



⑥ 第六章 关系数据理论

Principles of Database Systems





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

6.1 问题的提出

- ❖ 所有数据库应用系统设计的共同问题—— 如何把复杂的现实世界表达成合适的数据库模式？
- ❖ 具体到关系数据库应用系统中，就表现为下列问题：
 - ❖ 数据库中应该构造几个关系？
 - ❖ 每个关系又应该包括哪些属性？
- 这些问题属于数据库逻辑设计的问题。
- ❖ 本章讲述关系数据库设计理论，它是数据库逻辑设计的有力工具。主要包括三个方面的内容：
 - ❖ 数据依赖
 - ❖ 范式
 - ❖ 模式设计方法

6.1 问题的提出

- 一个完整的关系模式用五元组表示： $R(U, D, DOM, F)$ ， 其中：
 - R: 关系名称
 - U: 一组属性
 - D: U中属性所来自的域
 - DOM: 属性到域的映射
 - F: 属性组U上的一组数据依赖
- D和DOM对关系模式设计影响不大，起主要作用的是U和F，因此简化的关系模式表示为3元组： $R<U, F>$

表示：当且仅当U上的一个关系r满足F时，r称为关系模式 $R<U, F>$ 上的一个关系。
- 有时候，更简写成 $R(U)$ ，即 $R(A_1, A_2, \dots, A_n)$ 的形式

6.1 问题的提出

- **规范化理论**主要包括三个方面的内容：
 - **数据依赖**
 - **范式 (Normal Form)**
 - **模式设计**
- 其中，
 - **数据依赖**是一个关系内部属性与属性之间的一种约束关系。这种约束关系通过数据值的相等与否体现出数据间的相关关系。
 - **函数依赖 (Functional Dependancy, FD)** 和
 - **多值依赖 (Multi-Valued Depandancy, MVD)**
 - **函数依赖**起着核心的作用，是模式分解和模式设计的基础，范式是模式分解的标准。



6.1 问题的提出

- 关系模式的存储异常问题：
 - 数据库的逻辑设计为什么要遵循一定的规范化理论？
 - 什么是好的关系模式？
 - 某些不好的关系模式可能导致哪些问题？

6.1 问题的提出

函数依赖

—— 属性间的一种依赖关系。

设有关系SCD:

学号	姓名	年龄	所在系	系主任	课程号	成绩
S1	赵亦	18	计算机	刘伟	C1	90
S1	赵亦	18	计算机	刘伟	C2	82
S2	钱尔	19	信息	王平	C2	78
S2	钱尔	19	信息	王平	C4	98
S4	李思	21	自动化	陈浩	C6	NULL

语义定义如下:

- 一个系有若干名学生, 但一个学生只属于一个系;
- 一个系只有一名系主任;
- 一个学生可以选修多门功课, 每门课程可有若干学生选修;
- 每个学生学习课程有一个成绩。

思考: 关系模式SCD中存在哪些函数依赖?

学号 \rightarrow 姓名, 学号 \rightarrow 年龄, 学号 \rightarrow 所在系,

所在系 \rightarrow 系主任,

(学号, 课程号) \rightarrow 成绩

6.1 问题的提出

❖ 问题1：数据冗余

- **例如：**每个系名和系主任的名字存储的次数等于该系的学生人数乘以每个学生选修的课程门数，同时学生的姓名、年龄也都要重复存储多次，数据的冗余度很大。

学号	姓名	年龄	所在系	系主任	课程号	成绩
S1	赵亦	18	计算机	刘伟	C1	90
S1	赵亦	18	计算机	刘伟	C2	82
S2	钱尔	19	信息	王平	C2	78
S2	钱尔	19	信息	王平	C4	98
S4	李思	21	自动化	陈浩	C6	NULL

❖ 问题2：更新异常(update anomalies)

- **例如：**如果某系更换系主任，则属于该系的所有学生记录都要进行修改。

6.1 问题的提出

学号	姓名	年龄	所在系	系主任	课程号	成绩
S1	赵亦	18	计算机	刘伟	C1	90
S1	赵亦	18	计算机	刘伟	C2	82
S2	钱尔	19	信息	王平	C2	78
S2	钱尔	19	信息	王平	C4	98
S4	李思	21	自动化	陈浩	C6	NULL

❖ 问题3：插入异常(insertion anomalies)

- **例如：**此关系的码为(学号, 课程号)，依据实体完整性约束，缺少其中一部分时，元组不能插入。因此，如果某个学生没有选课，则这名学生的信息无法插入到数据库中。

❖ 问题4：删除异常(deletion anomalies)

- **例如：**某系学生全部毕业时，要删除全部学生的记录，则系名、系主任也随之删除。虽然这个系依然存在，但在数据库中却无法找到该系的信息。

6.1 问题的提出

产生上述问题的原因，直观地说，是因为关系SCD中“包罗万象”，内容太杂了。

- 解决方法:分解。

- 学生关系: **S(学号, 姓名, 年龄, 所在系)**

- 选课关系: **SC(学号, 课程号, 成绩)**

- 系关系: **D(系名, 系主任)**

- 与SCD相比，分解为三个关系模式后:

- 数据的冗余度明显降低;

- 如果某系更换系主任，只需修改关系D中的一条记录;

- 如果某个学生尚未选课，只要在S中添加一条学生记录，与选课关系无关，避免了插入异常;

- 当一个系的学生全部毕业时，只需在S中删除该系的全部学生记录，而D中该系的信息仍然保留，避免了删除异常。

6.1 问题的提出

❖ 一个好的关系模式应该具备以下**四个**条件：

❖ 尽可能少的数据冗余

❖ 没有更新异常

❖ 没有插入异常

❖ 没有删除异常

❖ 产生关系数据库中这些异常问题的根本原因：

❖ 关系模式中存在不合适的属性间联系

❖ 解决问题的方法：

❖ 采用**分解**策略消去不合适联系

学号	姓名	年龄	所在系	系主任	课程号	成绩
S1	赵亦	18	计算机	刘伟	C1	90
S1	赵亦	18	计算机	刘伟	C2	82
S2	钱尔	19	信息	王平	C2	78
S2	钱尔	19	信息	王平	C4	98
S4	李思	21	自动化	陈浩	C6	NULL

