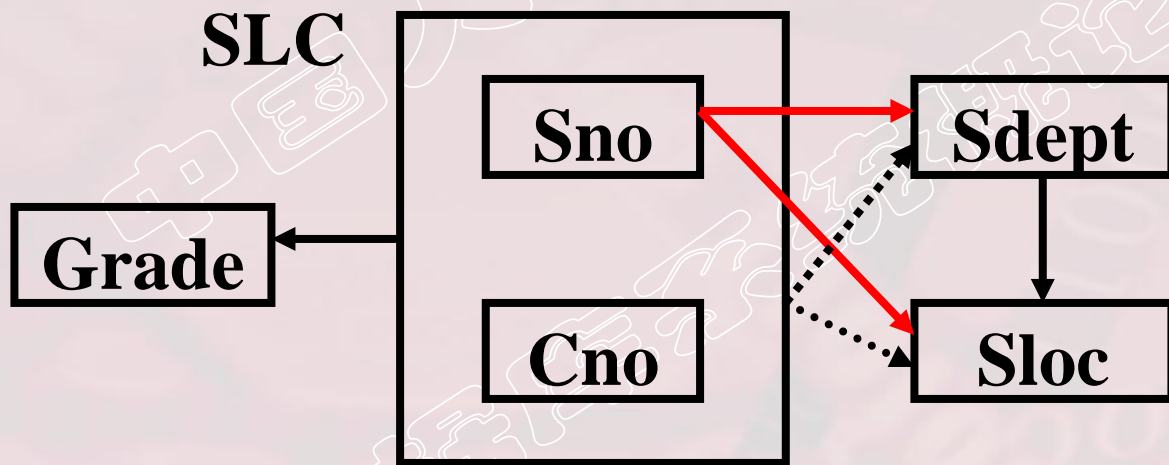
Sno	Sdept	Sloc	Cno	Grade
2014101	IS	S	3	89
2014101	IS	S	2	97
2014101	IS	S	5	88
2014103	IS	N	1	86
2014103	IS	N	3	92
2014104	IS	N	3	79
2014101	IS	S		
2014101	IS	S		
2014101	IS	S		
2014101	IS	S		
2014101	IS	S	8	75

因此，SLC不是一个好的关系模式。

6.2.4 2NF (续)

❖ 原因: SLC(Sno, Sdept, Sloc, Cno, Grade) 中

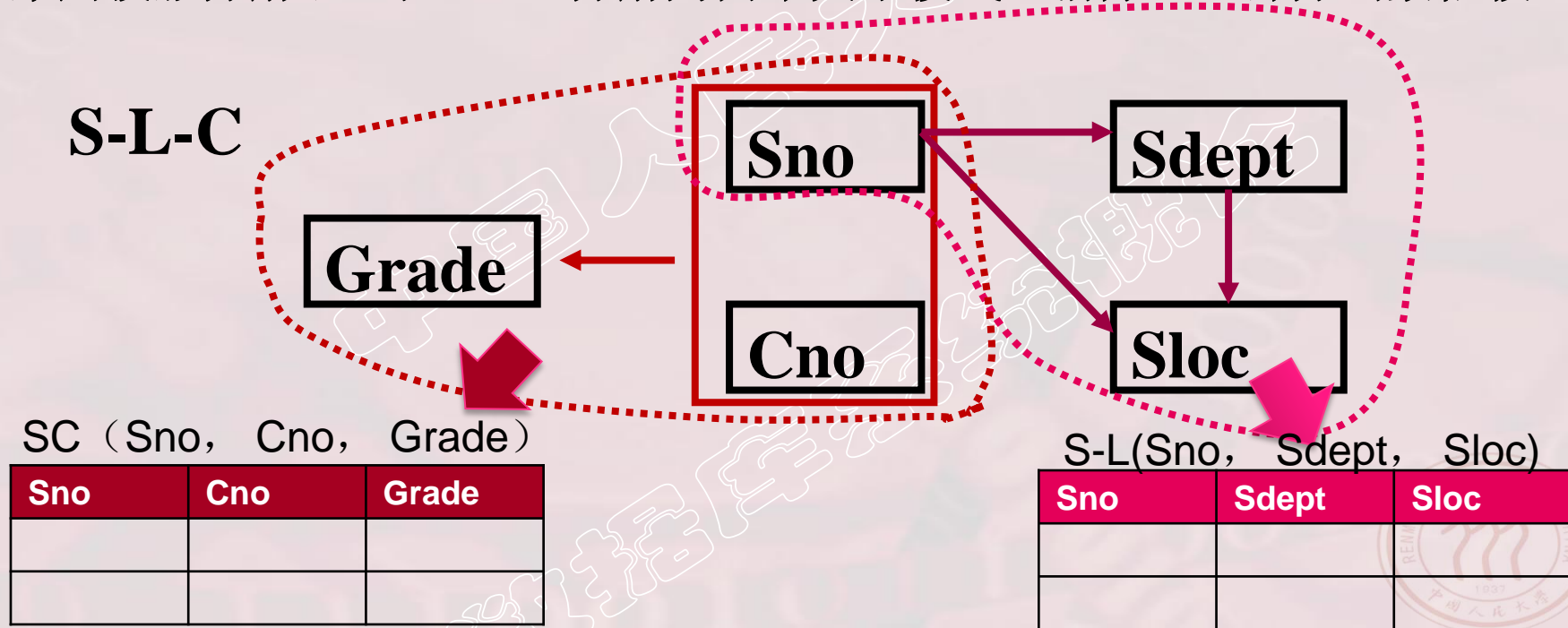
❖ **Sdept**、**Sloc**部分函数依赖于码。SLC的码为(Sno, Cno)



