

第十六讲

范式及无损分解

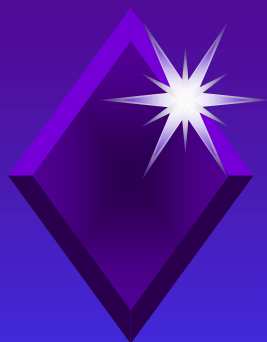


本讲主要目标



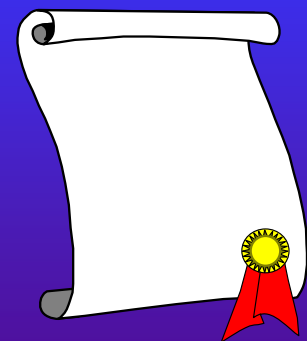
学完本讲后，你应该能够了解：

1. 第二范式、第三范式、BC范式的定义，以及范式之间的包含关系；
2. 对于一个关系模式，判断其最高属于第几范式，并使用基本的模式分解方法，将1NF分解为2NF，将2NF分解为3NF；
3. 分解的无损连接性和保持函数依赖性的概念；
4. 使用规范化方法设计数据库的逻辑结构时，并不是要使得到的数据库模式都达到最高范式，而是还需要平衡查询效率和更新代价₂₀



内容提纲

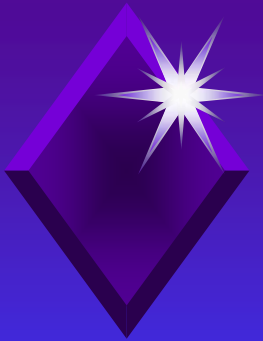
- 一. 范式：2NF、3NF和BCNF
- 二. 多值依赖
- 三. 范式：4NF
- 四. 无损分解
- 五. 规范化回顾
- 六. 其它设计考虑





范式：2NF、

 3NF和BCNF



范式：2NF、3NF和BCNF

1. 范式(Normal Forms)

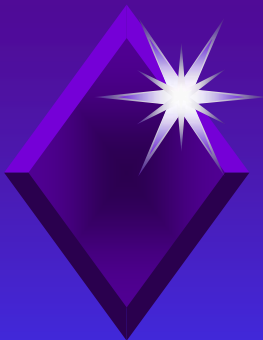
- 范式的定义：关系数据库中符合某一级别的关系模式的集合。所谓“第几范式”，是表示关系的某一种级别，R为第几范式就可以写成

$$R \in xNF$$

- 各范式之间的联系有

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

- 并不总是需要达到最高范式



范式：2NF、3NF和BCNF

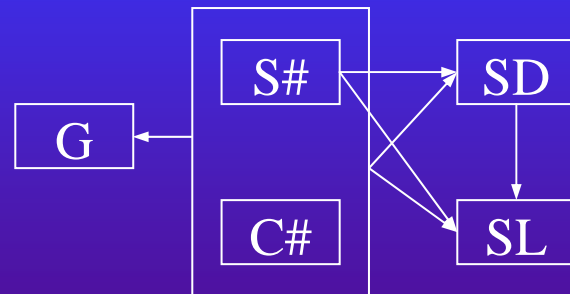
2. 实例

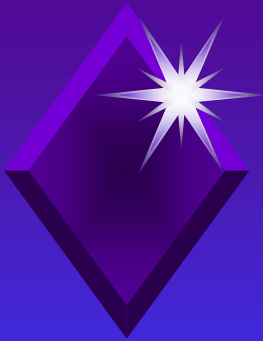
例，有一个关系模式S-L-C (S#, SD, SL, C#, G)，其中S#为学生的学号，SD为学生所在系，SL为学生的住处，并且每个系的学生住在同一个地方，C#为课程号，G为成绩。这里键为 (S#, C#)。

函数依赖有：

$$\begin{array}{ll} S\# \rightarrow SD, & (S\#, C\#) \xrightarrow{F} G \\ S\# \rightarrow SL, & \xrightarrow{P} SD \\ SD \rightarrow SL, & (S\#, C\#) \xrightarrow{P} SL \\ & (S\#, C\#) \xrightarrow{P} SL \end{array}$$

图示





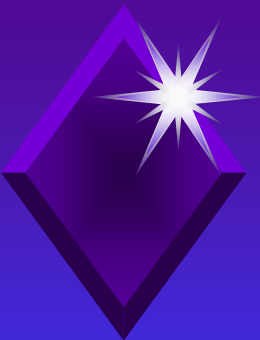
范式：2NF、3NF和BCNF

3. 第一范式 (1NF)

第一范式 (First Normal Form, 简称为 1NF)