6.1 问题的提出

- 分解后的关系模式是“好”的。
- 不过，一个好的关系模式**并不是**在任何情况下都是最优的。
- 如何按照一定的规范设计关系模式，将结构复杂的关系**分解**成结构简单的关系，从而把“不好”的关系数据库模式转变为“好”的关系数据库模式，这就是**关系的规范化**。
- 规范化又可以根据不同的要求而分成若干级别。
- 数据库模式的好坏和关系中各属性间的依赖关系有关，因此，我们先讨论属性间的依赖关系，然后再讨论关系规范化理论。

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

❖ 一个好的关系模式应该具备以下四个条件：

- ❖ 尽可能少的数据冗余
- ❖ 没有更新异常
- ❖ 没有插入异常
- ❖ 没有删除异常

6.2.1 函数依赖

定义6.1 设 $R(U)$ 是一个属性集 U 上的关系模式, $X, Y \subseteq U$, r 是 $R(U)$ 上的任意一个关系, 如果成立

对 $\forall t, s \in r$, 若 $t[X] = s[X]$, 则 $t[Y] = s[Y]$

则称 “ **X 函数确定 Y** ” 或 “ **Y 函数依赖于 X** ”, 记作 $X \rightarrow Y$ 。(或不存在 t, s , $t[X] = s[X]$ 而 $t[Y] \neq s[Y]$,)

X 称为这个函数依赖的**决定属性集**(Determinant)。

$$Y = f(x)$$

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

6.2.1 函数依赖

有关函数依赖的几点说明:

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

只能根据语义来确定一个函数依赖, 而不能按照其形式化定义来证明一个函数依赖是否成立。

例: 对于关系模式S, 当学生不存在重名的情况下, 可以得到:

姓名→年龄 或 姓名→性别。

函数依赖反映了一种语义完整性约束。

6.2.1 函数依赖

有关函数依赖的几点说明：

3) 函数依赖关系的存在**与时间无关**。

4) 属性间联系与函数依赖的对应关系

- **1:1联系**：存在函数依赖 $X \rightarrow Y$ ，且 $Y \rightarrow X$ ，即 $X \leftrightarrow Y$ 。
- **1:m联系**：存在函数依赖 $X \rightarrow Y$ ，但不存在 $Y \rightarrow X$ 。
- **m:n联系**：不存在函数依赖。

例：关系模式SCD中存在的函数依赖——属性间的联系：

- 学号 \rightarrow 姓名，学号 \rightarrow 年龄，学号 \rightarrow 所在系，1:m
- 所在系 \rightarrow 系主任，1:1
- (学号, 课程号) \rightarrow 成绩，学号和课程之间m:n

6.2.1 函数依赖

2. 平凡函数依赖与非平凡函数依赖

- 平凡函数依赖

设有关系模式 $R(U, F)$, X 、 Y 都是 U 的子集, 若 $X \rightarrow Y$, $Y \subseteq X$, 则称 $X \rightarrow Y$ 是平凡函数依赖。

- 非平凡函数依赖

设有关系模式 $R(U, F)$, X 、 Y 是 U 的子集, 若 $X \rightarrow Y$, 但 $Y \not\subseteq X$, 则称 $X \rightarrow Y$ 是非平凡函数依赖。

说明: 对于任何 R , 平凡函数依赖总是成立的。

若不特别声明, 我们总是讨论非平凡的函数依赖。

6.2.1 函数依赖

3. 完全函数依赖与部分函数依赖

定义6.2 设关系模式 $R(U)$, U 是属性全集, X 和 Y 是 U 的子集:

- 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 **完全函数依赖于** X , 记作 $X \xrightarrow{F} Y$ 。
- 如果对 X 的某个真子集 X' , 有 $X' \rightarrow Y$, 则称 Y **部分函数依赖于** X , 记作 $X \xrightarrow{P} Y$ 。

例:

在SCD中:

(学号, 课程号) \xrightarrow{F} 成绩。

而学号 \rightarrow 年龄, 所以 (学号, 课程号) \xrightarrow{P} 年龄。

6.2.1 函数依赖

4. 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 $Y \not\subseteq X$, $Y \not\rightarrow X$, $Z \not\subseteq Y$, 则称 Z 传递函数依赖于 X 。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: 在前述关系模式SCD中, 有:

$Sno \rightarrow Sdept$, $Sdept \rightarrow Mname$

系主任 $Mname$ 传递函数依赖于学号 Sno 。

6.2.2 码

定义6.4 设K为关系模式 $R\langle U, F \rangle$ 中的属性或属性组合。若 $K \xrightarrow{F} U$ ，则K称为R的一个**候选码** (Candidate Key)。若关系模式R有多个候选码，则选定其中的一个做为**主码** (Primary key)。

- ❖ 最简单情况：单个属性为码，称为**单属性码**；
- ❖ 一般情况：码为多个属性，称为**多属性码**；
- ❖ 最极端情况：全部属性为码，称为**全码**。
- **主码**—— 任意选候选码之一。
- **主属性**—— 包含在任何一个候选码中的属性。
- **非主属性**—— 不包含在任何一个候选码中的属性。
- **超码**—— 若 $K \xrightarrow{P} U$ ，则K称为超码。

