

练习

关系模式 $R(A,B,C,D,E)$ 的函数依赖有 $A,B \rightarrow C$ 和 $C,D \rightarrow E$. 如果 A, B 和 D 每个最多有三个不同的取值, 则 E 最多有多少个不同的取值?

A. 27

B. 9

C. 3

D. 81

数据库原理

第6章 关系数据库设计理论

辽东学院 鲁 琴

本节要点

数据库基础概念

数据库原理

关系数据库

数据库新技术

关系数据模型

关系数据语言

数据库设计

数据库管理

数据库设计步骤

概念设计工具

ER图

UML图

逻辑设计工具

关系数据库设计理论

数据依赖

1NF

2NF

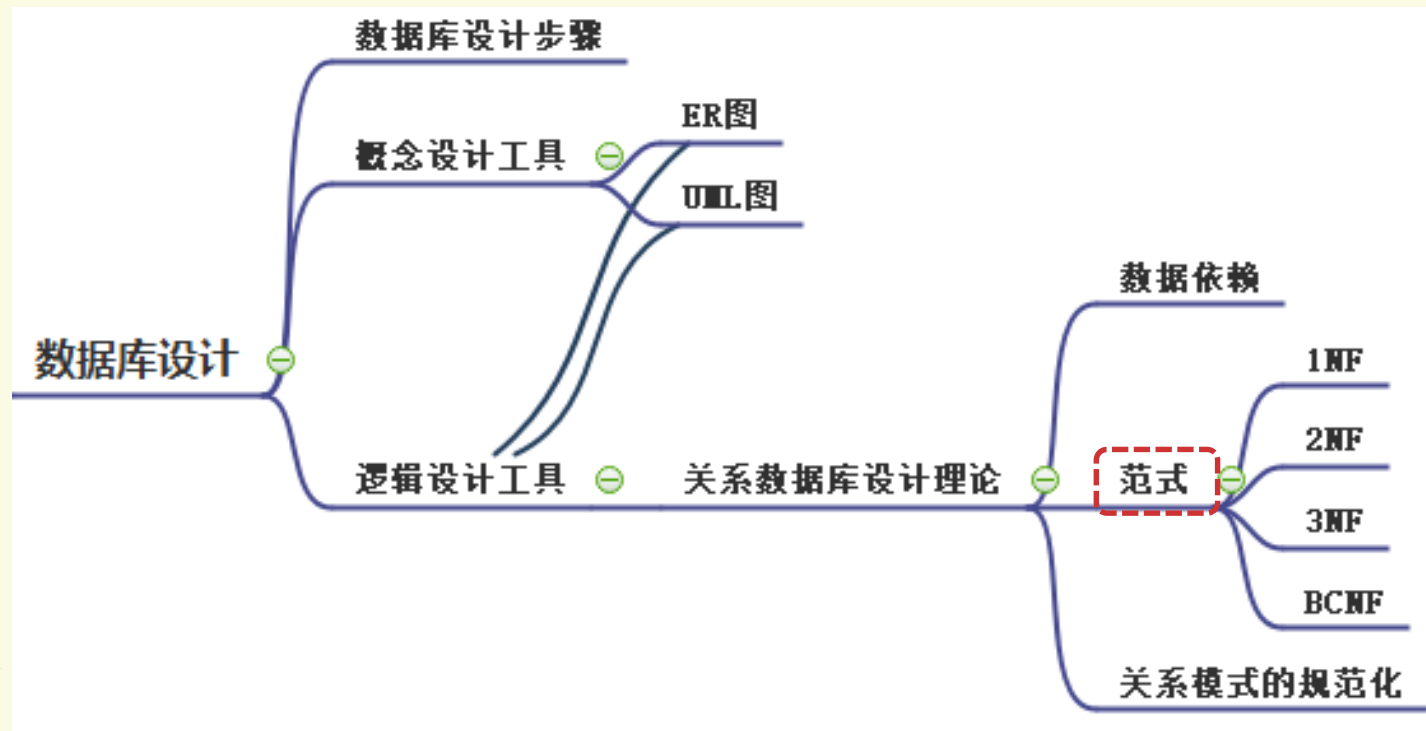
3NF

BCNF

关系模式的规范化

范式

第6章 关系数据库设计理论



第5章 关系数据库设计理论

6.1 数据依赖

6.2 范式

6.2.1 第一范式 (1NF)