—— 如果一张表不含有**多值属性** (有时称为**重复字段**) 和**内部结构** (比如记录类型) 的列, 则称该表为第一范式

- 关系模式 $S-L-C(S\#, SD, SL, C\#, G) \in 1NF$
- 关系模式 $S-L-C$ 存在更新异常、插入异常和删除异常

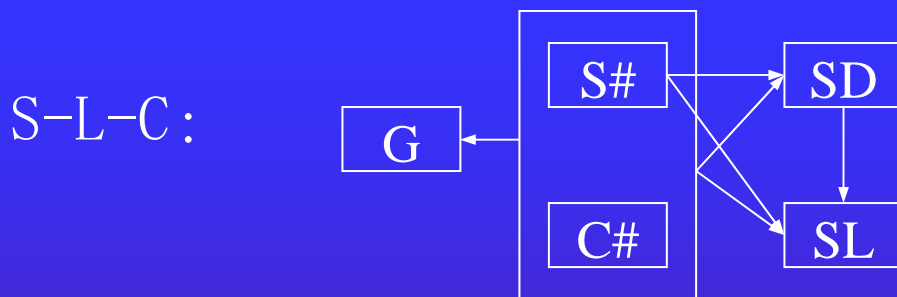


范式：2NF、3NF和BCNF

4. 第二范式 (2NF)

第二范式 (Second Normal Form, 简称2NF)

—— 若 $R \in 1NF$, 且每一个非主属性完全函数依赖于键, 则 $R \in 2NF$ 。

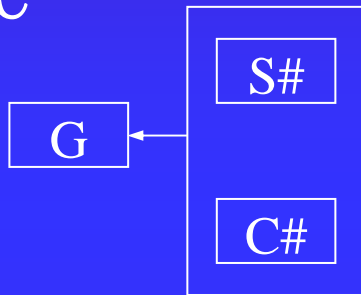


- 分析：存在非主属性对键的部分函数依赖
- 结果： $S-L-C \in 1NF$, 但 $S-L-C \notin 2NF$
- 分解：将S-L-C分解为： $S-L(S\#, SD, SL)$, 和 $S-C(S\#, C\#, G)$

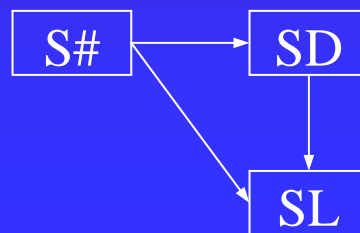
范式：2NF、3NF和BCNF

4. 第二范式 (2NF)

S-C



S-L

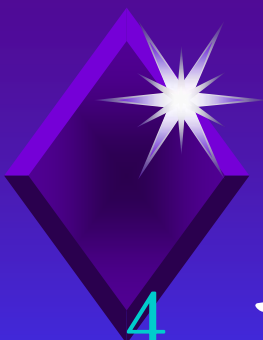


◆ 分析： S-C的键为 (S#, C#) ， S-L的键为S#， 不存在非主属性对键的部分函数依赖。

◆ 结果： $S-C \in 2NF$ ； $S-L \in 2NF$

◆ 问题：

S-C和S-L中消除了1NF
中的某些更新异常，但
仍然存在更新异常



范式：2NF、3NF和BCNF

4. 第二范式 (2NF)

在S-L (S#, SD, SL) 和S-C (S#, C#, G) 中

☆ 一个学生修很多课程



☆ 一个学生转系



☆ 插入一个尚未选课的学生信息



☆ 一个学生只选修了一门课，但现在决定不选了



☆ 一个系有很多学生，且同系学生住在一个地方



☆ 一个学生转系



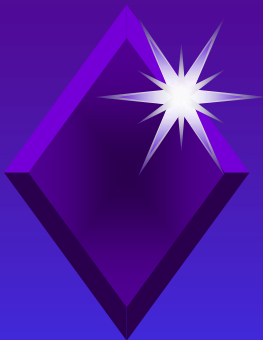
☆ 一个新系创建但新生尚未注册



☆ 一个系所有学生毕业了



结论：2NF可以消除一些1NF中存在的更新异常，但不能彻底消除更新异常



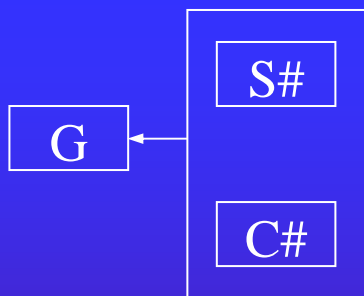
范式：2NF、3NF和BCNF

5. 第三范式 (3NF)

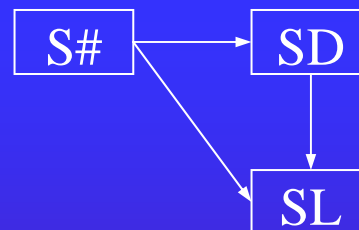
第三范式 (Third Normal Form, 简称3NF)