6.2.2 码

定义6.6 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。

- **主码和外部码一起提供了表示关系间联系的手段。**

6.2.3 范式

- **规范化的基本思路**——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。
- 这就要求设计出来的关系模式要**满足一定的条件**。我们把关系数据库的规范化过程中为不同程度的规范化要求设立的不同标准称为**范式** (Normal Form) 。
- 满足最基本规范化要求的关系模式叫**第一范式**，
- 在第一范式中进一步满足一些要求为**第二范式**，
- 通过**模式分解**将一个低级范式转换为若干个高级范式的过程称作**规范化**。

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

第一范式 1NF

- 如果关系模式R，其所有的属性均为简单属性，即每个属性都是不可再分的，则称R属于第一范式，简称**1NF**，记作 **$R \in 1NF$** 。
- 第一范式是最基本的规范形式，我们把满足这个条件的关系称为**规范化关系**。

- 例如：**以下关系是一个非规范化的关系。

XH	KM	CJ		
		test1	test2	test3
001	OS	80	78	90
002		85	82	95

S#	C#
S1	{C1, C2, C3}
S2	{C2,C4}

- 关系DBS中，所有关系模式都必须是1NF。

■ 将非1NF转换为1NF的方法:

① 去掉嵌套属性上层

② 重写行交叉处的值

Sno	Cno	test1	test2	test3
001	OS	80	78	90
002	OS	85	82	95

Sno	Cno
S1	C1
S1	C2
S1	C3
S2	C2
S2	C4

例: **SCD**(学号, 姓名, 年龄, 所在系, 系主任, 课程号, 成绩),
SCD \in 1NF

■ 1NF存在的问题:

数据冗余, 具有插入异常、删除异常、更新复杂。

■ 问题存在原因:

——非主属性部分函数依赖于候选码。

■ 例：SCD(学号, 姓名, 年龄, 所在系, 系主任, 课程号, 成绩)

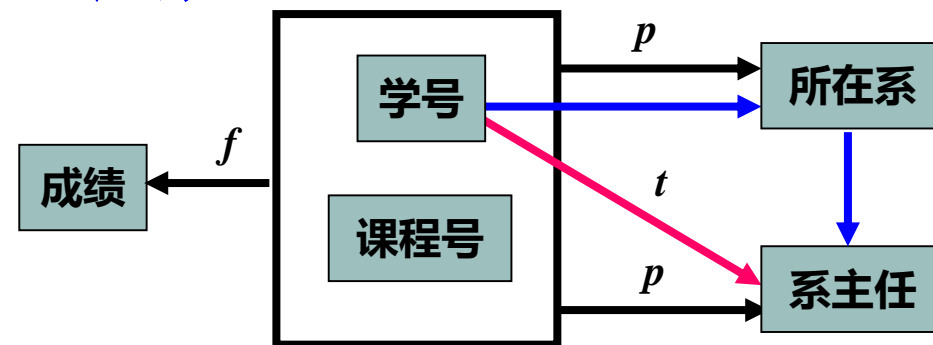
(学号, 课程号) \rightarrow 成绩, 学号 \rightarrow 姓名, 学号 \rightarrow 年龄,
学号 \rightarrow 所在系,

(学号, 课程号) \xrightarrow{p} 年龄

(学号, 课程号) \xrightarrow{p} 所在系,

(学号, 课程号) \xrightarrow{p} 姓名,

学号 \xrightarrow{t} 系主任, (学号, 课程号) \xrightarrow{p} 系主任



■ 解决方法：规范化(投影分解)

消除非主属性对码的部分FD。

- 所有完全FD于码的属性组成一个关系模式。
- 所有部分FD于码的属性组成一个关系模式。

6.2.4 2NF

- 如果关系模式 $R \in 1NF$ ，且每个非主属性都完全函数依赖于R的码，则称R属于**第二范式**（简称**2NF**），记作 $R \in 2NF$ 。

例：

SD

学号 Sno	姓名 Sname	年龄 Age	系别 Dept	系主任 Mname
S1	李勇	18	计算机	王平
S2	刘晨	17	计算机	王平
S3	王敏	18	自控系	刘伟

解决了哪些问题？

SC

学号 Sno	课程号 Cno	成绩 Score
S1	C1	85
S1	C2	82
S2	C4	78
S3	C2	87

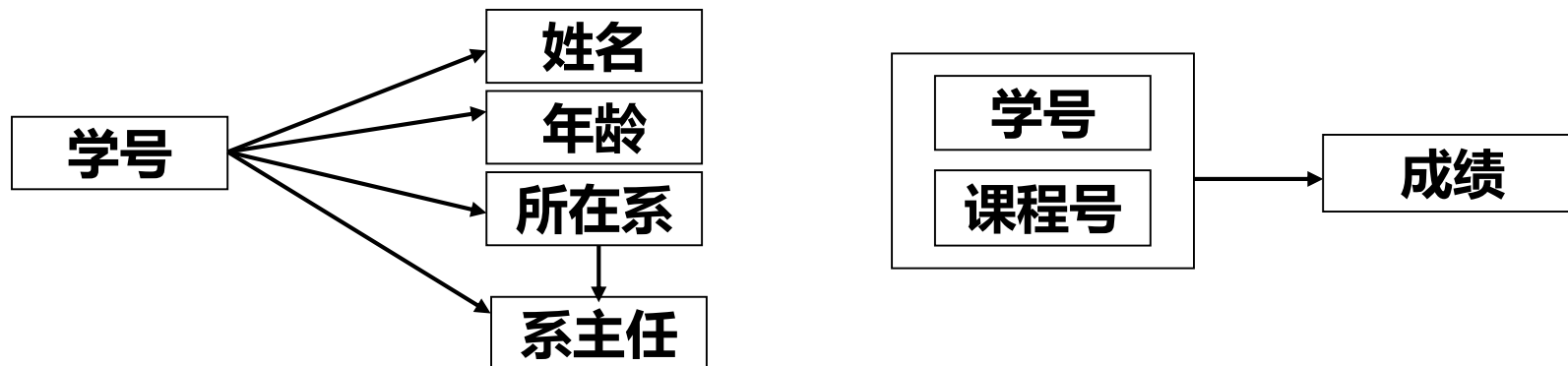
还有没有问题？

$SD \in 2NF$

$SC \in 2NF$

- 推论：**若 $R \in 1NF$ ，且其候选码为单个属性，则 $R \in 2NF$ 。 (Why?)