6.2.2 第二范式 (2NF)

6.2.3 第三范式 (3NF)

6.2.4 BC范式 (BCNF)

6.3 关系模式的规范化

6.2 范式

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**

范式是符合某一种级别的关系模式的集合

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)

各种范式之间关系：

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

某一关系模式R为第n范式，可简记为 $R \in nNF$

6.2.1 第一范式 (1NF)

定义6.6 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基本的要求

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

1NF

码是什么？

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

其中各属性分别为学号, 所在系, 宿舍楼, 课程号和成绩。

规定:

一个学生就属于一个系, 每个系的学生住在同一个宿舍楼,
一个学生学的每一门课程都有唯一的成績。

写出所有的函数依赖:

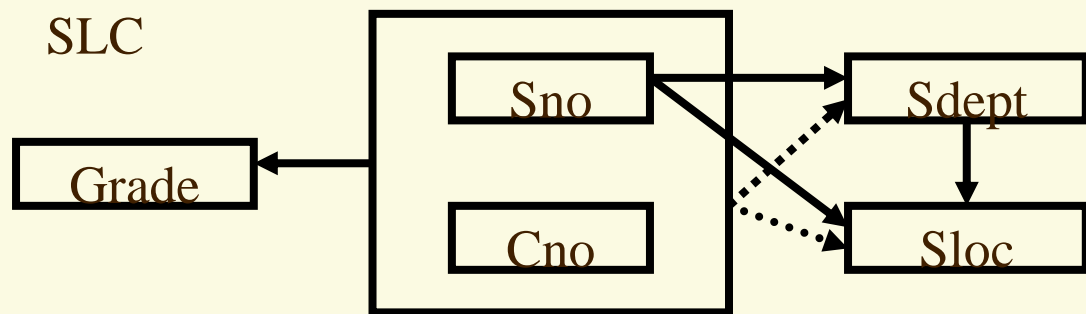
$Sno \rightarrow Sdept, Sdept \rightarrow Sloc, (Sno, Cno) \rightarrow Grade$

$(Sno, Cno) \rightarrow Sno, (Sno, Cno) \rightarrow Sdept$

$(Sno, Cno) \rightarrow Sloc, (Sno, Cno) \rightarrow Cno$

部分
函数依赖

SLC的函数依赖图



SLC的码为(Sno, Cno)

1. SLC满足第一范式
2. 非主属性Sdept和Sloc部分函数依赖于码(Sno, Cno)

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

SLC存在的问题

SLC不是一个好的关系模式

(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

(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, 此操作将导致该学生信息的整个元组都要删除

(3) 数据冗余度大

如果一个学生选修了10门课程, 那么他的Sdept和Sloc值就要重复存储了10次

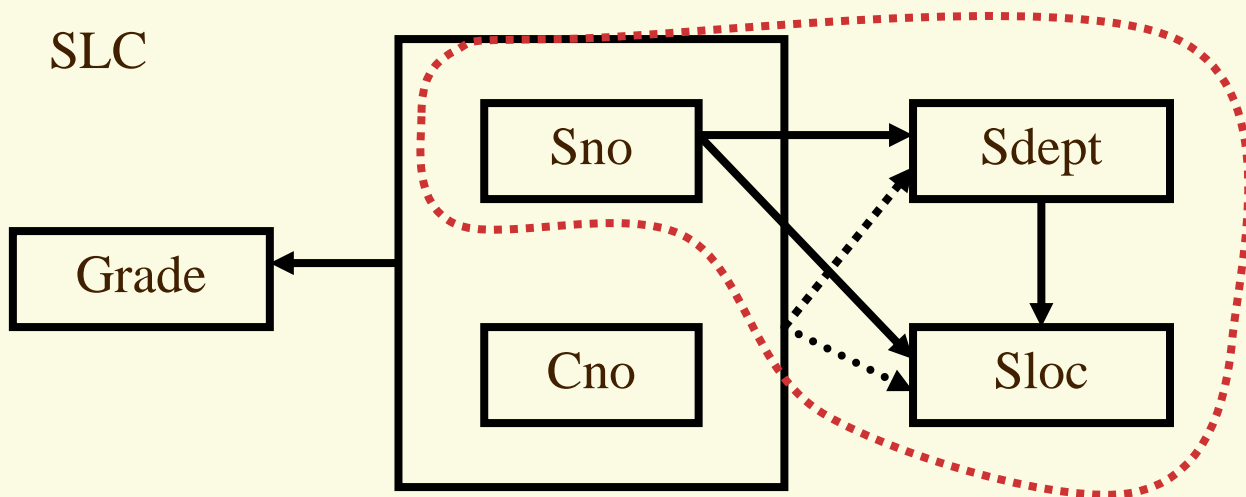
(4) 修改复杂

如学生转系, 在修改此学生元组的Sdept值的同时, 还可能需修改住处(Sloc) 如果这个学生选修了10门课, 则必须无遗漏地修改10个元组中全部Sdept、Sloc信息