—— 若 $R \in 2NF$ ，且每一个非主属性不传递函数依赖于键，则 $R \in 3NF$ 。

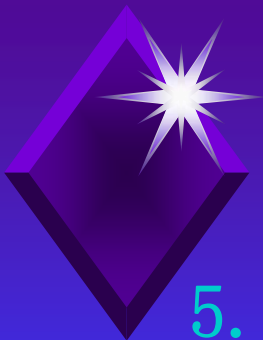
S-C



S-L



- 分析：S-C的键为 $(S\#, C\#)$ ，不存在非主属性对键的传递函数依赖。S-L的键为 $S\#$ ，存在非主属性对键的传递函数依赖 $S\# \rightarrow SL$ 。
- 结果： $S-C \in 3NF$ ； $S-L \notin 3NF$
- 分解：将S-L分解为：S-D ($S\#, SD$) 和 D-L (SD, SL)



范式：2NF、3NF和BCNF

5. 第三范式 (3NF)

S-D

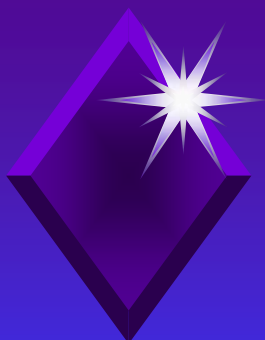
D-L



- 分析： S-D的键为S#， D-L的键为SD， 不存在非主属性对键的传递函数依赖。
- 结果： $S-D \in 3NF$ ； $D-L \in 3NF$
- 问题：

S-D和D-L中消除了前面的更新异常了吗？

3NF存在更新异常吗？



范式：2NF、3NF和BCNF

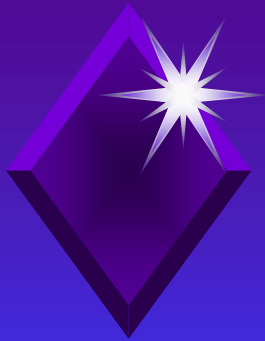
5. 第三范式 (3NF)

在S-D (S#, SD) 和 D-L (SD, SL) 中：

- ☆ 一个系有很多学生，且同系学生住在一个地方 ✓
- ☆ 一个学生转系 ✓
- ☆ 一个新系创建但新生尚未注册 ✓
- ☆ 一个系所有学生毕业了 ✓

3NF不能
彻底消
除更新
异常

结论：3NF可以消除一些2NF中存在的更新异常，



范式：2NF、3NF和BCNF

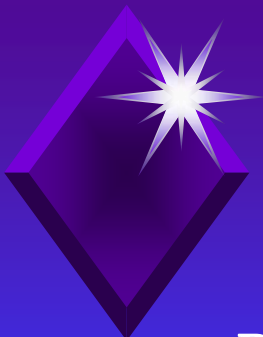
6. Boyce-Codd范式 (BCNF)

Boyce-Codd 范 式 (Boyce-Codd Normal Form, 简称 BCNF) —— $R \in 1NF$, 且每一个**决定因素都包含键**, 则 $R \in BCNF$ 。

在S-D ($S\#, SD$) 和 D-L (SD, SL) 中:

$S-D \in BCNF$, $D-L \in BCNF$

所有的3NF
都是BCNF ?



范式：2NF、3NF和BCNF

6. Boyce-Codd范式 (BCNF)

例 关系模式STJ (S, T, J) 中, S表示学生, T表示教师, J表示课程。每一教师只教一门课。每门课有若干教师, 某一学生选定某门课, 就对应一个固定的教师。

由语义可以得到函数依赖

$$T \rightarrow J ; \quad (S, J) \rightarrow T ; \quad (S, T) \rightarrow J$$

键为 (S, J) , (S, T)

分析:

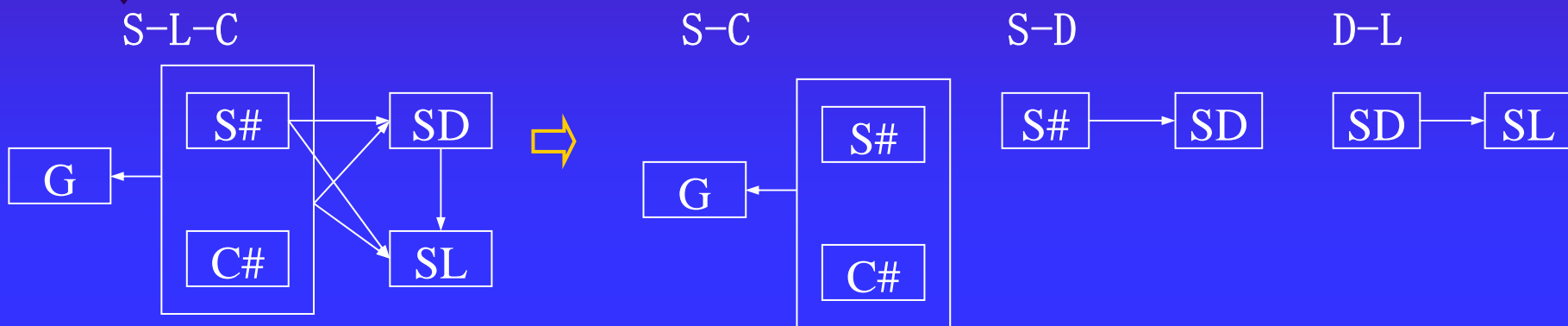
(1) 没有非主属性, 所以, 不存在非主属性对键的部分与传递依赖, 因此, $STJ \in 3NF$