- SCD分解后，SD和SC的函数依赖分别如下图所示：



1NF→2NF,解决的问题:

- 消除了一些数据冗余，如：姓名、年龄不需要重复存储多次。
- 插入异常部分解决，如：学生的基本信息与选课信息分开插入。
- 删除异常部分解决，如：删除某学生选课信息，而学生其它信息不受影响。

仍然存在的问题:

- 冗余仍然存在，如：系主任。
- 更新复杂，如：更换系主任。
- 插入异常，如：没有学生，系主任无法插入。
- 删除异常，如：删除某系全体学生，删除系主任信息。

症由：非主属性对码的传递依赖： $sno \rightarrow dept, dept \rightarrow mname$

解决方法：投影分解，消去非主属性对码的传递依赖。

6.2.5 3NF

- 如果关系模式 $R(U, F) \in 1NF$ ，其中不存在码 X 、属性组 Y 及非主属性 Z ($Z \subseteq Y$) 使得 $X \rightarrow Y$ ($Y \rightarrow X$) / , $Y \rightarrow Z$ 成立，则称 R 属于**第三范式** (Third Normal Form)，简称**3NF**，记作 **$R \in 3NF$** 。
- 若 $R \in 3NF$ ，则每一非主属性既不部分依赖于码也不传递依赖于码。
- 若 $R \in 3NF$ ，则必有 $R \in 2NF$ 。
- 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。

■ **试证：如果 $R \in 3NF$ ，则 $R \in 2NF$ 。**

■ **证明：（反证法）**

假设 $R \in 3NF$ ，且 $R \notin 2NF$ 。

$\therefore R \notin 2NF$

\therefore 存在非主属性 Z 对码 X 的部分依赖，即存在一个 X 的真子集 X' ，使得
 $X' \rightarrow Z$ ；

$\therefore X \rightarrow X'$ ， $X' \rightarrow X$ ， $X' \rightarrow Z$ ，说明 R 中存在非主属性 Z 对码 X 的传递依赖，则根据第三范式的定义 $R \notin 3NF$ ，这与假设相矛盾。

\therefore 若 $R \in 3NF$ ，则 $R \in 2NF$ 。（证毕）

- 以2NF关系模式SD为例，说明3NF规范化的过程。
 SD(学号, 姓名, 年龄, 所在系, 系主任)
- 分析SD的属性，可以发现关系SD实际上描述了两个实体：
 - 一个为学生实体，属性有学号，姓名，年龄，所在系；
 - 一个是系的实体，其属性有系名和系主任。
- 可以将SD分解成如下两个关系：
 - S(学号, 姓名, 年龄, 所在系)，描述学生实体；
 - D(系名, 系主任)，描述系的实体。
- 分解后的两个关系S和D，码分别为学号和系名，不存在非主属性对码的传递函数依赖。因此， $S \in 3NF$ ， $D \in 3NF$ 。
- 分解为3NF后：
 - 数据冗余降低。
 - 不存在插入异常、不存在删除异常、不存在更新异常。

- **推论1:** 若 $R \in 2NF$, 且至多存在一个非主属性, 则 $R \in 3NF$
- **推论2:** 任何二元关系模式 $R(A,B)$ 必为3NF。
- **说明:**
 - 部分 FD 和传递 FD 是冗余及操作异常的重要根源。
 - 3NF 不存在非主属性对候选码的部分 FD 和传递 FD 。
 - 3NF 消去了大部分冗余及操作异常。
 - 但并非所有的3NF都能完全消除冗余及操作异常。



课堂练习：试证明，若关系模式R是二元关系，则 $R \in 3NF$ 。

证明：

假设R的两个属性分别为A和B，分两种情况证明：

1) 若R的码为单属性码

- ∵ R的码不可再分
- ∴ 不存在非主属性对码的部分依赖
- 又∵ R中只有两个属性
- ∴ 不存在非主属性对码的传递依赖
- ∴ $R \in 3NF$ 。

2) 若R的码为全码

- ∵ R中没有非主属性
- ∴ 不可能存在非主属性对码的部分依赖或传递依赖
- ∴ $R \in 3NF$ 。

案例：规范化的过程

■ 某书店购书情况汇总登记表：

订单号 NO	订户代号 C#	姓名 CN	地址 CA	书号 B#	书名 BN	出版单位 EU	单价 UP	定购数量 QUA
00001	0253	清华大学 自动化系	北京	02164	计算机语言	教育出版社	2.0	50
				01003	BASIC语言	科学出版社	1.1	50
				06372	FORTRAN语言	清华出版社	1.5	30
0002	0372	陈刚	上海	01083	电子线路	邮电出版社	1.4	1
				02954	自动控制原理	国防出版社	0.95	1
0003	2234	天津大学 计算机系	天津	02954	自动控制原理	国防出版社	0.95	20
				02164	计算机语言	教育出版社	2.0	50
				01083	电子线路	邮电出版社	1.4	50
0004	2523	北京钢铁学院 图书馆	北京	01003	BASIC语言	科学出版社	1.1	30
				06372	FORTRAN语言	清华出版社	1.5	30

案例：规范化的过程(2)