6.2.4 2NF (续)

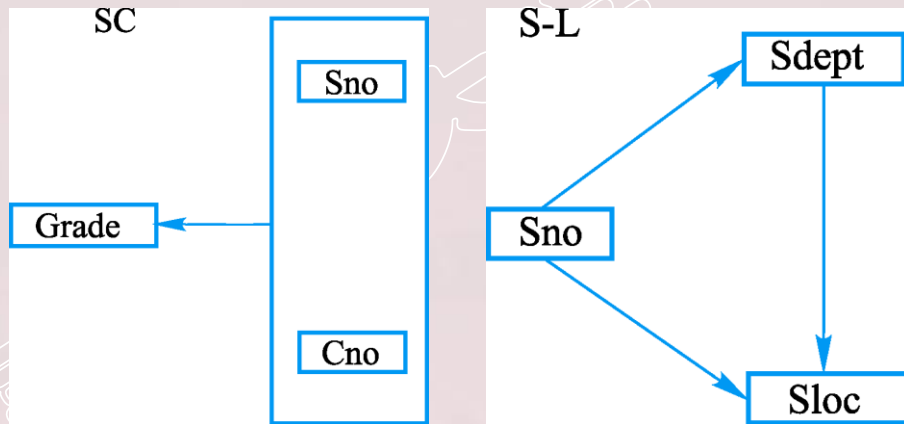
❖ 解决方法

采用投影分解法，把**S-L-C**分解为两个关系模式，消除这些部分函数依赖



6.2.4 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
- 非主属性对码都是完全函数依赖了。他们都是2NF。
- 从而使上述四个问题在一定程度上得到了一定的解决。



6.2.4 2NF (续)

SC

Sno	Cno	Grade

S-L

Sno	Sdept	Sloc

- (1) 由于学生选修课程的情况与学生的基本情况是分开存储在两个关系中的，在**S-L**关系中可以插入尚未选课的学生。
- (2) 删除一个学生的所有选课记录，只是**SC**关系中没有关于该学生的记录了，**S-L**关系中关于该学生的记录不受影响。
- (3) 不论一个学生选多少门课程，他的**Sdept**和**Sloc**值都只存储1次。这就大大降低了数据冗余。
- (4) 学生转系只需修改**S-L**关系中该学生元组的**Sdept**值和**Sloc**值，由于**Sdept**、**Sloc**并未重复存储，因此减化了修改操作。

6.2 规范化

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 第二范式 (2NF)

6.2.5 第三范式 (3NF)

6.2.6 BC范式 (BCNF)

*6.2.7 多值依赖

*6.2.8 第四范式 (4NF)

6.2.9 规范化小结



第六章 关系数据理论



6.2 规范化

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 第二范式 (2NF)

6.2.5 第三范式 (3NF)

6.2.6 BC范式 (BCNF)

*6.2.7 多值依赖

*6.2.8 第四范式 (4NF)

6.2.9 规范化小结



6.2.5 第三范式 (3NF)