(2) 在 $T \rightarrow J$ 中, 决定因素 T 不包含键, 因此, $STJ \notin BCNF$

结论: 并非所有的3NF都是BCNF

范式：2NF、3NF和BCNF

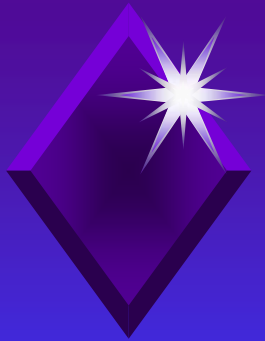
小结：



- ◆ 分析：S-C的键为 (S#, C#)，决定因素包含键；
S-D的键为S#，决定因素包含键；
D-L的键为SD，决定因素包含键。

- ◆ 结果：S-C \in BCNF, S-D \in BCNF, D-L \in BCNF

因此，S-L-C分解成了三个BCNF：S-C (S#, C#, G)，
S-D (S#, SD)，D-L (SD, SL)。



范式：2NF、3NF和BCNF

如果一个模式属于BCNF，在函数依赖的范畴内，彻底消除了更新异常吗？

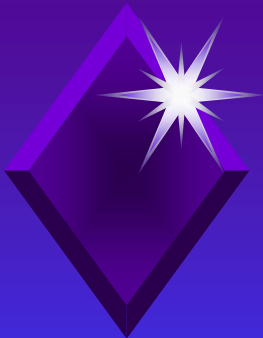
Yes



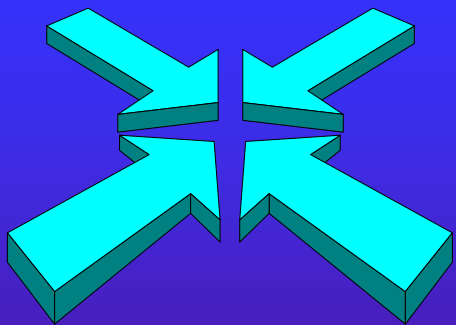
如果一个模式属于BCNF，在数据依赖的范畴内，彻底消除了更新异常吗？

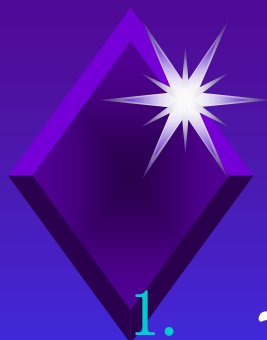
No

其它的数据依赖也会产生更新异常



多值依赖





多

值

依

赖

实例

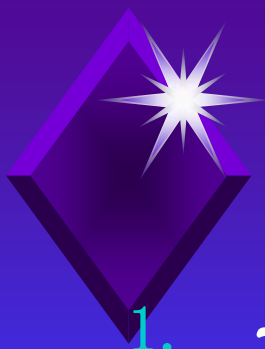
关系模式TEACHING：学校某一门课由多个教员讲授，他们使用相同的一套参考书。每个教员可以讲授多门课程，每种参考书可以供多门课程使用。

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

有函数
依赖
吗？

NO

有更新
异常
吗？



多

值

依

赖

1. 实例

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	李勇	物理习题集
物理	王军	普通物理学
物理	王军	光学原理
物理	王军	物理习题集
数学	李勇	数学分析
数学	李勇	微分方程
数学	李勇	高等代数
数学	张平	数学分析
数学	张平	微分方程
数学	张平	高等代数