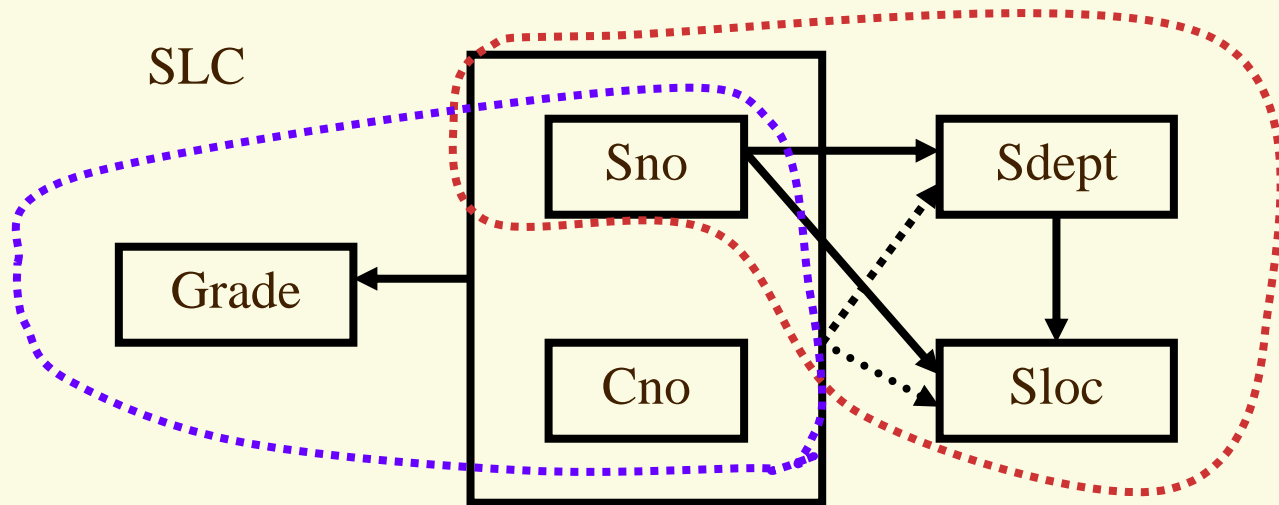
SLC不是一个好的关系模式

原因

Sdept、**Sloc**部分函数依赖于码。



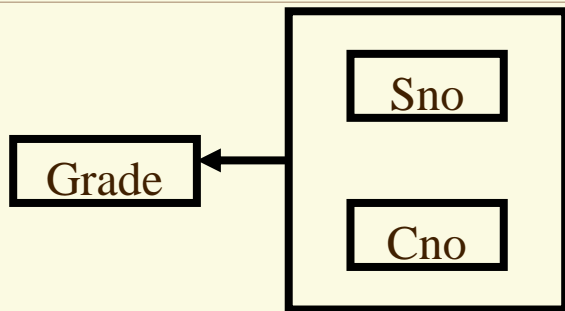
解决方法：分解



SC (Sno, Cno, Grade)

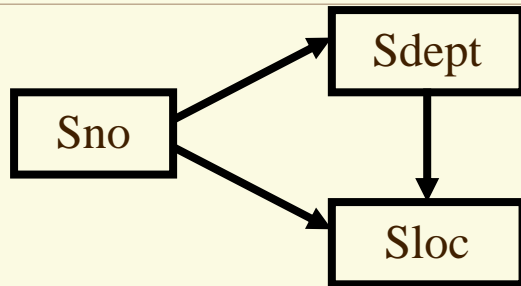
SL (Sno, Sdept, Sloc)

函数依赖图



SC (Sno, Cno, Grade)

解决了什么问题?



SL (Sno, Sdept, Sloc)

插入异常?

删除异常?