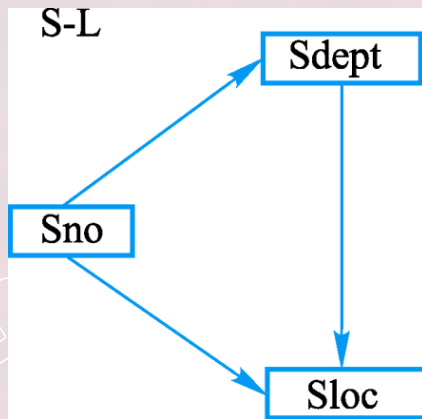
❖ 2NF还有什么问题？

- 采用投影分解法，把S-L-C分解为两个关系模式：SC和S-L，消除了S-L-C中非主属性对码的部分函数依赖。
- 一般地，如果把1NF关系模式通过投影分解方法，消除非主属性对码的部分函数依赖，分解为多个2NF的关系模式。
- 可以在一定程度上减轻原1NF关系模式中存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。
- 但是还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 第三范式 (3NF)

■ 2NF关系模式S-L(Sno, Sdept, Sloc)中
函数依赖:



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

6.2.5 第三范式 (3NF)

S-L关系存在的问题:

(1) 插入异常

如果某个系因种种原因（例如刚刚成立），目前暂时没有在校学生，我们就无法把这个系的信息，如**MA, S**，存入数据库。

Sno	Sdept	Sloc
2014101	IS	N
2014102	IS	N
2014103	IS	N
2014104	IS	N
null	MA	S



6.2.5 第三范式 (3NF)

(2) 删除异常

如果某个系（如**IS**）的学生全部毕业了，我们在删除该系学生信息的同时，把这个系的信息，如**IS, N**，也丢掉了。

Sno	Sdept	Sloc
2014101	IS	N
2014102	IS	N
2014103	IS	N
2014104	IS	N
2014105	PH	S
2014106	PH	S



6.2.5 第三范式 (3NF)

(3) 数据冗余度大

每一个系的学生都住在同一个地方，关于系的住处的信息却重复出现，**重复次数与该系学生人数**相同。

Sno	Sdept	Sloc
2014101	IS	N
2014102	IS	N
2014103	IS	N
2014104	IS	N
2014105	PH	S
2014106	PH	S
2014107	PH	S
2014108	PH	S
.....

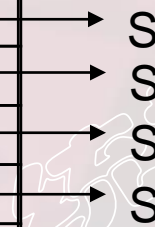


6.2.5 第三范式 (3NF)

(4) 修改复杂

学校调整学生住处时，由于关于每个系的住处信息是重复存储的，修改时**必须同时更新该系所有学生的Sloc**属性值。

Sno	Sdept	Sloc
2014101	IS	S
2014102	IS	S
2014103	IS	S
2014104	IS	S
2014105	PH	S
2014106	PH	S
2014107	PH	S
2014108	PH	S
.....



6.2.5 第三范式 (3NF)

(4) 修改复杂

学校调整学生住处时，由于关于每个系的住处信息是重复存储的，修改时**必须同时更新该系所有学生的Sloc**属性值。

Sno	Sdept	Sloc
2014101	IS	S
2014102	IS	S
2014103	IS	S
2014104	IS	S
2014105	PH	S
2014106	PH	S
2014107	PH	S
2014108	PH	S
.....

所以，
S-L仍不是一个好的关系模式。

6.2.5 第三范式 (3NF)

❖ 原因:

S-L中**Sloc**传递函数依赖于**Sno**,
即: 非主属性传递函数依赖码

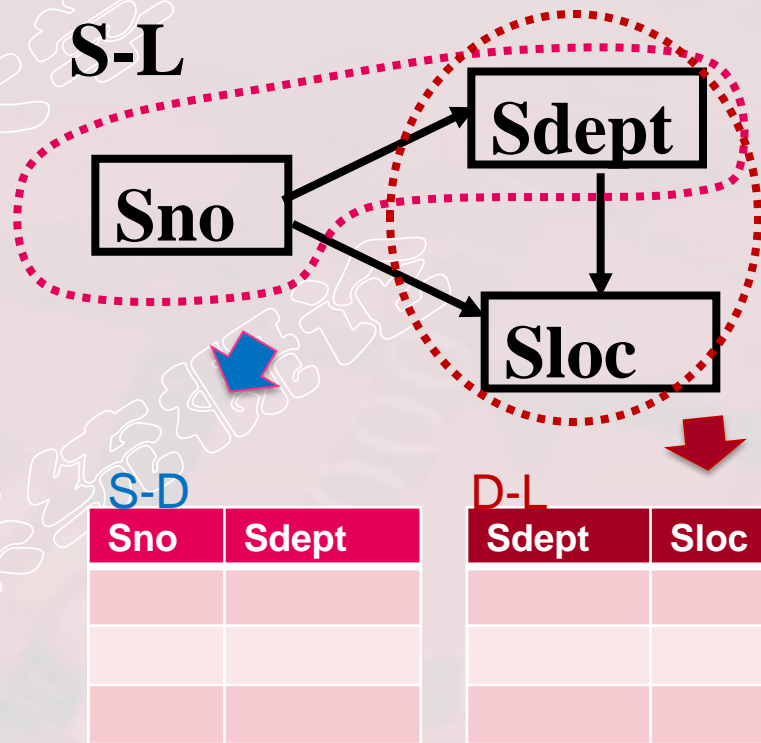
❖ 解决方法

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**



6.2.5 第三范式 (3NF)

S-D	Sno	Sdept	D-L	Sdept	Sloc

➤ 异常的情况得到改善:

(1) D-L关系中可以插入系的信息，即使还没有在校学生。

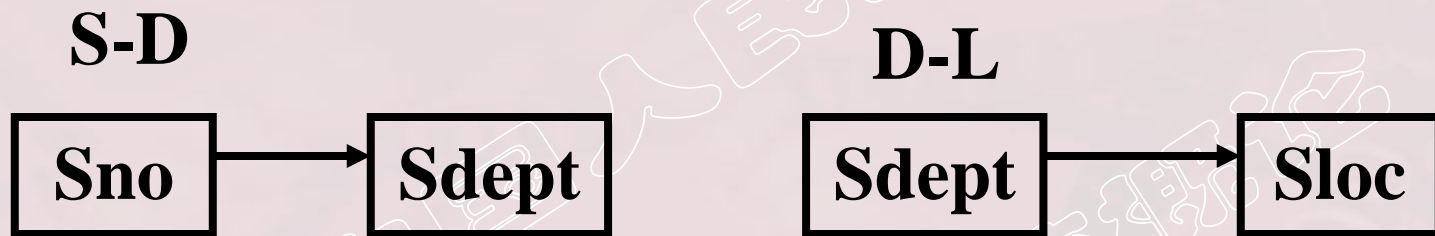
(2) 某个系的学生全部毕业了，只是删除S-D关系中的相应元组，D-L关系中关于该系的信息仍存在。

(3) 关于系的住处的信息只在D-L关系中存储一次。

(4) 当学校调整某个系的学生住处时，只需修改D-L关系中一个元组的Sloc属性值。

6.2.5 第三范式 (3NF)

- S-D的码为Sno, D-L的码为Sdept.



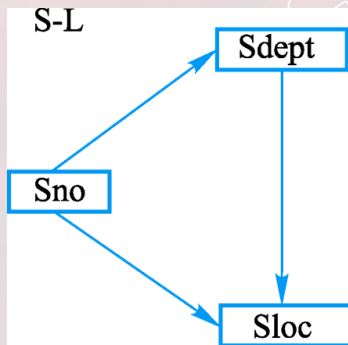
在分解后的关系模式中既没有非主属性对码的部分函数依赖，也没有非主属性对码的传递函数依赖，进一步解决了上述四个问题。

6.2.5 第三范式 (3NF)

❖ 3NF的定义

定义6.7 关系模式 $R\langle U, F \rangle \in 1NF$, 若 R 中不存在这样的码 X 、属性组 Y 及非主属性 Z ($Y \not\rightarrow Z$), 使得 $X \rightarrow Y$, $Y \rightarrow Z$, $Y \twoheadrightarrow X$, 成立, 则称 $R\langle U, F \rangle \in 3NF$ 。

例: S-D ($Sno, Sdept$) $\in 3NF$
D-L ($Sdept, Sloc$) $\in 3NF$



S-L 不存在部分函数依赖, 但是存在传递函数, 所以
 $S-L(Sno, Sdept, Sloc) \in 2NF$
 $S-L(Sno, Sdept, Sloc) \notin 3NF$

6.2.5 第三范式（续）

❖ 3NF的一些性质：

- 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 $R \in 2NF$ 。
- 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 第三范式

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 第二范式 (2NF)

6.2.5 第三范式 (3NF)

6.2.6 BC范式 (BCNF)

*6.2.7 多值依赖

*6.2.8 第四范式 (4NF)

6.2.9 规范化小结



第六章 关系数据理论