★ 冗余

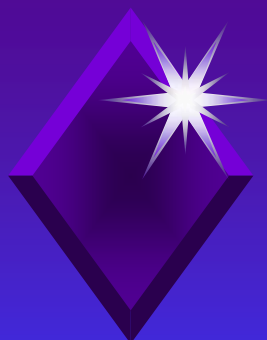
★ 某一课程增加一名教员

-- 必须插入多个元组

★ 某一门课要去掉一本参考书

-- 必须删除多个元组

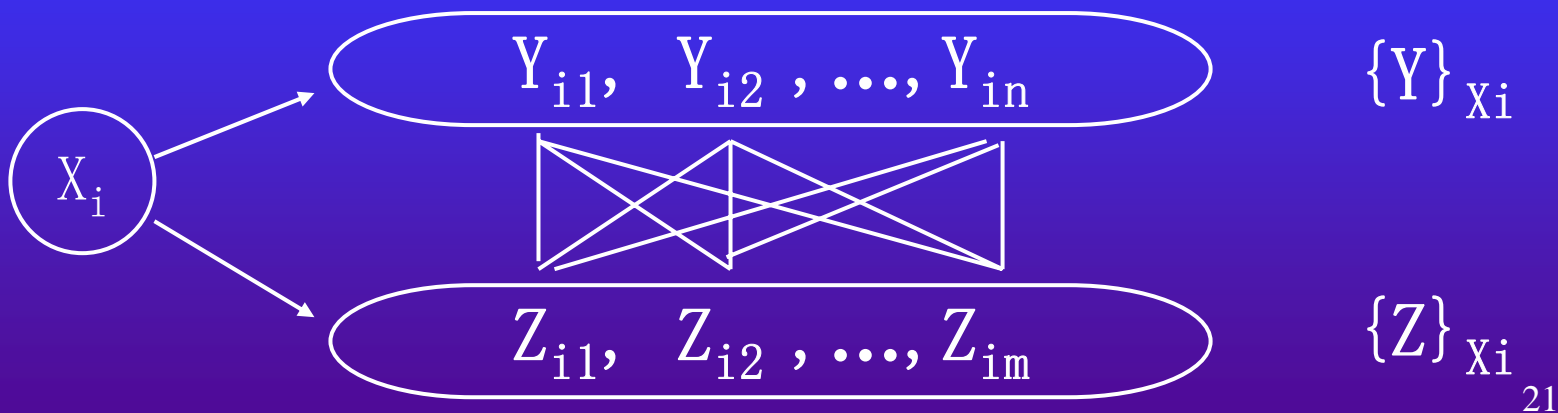
因为存在
多值依赖

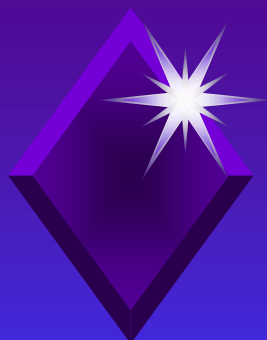


多 值 依 赖

2. 多值依赖的定义

设 $R(U)$ 是属性集 U 上的一个关系模式。 X, Y, Z 是 U 的子集, 并且 $Z = U - X - Y$ 。关系模式 $R(U)$ 中多值依赖 $X \twoheadrightarrow Y$ 成立, 当且仅当对 $R(U)$ 的任一关系 r , 给定的一对 (x, z) 值, 有一组 Y 的值, 这组值仅仅决定于 x 值而与 z 值无关。