数据冗余度大?

修改复杂?

解决的问题

- (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SL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，SL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SL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 范式

6.2.1 第一范式 (1NF)

6.2.2 第二范式 (2NF)

6.2.3 第三范式 (3NF)

6.2.4 BC范式 (BCNF)

6.2.2 第二范式 (2NF)

定义6.7 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：SLC(Sno, Sdept, Sloc, Cno, Grade) $\in 1NF$

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

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

第二范式（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。

SL的问题

2NF关系模式SL(Sno, Sdept, Sloc)中

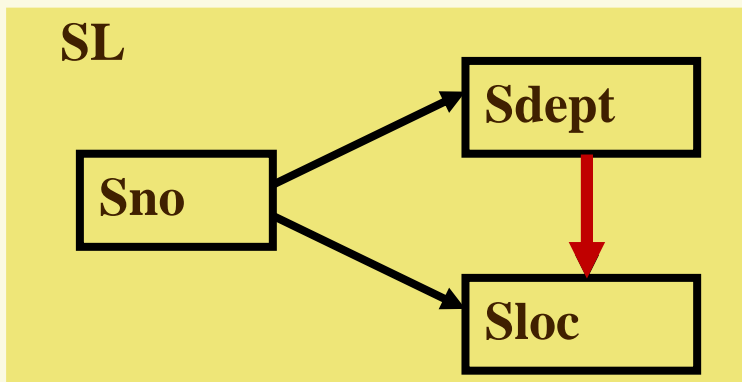
函数依赖:

$Sno \rightarrow Sdept$

$Sno \rightarrow Sloc$

$Sdept \rightarrow Sloc$

SL的码是**Sno**



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**。

SL仍不是一个好的关系模式

(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存入数据库。

(2) 删除异常

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，把这个系的信息也丢掉了。

(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。

(4) 修改复杂

当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。

第二范式 (续)

原因

Sloc传递函数依赖于Sno

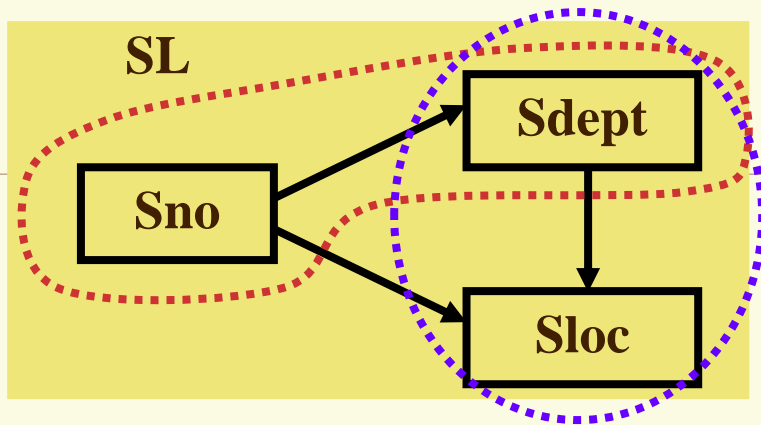
解决方法

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)

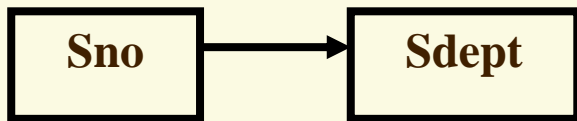
DL (Sdept, Sloc)

SD的码为Sno， DL的码为Sdept。



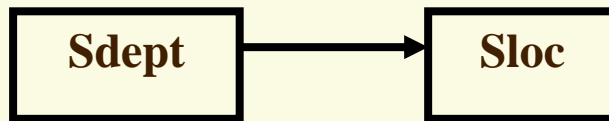
第二范式（续）

SD的码为**Sno**



SD

DL的码为**Sdept**



DL

解决问题

- (1) DL关系中可以**插入**无在校学生的系的信息。
- (2) 某个系的学生全部毕业了，只是**删除**SD关系中的相应元组，DL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DL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**DL关系中一个相应元组的Sloc属性值。

6.2 范式

6.2.1 第一范式 (1NF)

6.2.2 第二范式 (2NF)

6.2.3 第三范式 (3NF)

6.2.4 BC范式 (BCNF)

6.2.1 第一范式 (1NF)

定义6.6 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**最基本**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。

6.2.2 第二范式 (2NF)