消除重复组后，关系模式满足1NF的要求。

订单号 NO	订户代号 C#	姓名 CN	地址 CA	书号 B#	书名 BN	出版单位 EU	单价 UP	订购数量 QUA
00001	0253	清华大学自动化系	北京	02164	计算机语言	教育出版社	2.0	50
00001	0253	清华大学自动化系	北京	01003	BASIC语言	科学出版社	1.1	50
00001	0253	清华大学自动化系	北京	06372	FORTRAN语言	清华出版社	1.5	30
0002	0372	陈刚	上海	01083	电子线路	邮电出版社	1.4	1
0002	0372	陈刚	上海	02954	自动控制原理	国防出版社	0.95	1
0003	2234	天津大学计算机系	天津	02954	自动控制原理	国防出版社	0.95	20
0003	2234	天津大学计算机系	天津	02164	计算机语言	教育出版社	2.0	50
0003	2234	天津大学计算机系	天津	01083	电子线路	邮电出版社	1.4	50
0003	2234	天津大学计算机系	天津	01003	BASIC语言	科学出版社	1.1	20
0004	2523	北京钢铁学院图书馆	北京	01003	BASIC语言	科学出版社	1.1	30
0004	2523	北京钢铁学院图书馆	北京	06372	FORTRAN语言	清华出版社	1.5	30

根据分析可以得到一组函数依赖：

$F = \{ NO \rightarrow C\#, C\# \rightarrow CN, C\# \rightarrow CA, B\# \rightarrow BN, B\# \rightarrow EU, B\# \rightarrow UP, (NO, B\#) \rightarrow QUA \}$,

表中 $(NO, B\#)$ 为主码。

案例：规范化的过程(3)

- **消除部分函数依赖**，将其分解成三个关系，使每一个非主属性都完全依赖于主码，满足2NF的要求。

订单号 NO	订户代号 C#	姓名 CN	地址 CA	书号 B#	书名 BN	出版单位 EU	单价 UP
00001	0253	清华大学自动化系	北京	02164	计算机语言	教育出版社	2.0
00002	0372	陈刚	上海	01003	BASIC语言	科学出版社	1.1
00003	2234	天津大学计算机系	天津	06372	FORTRAN语言	清华出版社	1.5
00004	2523	北京钢铁学院图书馆	北京	01083	电子线路	邮电出版社	1.4
				02954	自动控制原理	国防出版社	0.95
订单号 NO	书号 B#	定购数量 QUA					
00001	02164	50					
00001	01003	50					
00001	06372	30					
00002	01083	1					
00002	02954	1					
00003	02954	20					
00003	02164	50					
00003	01083	50					
00003	01003	20					
00004	01003	30					
00004	06372	30					

各自函数依赖：

$F1=\{NO \rightarrow C\#, C\# \rightarrow CN, C\# \rightarrow CA\}$ ，NO为主码。

$F2=\{B\# \rightarrow BN, B\# \rightarrow EU, B\# \rightarrow UP\}$ ，B#为主码。

$F3=\{(NO, B\#) \rightarrow QUA\}$ ，(NO, B#) 为主码。

案例：规范化的过程(4)



- 进一步消除传递函数依赖，满足3NF的要求。

订户代号 C#	姓名 CN	地址 CA	书号 B#	书名 BN	出版单位 EU	单价 UP
0253	清华大学自	北京	02164	计算机语言	教育出版社	2.0
0372	陈刚	上海	01003	BASIC语言	科学出版社	1.1
2234	天津大学计	天津	06372	FORTRAN语言	清华出版社	1.5
2523	北京钢铁学	北京	01083	电子线路	邮电出版社	1.4
			02954	自动控制原理	国防出版社	0.95

订单号 NO	书号 B#	定购数量 QUA
00001	02164	50
00001	01003	50
00001	06372	30
0002	01083	1
0002	02954	1
0003	02954	20
0003	02164	50
0003	01083	50
0003	01003	20
0004	01003	30
0004	06372	30

订单号 NO	订户代号 C#
00001	0253
0002	0372
0003	2234
0004	2523

各自函数依赖：

$F11 = \{ NO \rightarrow C\# \}$ ，NO为主码。

$F12 = \{ C\# \rightarrow CN, C\# \rightarrow CA \}$ ，C#为主码。

3NF可能存在的问题

■ 语义:

- ❖ 每门课可由多个教师讲;
- ❖ 每位教师只讲一门课;
- ❖ 学生一旦选定一门课,
就确定了一个固定教师。

■ 解:

- 候选码: (学号, 课程), (学号, 教师)
- 非主属性: 无
- 函数依赖: (学号, 课程) \rightarrow 教师, 教师 \rightarrow 课程
- 思考: $R \in 3NF$?
 \therefore 不存在非主属性, $\therefore R \in 3NF$ 。

R

学号	课程	教师
S1	OS	于得水
S2	OS	马千里
S4	DB	牛得草
S5	DB	牛得草
S5	OS	于得水
.....

3NF可能存在的问题

1) 数据冗余

多个学生选同一个教师的同一门课时，
课程，教师重复。

2) 更新异常

某门课程改名，则修改多个元组。

3) 插入异常

候选码为（学号，课程）和（学号，教师），则：

教师开设某门课程，尚无学生选修，因为无学号信息，则课程、教师信息不能进入数据库。

4) 删除异常

删去学生选课信息，丢失教师授课信息。

R

学号	课程	教师
S1	OS	于得水
S2	OS	马千里
S4	DB	牛得草
S5	DB	牛得草
S5	OS	于得水
.....

3NF可能存在的问题

- 存在上述问题的原因：
存在主属性对候选码的不良依赖。
 \therefore 教师 \rightarrow 课程 \therefore (学号, 教师) \rightarrow 课程
- 解决方法：投影分解
——消去主属性对候选码的部分fd。
- 可将R分解为：

学号	教师
S1	于得水
S2	马千里
.....

教师	课程
于得水	OS
马千里	OS
牛得草	DB
.....