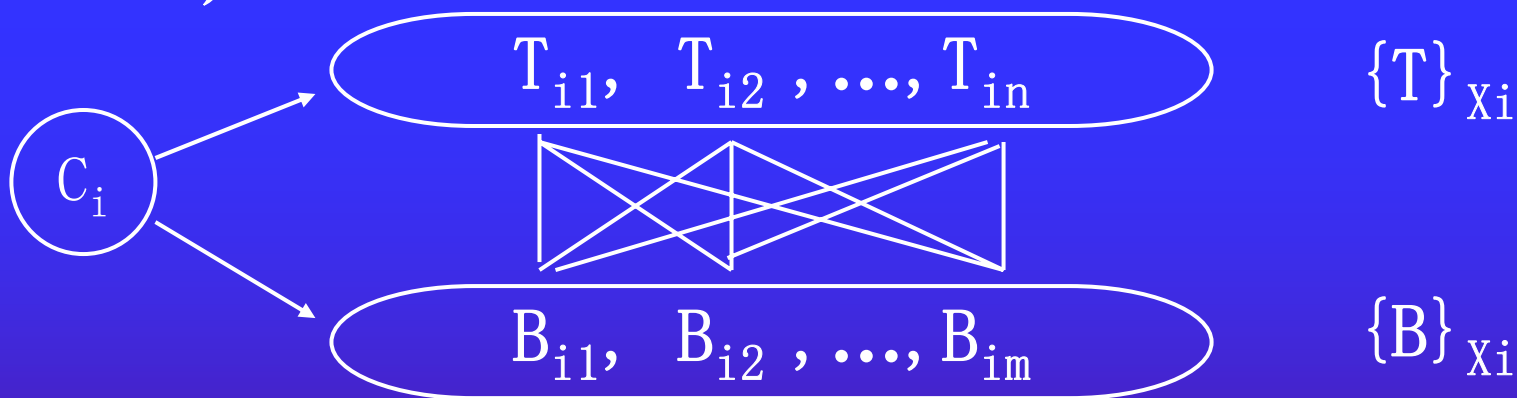




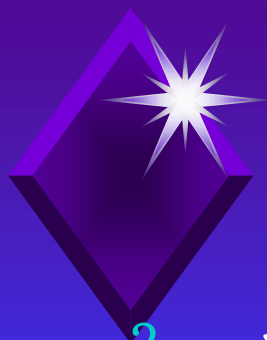
多 值 依 赖

2. 多值依赖的定义

例 在关系模式TEACHING (C, T, B)
中,



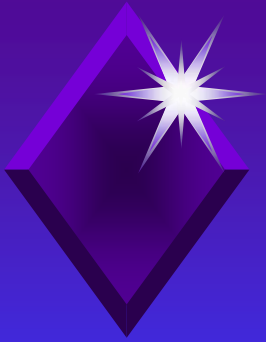
因此, 存在多值依赖 $C \twoheadrightarrow T$



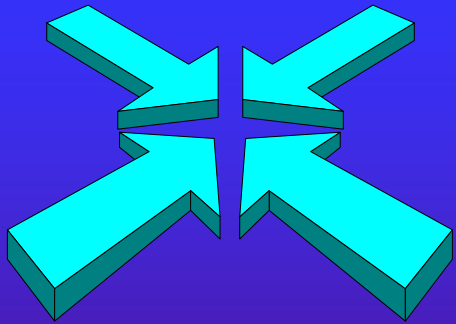
多 值 依 赖

3. 多值依赖的性质

- 对称性
- 传递性
- 多值依赖的有效性与属性集的范围有关
- 若 $X \twoheadrightarrow Y$ ，而 $Z = \emptyset$ ，则称 $X \twoheadrightarrow Y$ 为平凡的多值依赖。



范式：4NF





范式：4NF

➤ 4NF的定义

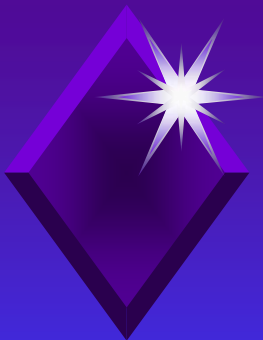
关系模式 $R(U) \in 1NF$ ，如果对于每个**非平凡多值依赖** $X \twoheadrightarrow Y$ ($X \not\subseteq Y$)， X 都含有键，则称 $R(U) \in 4NF$ 。

➤ 定理

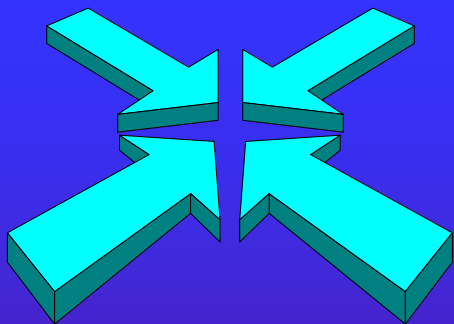
如果一个关系模式是4NF，则必为BCNF。

例 在关系模式 $TEACHING(C, T, B)$ 中，存在非平凡的多值依赖 $C \twoheadrightarrow T$ 和 $C \twoheadrightarrow B$ ，键为 (C, T, B) ，而 C 不是键，因此， $TEACHING \notin 4NF$ 。

将 $TEACHING$ 分解为 $T1(C, T)$ 和 $T2(C, B)$ ，只存在平凡的多值依赖 $C \twoheadrightarrow T$ 和 $C \twoheadrightarrow B$ ，因此
 $T1 \in 4NF$, $T2 \in 4NF$



无损分解





无损分解

- 模式分解的**属性等价性** —— 如果关系模式 $R(A)$ 被分解为关系模式 $R_1(A_1), R_2(A_2), \dots, R_n(A_n)$, 且

$$A = A_1 \cup A_2 \cup \dots \cup A_n$$

则该分解是属性等价的分解

- 模式分解的**无损连接性** —— 如果关系模式 $R(A)$ 被分解为关系模式 $R_1(A_1), R_2(A_2), \dots, R_n(A_n)$, 且

$$R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_n$$

则该分解是**无损连接**的分解

在对关系模式进行分解的过程中, 满足属性等价且有冗余属性的分解一定具有无损连接性吗?