定义6.7 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：SLC(Sno, Sdept, Sloc, Cno, Grade) $\in 1NF$

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

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

定义6.8 第三范式 (3NF)

关系模式 $R \in 1NF$,且不存在非主属性对码的传递函数依赖, 则称 $R \in 3NF$ 。

例, $SL(Sno, Sdept, Sloc) \in 2NF$

$SD(Sno, Sdept) \in 3NF$

$DL(Sdept, Sloc) \in 3NF$

6.2 规范化

6.2.1 第一范式 (1NF)

6.2.2 第二范式 (2NF)

6.2.3 第三范式 (3NF)

6.2.4 BC范式 (BCNF)

第三范式（续）

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 $2NF$ 。

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，在有些情况下，不能完全消除关系模式中的各种异常情况和数据冗余。

6.2.4 BC范式 (BCNF)

BCNF (Boyce Codd Normal Form) 是由Boyce和Codd提出的, 比3NF更进了一步。通常认为**BCNF是修正的第三范式**。

定义6.9 设关系模式 $R \in 1NF$, 如果对于 R 的每个函数依赖 $X \rightarrow Y$, 若 Y 不属于 X , 则 X 必含有候选码, 那么 $R \in BCNF$ 。换句话说, 在关系模式 R 中, 如果每一个**决定属性集**都包含候选码, 则 $R \in BCNF$ 。

BC范式（续）

采用投影分解法将一个3NF的关系分解为多个BCNF的关系，可以进一步解决原3NF关系中可能存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。

BCNF的关系模式所具有的性质

1. 所有非主属性都完全函数依赖于每个候选码
2. 所有主属性都完全函数依赖于每个不包含它的候选码
3. 没有任何属性完全函数依赖于非码的任何一组属性

BC范式（续）

3NF与BCNF的关系

- 如果关系模式 $R \in \text{BCNF}$ ，必定有 $R \in \text{3NF}$ 。
- 如果 $R \in \text{3NF}$ ，且 R 只有一个候选码，则 R 必属于 BCNF 。

如果一个关系数据库中的所有关系模式都属于 BCNF ，那么在函数依赖范畴内，它已实现了模式的彻底分解，达到了最高的规范化程度，消除了插入异常和删除异常。

BC范式

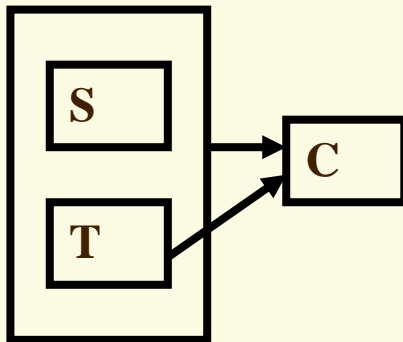
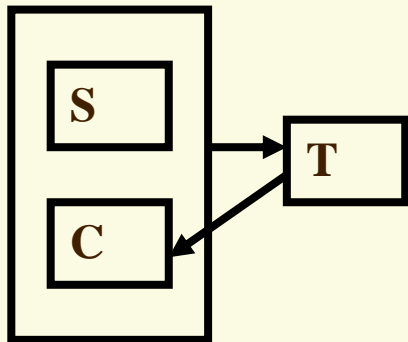
在关系模式STC (S, T, C) 中, S表示学生, T表示教师, C表示课程。

函数依赖:

- ✓ 假设每一教师只教一门课,每门课由若干教师教.
- ✓ 但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。于是有:

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

STC函数依赖图



STC

STC \in 3NF

(S, C)和(S, T)都可以作为候选码。

STC \in 3NF

T	C	S
T1	C1	S1
T1	C1	S2
T2	C1	S3
T2	C1	S4
T3	C2	S2
T4	C2	S2
T4	C3	S2

有什么问题？

STC存在的问题

$STC \in 3NF$ ，但它仍不是一个理想的关系模式

(1) 插入异常

如果某个教师开设了某门课程，但尚未有学生选修，则有关信息也无法存入数据库中

(2) 删除异常

如果选修过某门课程的学生全部毕业了，在删除这些学生元组的同时，相应教师开设该门课程的信息也同时丢掉了

(3) 数据冗余度大

虽然一个教师只教一门课，但每个选修该教师该门课程的学生元组都要记录这一信息

(4) 修改复杂

某个教师开设的某门课程改名后，所有选修了该教师该门课程的学生元组都要进行相应修改

原因

$T \rightarrow C$ ，即主属性C依赖于T，即主属性C部分依赖于码(S, T)。

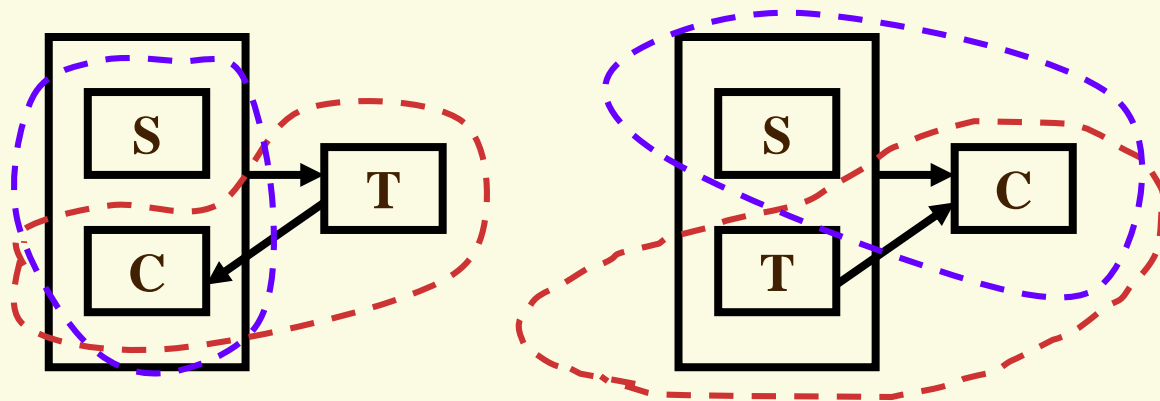
解决方法

采用投影分解法，将STC分解为二个关系模式：

SC(S, C)

TC(T, C)

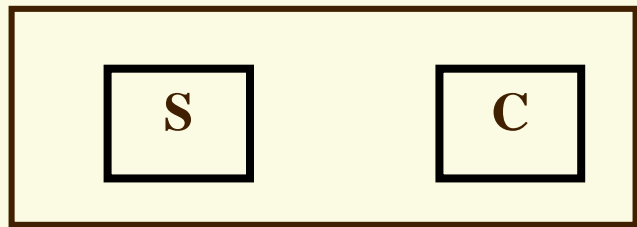
STC函数依赖图



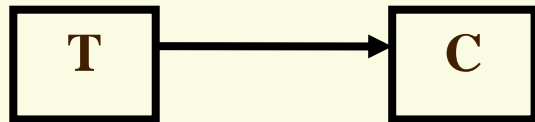
STC

BC范式

SC的码为 (S, C)，TC的码为T。



ST



TC

BCNF

在分解后的关系模式中没有任何属性对码的部分函数依赖和传递函数依赖。它解决了上述四个问题：

- (1) TC关系中可以存储所开课程尚未有学生选修的教师信息。
- (2) 选修过某门课程的学生全部毕业了，只是删除SC关系中的相应元组，不会影响TC关系中相应教师开设该门课程的信息。
- (3) 关于每个教师开设课程的信息只在TC关系中存储一次。
- (4) 某个教师开设的某门课程改名后，只需修改TC关系中的一个相应元组即可。

BCNF

如果一个关系数据库中所有关系模式都属于3NF，则已在很大程度上消除了插入异常和删除异常，但由于可能存在主属性对候选键的部分依赖和传递依赖，因此关系模式的分离仍不够彻底。

如果一个关系数据库中所有关系模式都属于BCNF，那么在函数依赖的范畴内，已经实现了模式的彻底分解，消除了产生插入异常和删除异常的根源，而且数据冗余也减少到极小程度。

练习

如果一个数据库包括学生选课信息，学生包括学生ID(唯一)和姓名 (name)，课程包括课程号(courseNum,唯一)和课程名 (title)，学生在一定的时间(year)选修课程会得到一个分数(grade)，下面哪个数据库模式设计合理？

- A. Took(sID, name, courseNum, title, year, grade)
- B. Course(courseNum, title, year), Took(sID, courseNum, grade)
- C. Student(sID, name), Course(courseNum, title), Took(sID, courseNum, year, grade)
- D. Student(sID, name), Course(courseNum, title), Took(name, title, year, grade)

本节小结