- 分解后存在的问题得到解决，新关系模式满足BC范式。

6.2.6 BC范式

定义：如果关系模式 $R(U, F) \in 1NF$ ，且对于所有的非平凡函数依赖 $X \rightarrow Y$ ， X 必包含了 R 的一个码，则称 R 属于**BC范式 (Boyce-Codd Normal Form)**，记作 $R \in BCNF$ 。

定理：如果 $R \in BCNF$ ，则 $R \in 3NF$ 。

证明：（反证法）

假设 $R \in BCNF$ ，且 $R \notin 3NF$ 。

$\because R \notin 3NF$ ， \therefore 存在码 X 、属性组 Y 及非主属性 Z ($Z \notin Y$) 使得 $X \rightarrow Y$ ($Y \not\rightarrow X$)， $Y \rightarrow Z$ 成立；

$\because Y \rightarrow X$ ， $\therefore Y$ 不包含码，即非平凡函数依赖 $Y \rightarrow Z$ 的决定部分不包含码。

根据BC范式的定义， $R \notin BCNF$ ，这与假设相矛盾。

$\therefore R \in 3NF$ 。（证毕）

- 关系模式 $R(\text{Sno}, \text{Cno}, \text{ORDER})$ 表示学生选修课程的成绩排名（假设每门课中没有并列名次），
- 则由语义有函数依赖：
 $(\text{Sno}, \text{Cno}) \rightarrow \text{ORDER}$, $(\text{Cno}, \text{ORDER}) \rightarrow \text{Sno}$,
请问：R属于3NF吗？ R属于BCNF吗？

■ 解答：

- (Sno, Cno) 或者 $(\text{Cno}, \text{ORDER})$ 都可以作为候选码
- 不存在非主属性对码传递依赖或部分依赖， $SCO \in 3NF$
- 没有其他决定因素， $R \in BCNF$

■ BCNF的性质:

- 1) 所有非主属性都完全依赖于候选码;
 - 2) 所有非主属性都不传递依赖于候选码;
 - 3) 所有主属性都完全依赖于不包含它的候选码;
 - 4) 所有主属性都不传递依赖于候选码。
- ❖ 如果一个关系DB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, 但仍然可能存在主属性对候选键的部分依赖和传递依赖。
 - ❖ 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, 完全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。
 - ❖ 但是除了函数依赖, 还存在另一种数据依赖——多值依赖。

6.2.7 多值依赖

1.问题的提出

设学校中某一门课程C由多个教师T讲授，他们使用相同的一套参考书B，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。

考察关系模式：

Teaching(C, T, B)

课程C

教师T

参考书B

课程C	教员T	参考书B
物理	李勇 王军	普通物理学 光学原理
数学	李勇 张平	数学分析 高等代数
计算数学	张平 周峰	数学分析

Teaching(C, T, B) 规范化的二维表格:

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	王军	普通物理学
物理	王军	光学原理
数学	李勇	数学分析
数学	李勇	高等代数
数学	张平	数学分析
数学	张平	高等代数
计算数学	张平	数学分析
计算数学	周峰	数学分析

- **分析:**
 - 具有唯一候选码(C,T,B), 即全码
 - $\text{Teaching} \in \text{BCNF}$
 - Teaching模式是否存在不良特性?
 - 数据冗余
 - 插入异常
 - 删除异常
 - 更新异常
- **症由:** 多值依赖

2. 定义:

定义6.15(描述型) 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 $Z = U - X - Y$, 多值依赖 $X \twoheadrightarrow Y$ 成立, 当且仅当:

- 1) 对 R 的任一关系 r , r 在 (X, Z) 上的每个值对应一组 Y 的值;
- 2) 这组值仅仅决定于 X 值而与 Z 值无关。

例: Teaching (C, T, B) 有

$C \twoheadrightarrow T$ 和 $C \twoheadrightarrow B$

对于 C 的每一个值, T 有一组值与之对应, 而不论 B 取何值

多值依赖



定义6.15' (形式化) 在 $R(U)$ 的任一关系 r 中, 如果存在元组 s, t 使得 $s[X] = t[X]$, 那么就必然存在元组 $u, v \in r$, (v, u 可以与 s, t 相同), 使得:

$$v[X] = u[X] = s[X] = t[X]$$

$$v[Y] = s[Y], u[Y] = t[Y]$$

$$v[Z] = t[Z], u[Z] = s[Z]$$

	X	$Z=U-X-Y$	Y
s	X	Z1	Y1
t	X	Z2	Y2
u	X	Z1	Y2
v	X	Z2	Y1

(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, $Z = U - X - Y$ 。

或: r 中若存在 $(x, y1, z1)$ 和 $(x, y2, z2)$, 则必存在 $(x, y2, z1)$ 和 $(x, y1, z2)$

多值依赖

- 3. 平凡多值依赖与非平凡多值依赖:
 - 若 $X \twoheadrightarrow Y$, 而 $Z = \varphi$, 即 $U = X \cup Y$: 则称

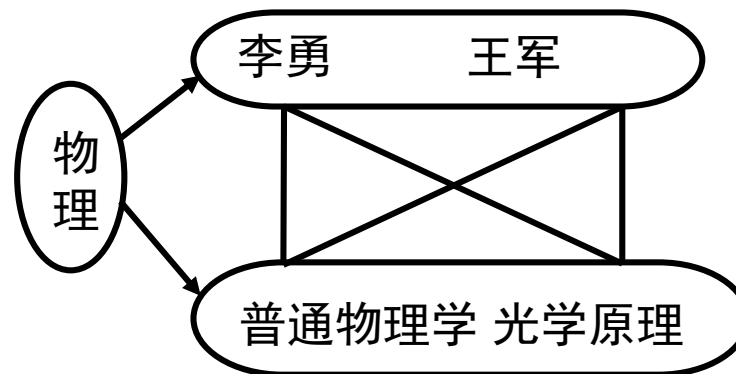
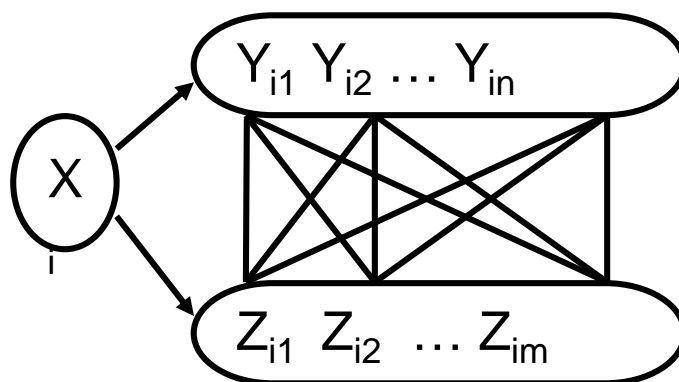
$X \twoheadrightarrow Y$ 为平凡的多值依赖