无损分解

A	B	C
a ₁	b ₁	c ₁
a ₂	b ₂	c ₁

分解



A	C
a ₁	c ₁
a ₂	c ₁

B	C
b ₁	c ₁
b ₂	c ₁

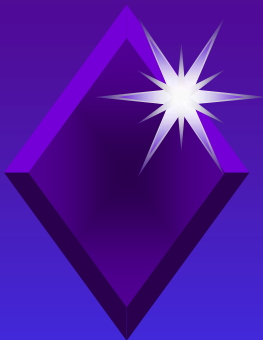
连接



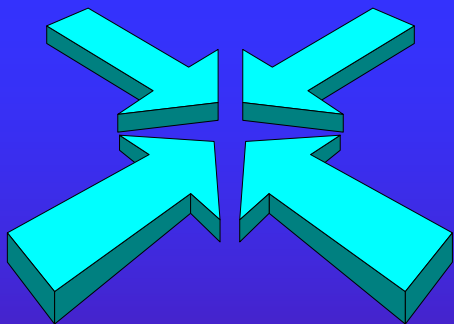
A	B	C
a ₁	b ₁	c ₁
a ₁	b ₂	c ₁
a ₂	b ₁	c ₁
a ₂	b ₂	c ₁

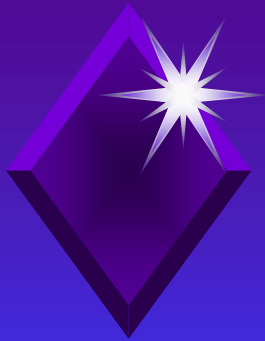


**结论：满足属性等价且有冗余属性的分解
不一定具有无损连接性**



规范化回顾

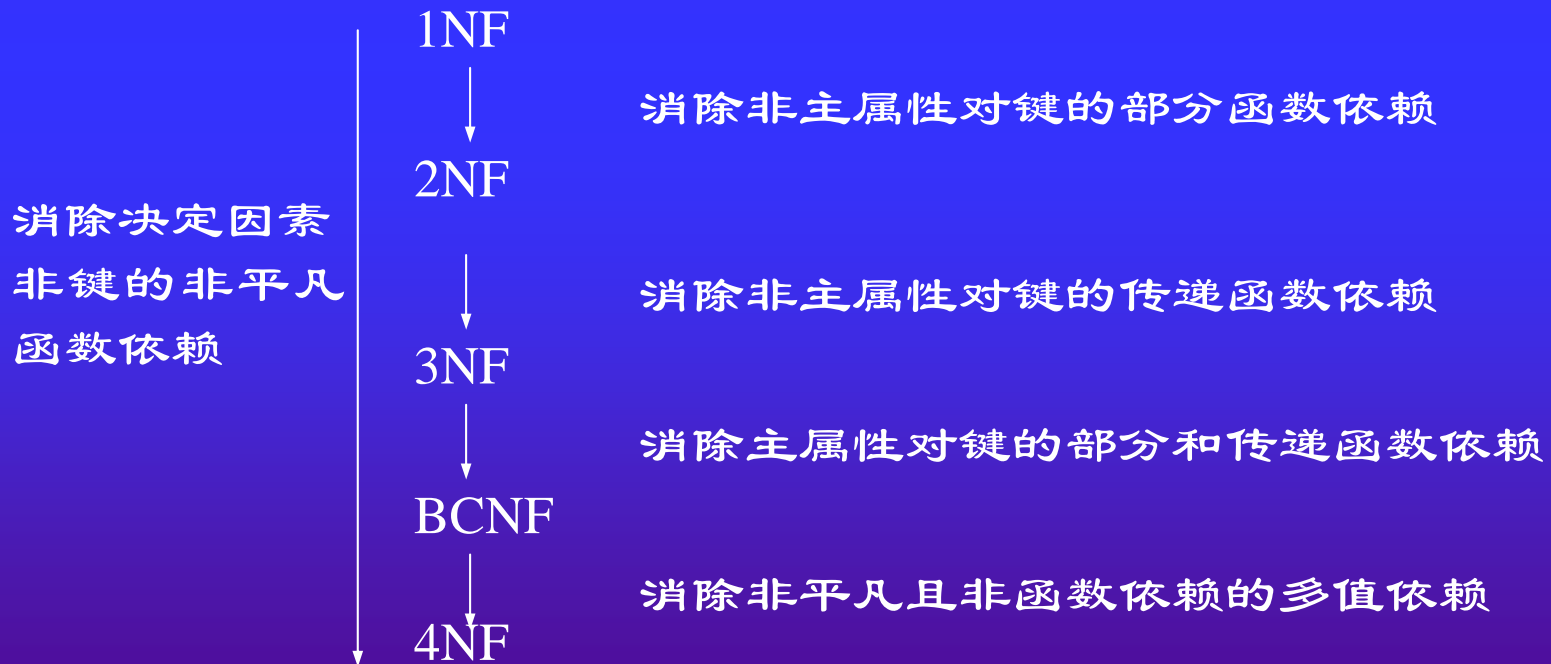


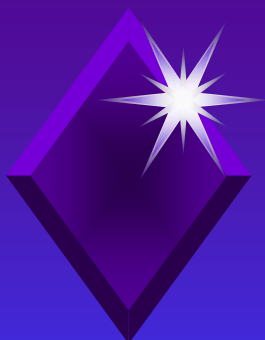


规范化回顾

1. 规范化过程

规范化过程是通过关系模式的分解来实现的。





规范化回顾

2. 模式分解

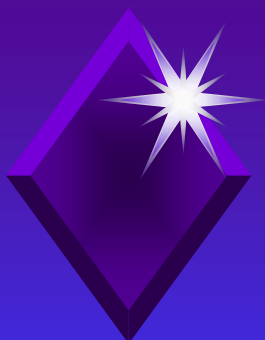
关系模式是五元组 $R (U, D, \text{dom}, F)$ 。

模式的分解包括三个方面：

- (1) 属性的分解
- (2) 函数依赖的分解
- (3) 关系的分解

分解后产生的模式应与原模式等价：

- (1) 分解后子模式的属性集与原模式属性集相同
- (2) 保持函数依赖
- (3) 无损连接性



规 范 化 回 顾

3. 实例

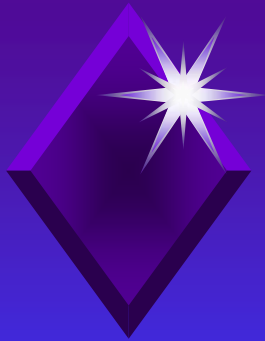
有一个关系模式 $R(A, B, C)$ ，存在函数依赖 $F = \{A \rightarrow B, B \rightarrow C\}$ ，下面的几个分解中，哪一个最好？

$$\rho_1 = \{R_1(A), R_2(B), R_3(C)\}$$

$$\rho_2 = \{R_4(A, B), R_5(A, C)\}$$

$$\rho_3 = \{R_4(A, B), R_6(B, C)\}$$

$$\rho_4 = \{R_5(A, C), R_6(B, C)\}$$



规范化回顾

3. 实例

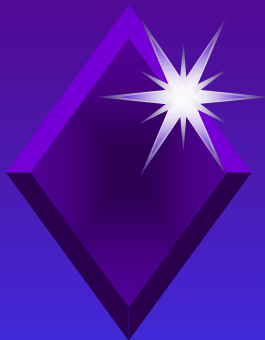
$\rho_1 = \{R_1(A), R_2(B), R_3(C)\}$

$\rho_2 = \{R_4(A, B), R_5(A, C)\}$

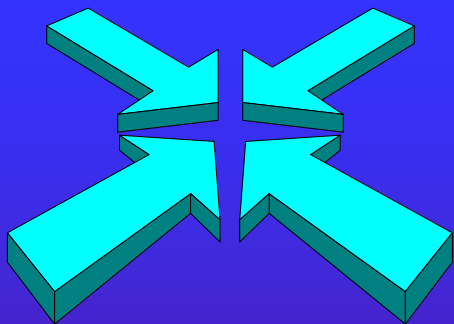
$\rho_3 = \{R_4(A, B), R_6(B, C)\}$

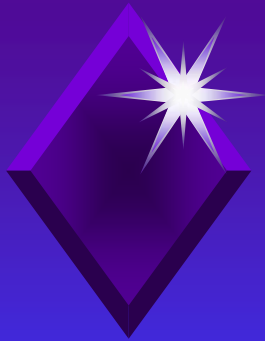
$\rho_4 = \{R_5(A, C), R_6(B, C)\}$

分解	属性等价	保持无损连接性	保持函数依赖性
ρ_1	√	×	×
ρ_2	√	√	×
ρ_3	√	√	√
ρ_4	√	×	×



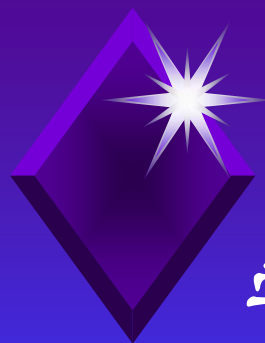
其它设计考虑





其它设计考虑

- 反规范化设计是为了提高查询效率
- 进行反规范化设计后，需要采取措施，处理可能出现的更新异常



本讲主要目标



学完本讲后，你应该能够了解：

1. 第二范式、第三范式、BC范式的定义，以及范式之间的包含关系；
2. 对于一个关系模式，判断其最高属于第几范式，并使用基本的模式分解方法，将1NF分解为2NF，将2NF分解为3NF；
3. 分解的无损连接性和保持函数依赖性的概念；
4. 使用规范化方法设计数据库的逻辑结构时，并不是要使得到的数据库模式都达到最高范式，而是还需要平衡查询效率和更新代价。

问 题 讨 论

1. 实际上，进行规范化的目的是为了**避免更新异常和提高更新代价**；而有时规范化产生的高范式会降低有些查询的代价。这种说法对吗？你在设计数据库模式时，会采取什么方法？