否则称 $X \twoheadrightarrow Y$ 为非平凡的多值依赖。

- 4. MVD性质:

- 多值依赖具有对称性

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



■ MVD性质(续):

- 函数依赖是多值依赖的特例，即

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

(此时, s 、 t 在 X 上的投影相等，则在 Y 上的投影也必然相等，该投影与 s 和 t 在 $Z = U - X - Y$ 上的投影无关。)

- 多值依赖具有传递性:

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

- 并规则:若 $X \twoheadrightarrow Y$ ， $X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow (Y \cup Z)$

- 交规则:若 $X \twoheadrightarrow Y$ ， $X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow (Y \cap Z)$

- 差规则:

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

5. 多值依赖与函数依赖的区别:

- 函数依赖规定某些元组不能出现在关系中，多值依赖要求某种形式的其它元组必须在关系中。
- 有效性与属性集的范围不同：
 - $X \rightarrow Y$ 的有效性仅决定于 X 、 Y 属性集上的值；
 - $X \twoheadrightarrow Y$ 不仅涉及属性组 X 和 Y ，而且涉及 U 中其余属性 Z 。
 - 若 $X \rightarrow Y$ 在 $R(U)$ 上成立，则对于任何 $Y' \subseteq Y$ ，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。

- 5. 多值依赖与函数依赖的区别（续）：
 - 若 $X \twoheadrightarrow Y$ 在 U 上成立，则在 W ($X \cup Y \subseteq W \subseteq U$) 上一定成立，称 $X \twoheadrightarrow Y$ 为 $R(U)$ 的**嵌入型多值依赖**。反之则不然，即 $X \twoheadrightarrow Y$ 在 W ($W \subset U$) **上成立**，在 U 上并**不**一定成立。
 - 只要在 $R(U)$ 的任何一个关系 r 中，元组在 X 和 Y 上的值满足定义5.1（函数依赖），则函数依赖 $X \rightarrow Y$ 在任何属性集 W ($X \cup Y \subseteq W \subseteq U$) 上成立。

例：

A	B	C	D
a1	b1	c1	d1
a1	b1	c2	d1
a1	b2	c1	d2
a1	b2	c2	d2

A	B	C	D
a1	b1	c1	d1
a1	b1	c1	d2
a1	b2	c2	d1
a1	b2	c2	d2

$A \twoheadrightarrow B$ 在
 $\{ABC\}$ 上成立，
而在 $\{ABCD\}$ 上
不成立

$A \twoheadrightarrow BC$ 成立
 $A \twoheadrightarrow B$ 不成立

多值依赖



- 例：找出关系上所满足的多值依赖：

A	B	C
a1	b1	c1
a1	b1	c2
a2	b1	c1
a2	b1	c3

$A \twoheadrightarrow C$

$A \twoheadrightarrow B$

$C \twoheadrightarrow B$

$C \twoheadrightarrow A$

事实上 $A \rightarrow B, C \rightarrow B$

$B \twoheadrightarrow C$? 若使 $B \twoheadrightarrow C$, 需加入哪些元组?

a1	b1	c3
a2	b1	c2

加入后有没有破坏前述
依赖关系?

- 6. 不当MVD的关系式引起的弊端及对策:
 - Teaching(C,T,B) 分解为: T(C,T) 和 B(C,B)

课程C	教员T
物理	李勇
物理	王军
数学	李勇
数学	张平
计算数学	张平
计算数学	周峰

课程C	参考书B
物理	普通物理学
物理	光学原理
数学	数学分析
数学	高等代数
计算数学	数学分析

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** $X \twoheadrightarrow Y$ ($Y \subseteq X$)，X都含有候选码，则 $R \in 4NF$ 。

不允许有非平凡且非函数依赖的多值依赖。

允许的是非平凡多值依赖，即函数依赖。

如果 $R \in 4NF$ ，则 $R \in BCNF$ 。

[例] 关系模式: Teaching(C,T,B) $\in 4NF$ ，码为(C, T, B)

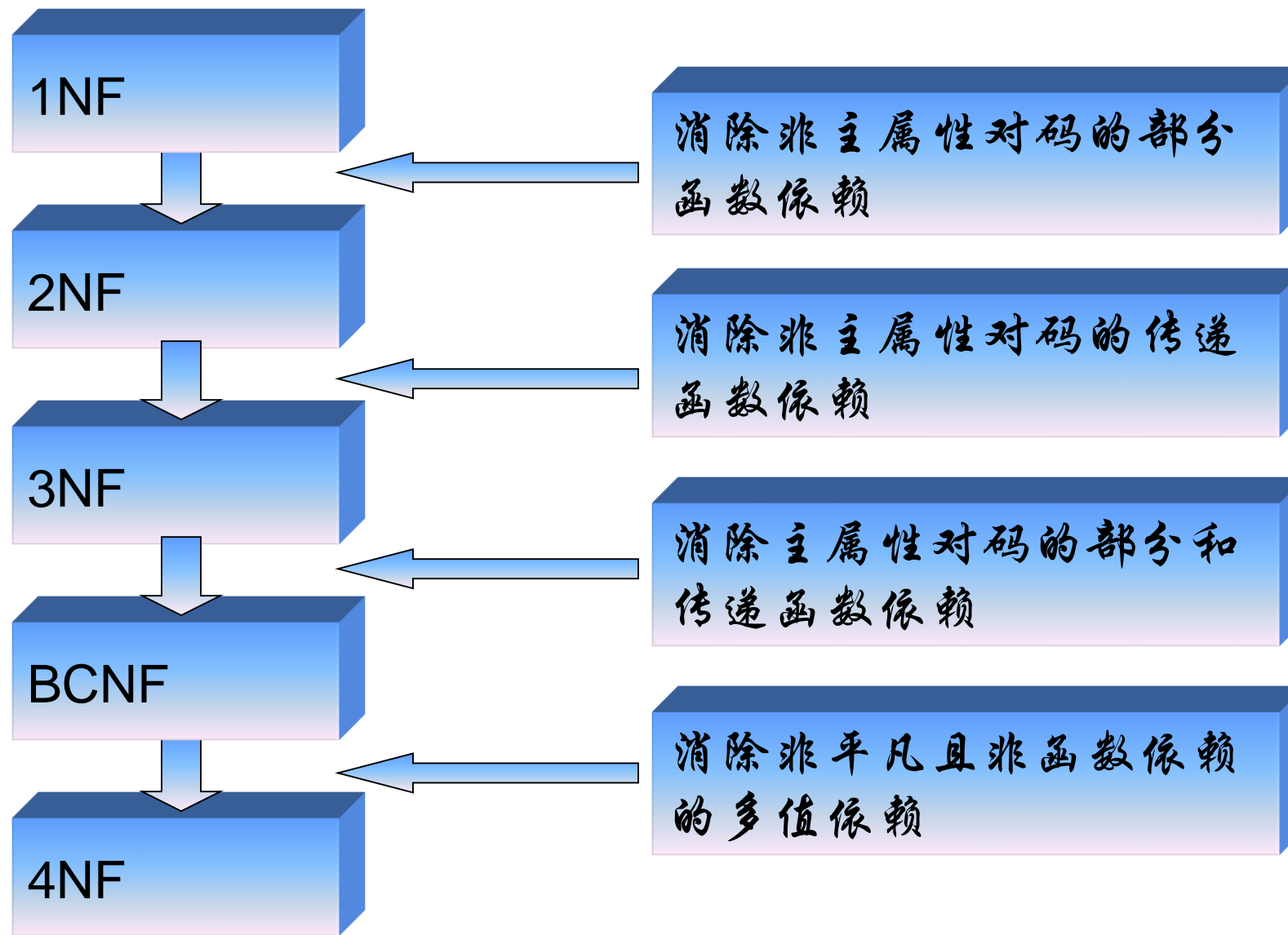
- 存在MVD: $C \twoheadrightarrow T$, 且C不是候选码
- 分解成两个关系模式:

$T(C, T) \in 4NF$ (码:CT)

$B(C, B) \in 4NF$ (码:CB)

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

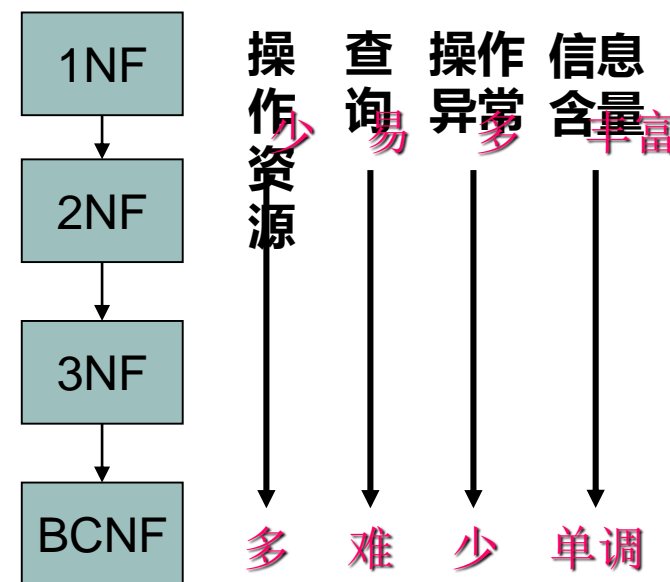
6.2.9 关系规范化小结



结论



- 1) 3NF必定为2NF和1NF，反之不一定；
- 2) 4NF必为BCNF，BCNF必为3NF，反之不一定；
- 3) 3NF已在很大程度上控制了数据冗余、很大程度上消去了插入和删除操作异常；但，3NF分解仍不够彻底（可能存在主属性对候选码的部分fd和传递fd）；
- 5) 在fd范围内，BCNF下已完全消去了插入删除异常；
- 6) 4NF是多值依赖范畴内最高程度的规范化；
- 7) 范式并非越高越好。理论上数据库设计一般应规范到3NF，但实际应用中为减少连接运算，提高查询效率，不一定都达到3NF。



- 8) 适可而止（垃圾，连接运算）；
- 9) 分解不唯一。

关系模式的规范化

反规范化——在数据库中，为了提高查询效率而有意降低关系规范化程度的一种技术。

例如：一般选课数据库的关系模式如下：

class(clno, dept, major)

student(sno, ID-card, sname, sex, birthdate, clno)

course(cno, cname, score, pcno)

sc(sno, cno, grade)

但对于查分系统，为提高查询效率，可设计如下关系模式：

students (sno, ID-card, sname, cno, grade)

思考：

- 任何一个二目关系模式 $R(A, B)$ 一定属于BCNF吗？一定属于4NF吗？
- 关系模式 $R(A, B, C)$ ， $AB \twoheadrightarrow C$ 一定成立吗？
- 一个全是主属性的关系模式一定可以达到第几范式？
- 一个全码的关系模式一定可以达到第几范式？