

第五章 关系数据库理论

什么是好的数据库设计

- 体现客观世界的信息
- 无过度的冗余
- 无插入异常
- 无更新复杂
- 无删除异常

一个异常的例子

Sno	Sname	Sage	ssex	sdept	cno	cname	credit	grade
95001	张三	25	M	CS	C01	数据库	3	95
95001	张三	25	M	CS	C02	网络技术	3	80
95001	张三	25	M	CS	C03	英语	4	76
95002	李四	23	F	MA	C01	数据库	3	80
95002	李四	23	F	MA	C03	英语	4	80

过度冗余——数据重复

更新异常——更新代价大、可能导致数据不一致

删除异常——部分信息的删除可能导致信息的丢失

插入异常——必须有完整信息

数据依赖

- 数据依赖
 - 函数依赖(Functional Dependency,FD)
 - 多值依赖(Multivalued Dependency,MVD)
- 数据依赖是针对数据模式，而不是特定的实例
- 是属性间的关联
- 是一种约束

函数依赖

- 假设给定 X 属性的值，就知道 Y 的值，那么 X 函数决定 Y ，记作 $X \rightarrow Y$
- 如果 R 的两个元组在属性 $A_1, A_2, \dots, A_n$ 上一致，则它们在另一个属性 B 上也一致，那么 $A_1, A_2, \dots, A_n$ 函数决定 B ，记作 $A_1 A_2 \dots A_n \rightarrow B$

函数依赖的种类

- 平凡的函数依赖

- 如果对于函数依赖 $A_1A_2\dots A_n \rightarrow B$, B 是 A 中的一个, 则该函数依赖就是平凡的函数依赖
- 平凡依赖是永恒的
- $sno, sname \rightarrow sname$

函数依赖种类（续）

- 非平凡

- 如果B中至少有一个属性不在A中
- $sno, sname \rightarrow sname, sdept$

- 完全非平凡

- 如果B中没有有一个属性在A中
- $sno, cno \rightarrow grade$

函数依赖规则

- 分解规则

$$A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$$

等价于 $A_1 A_2 \dots A_n \rightarrow B_i$

- 合并规则

$$A_1 A_2 \dots A_n \rightarrow B_i$$

等价于 $A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$

函数依赖规则

- 传递规则

若： $A \rightarrow B$ 、 $B \rightarrow C$

则： $A \rightarrow C$

- 增补规则

若： $A \rightarrow B$

则： $AC \rightarrow BC$

关系的码：从函数依赖的角度

- 如果一个或多个属性的集合 $\{A_1A_2...A_n\}$ 满足如下条件，则该集合为关系R的码：
 - 这些属性函数决定该关系的所有其他属性
 - $\{A_1A_2...A_n\}$ 的任何真子集都不能函数决定R的所有其他属性

关系的码

- **主码**：选定某一个码为主码。
- **主属性**：包含在任何一个候选码中的属性。
- **全码**：整个属性组都是码
- **外码**：该属性或属性组是另外一个关系的主码。

计算属性的闭包

- 属性集 $\{A_1, A_2, \dots, A_n\}$ 的闭包记为 $\{A_1, A_2, \dots, A_n\}^+$
 - 所有可以由 $A_1, A_2, \dots, A_n$ 根据函数依赖关系推导出来的属性集。

闭包的意义

- 检验给定的属性是否可以由已知的属性推导出来。
- 从给定的函数依赖，可以推导出蕴涵的函数依赖。

计算属性闭包的算法

■ 求闭包

closure = X;

repeat until there's no change:{

if there's an FD $U \rightarrow V$ in F such that $U \in \text{closure}$,
then set closure = closure $\cup V$ }

计算属性闭包： 举例

- 例： 对关系R (A,B,C,D,E,F) ,给定函数依赖 $AB \rightarrow C, BC \rightarrow AD, D \rightarrow E, CF \rightarrow B$,求 $\{A, B\}^+$

$$\{AB\}^+ = AB$$

$$\{AB\}^+ = ABC \quad // AB \rightarrow C$$

$$\{AB\}^+ = ABCD \quad // BC \rightarrow AD$$

$$\{AB\}^+ = ABCDE \quad // D \rightarrow E$$

关系的闭包

例：关系 $R(A,B,C)$, $F: \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$,
求关系 $R(A,B,C)$ 的闭包：

$\{A\}^+ + \{B\}^+ + \{C\}^+ + \{AB\}^+ + \{AC\}^+ + \{BC\}^+ +$
 $\{ABC\}^+$

计算关系的闭包,可以知道所有隐含的函数依赖关系,
如 $\{A\}^+ = \{A, B, C\}$ 可以得出: $A \rightarrow C, A \rightarrow B$

计算属性的闭包

- 目的

- 验证某一个函数依赖是否在关系的函数依赖集里

- 例：对于关系 $R(A,B,C,D,E)$

$F=\{A \rightarrow B, B \rightarrow C, CD \rightarrow E\}$ 是否具有 $A \rightarrow E$

只需求出 A 的闭包,是否包含 E ?

范式(Normal Forms)

- 异常的原因
 - 数据依赖的约束
- 解决方法
 - 数据库设计的规范化→分解

范式(Normal Forms)

■ 规范化

- 一个关系满足某个范式所规定的一系列条件时,它就属于该范式
- 可以用规范化要求来设计数据库
- 也可以用来验证设计结果的合理性,用其指导优化过程

■ $1NF \rightarrow 2NF \rightarrow 3NF \rightarrow BCNF \rightarrow 4NF$

范式——第一范式(1NF)

- 当且仅当一个关系R中，每一个元组的每一个属性只含有一个值时，该关系属于第一范式。
- 要求属性是原子的

范式——第二范式(2NF)

- 对于关系R，若 $R \in 1NF$ ，且每一个非主属性完全函数依赖于码，则 $R \in 2NF$ 。

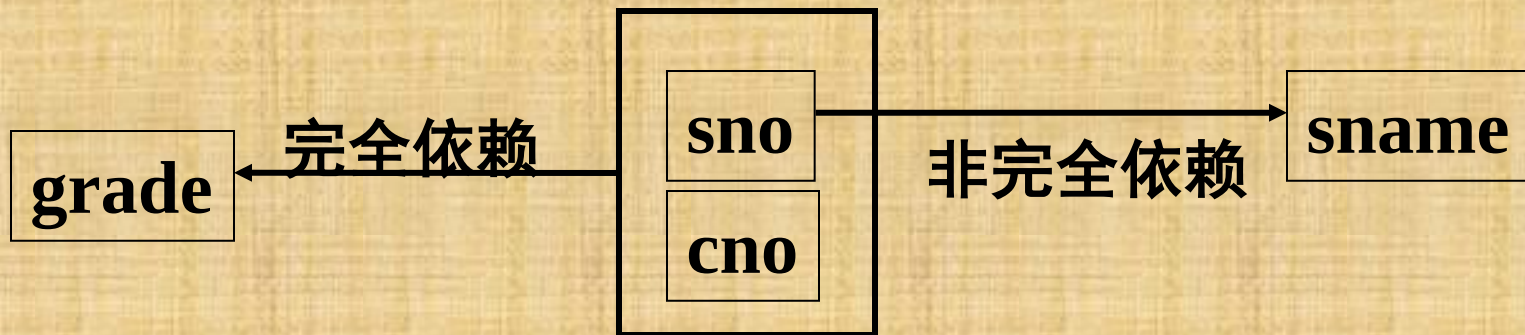
- 不能部分依赖于码

sc(sno,sname,cno,grade)

sno,cno $\rightarrow$ grade

sno $\rightarrow$ sname

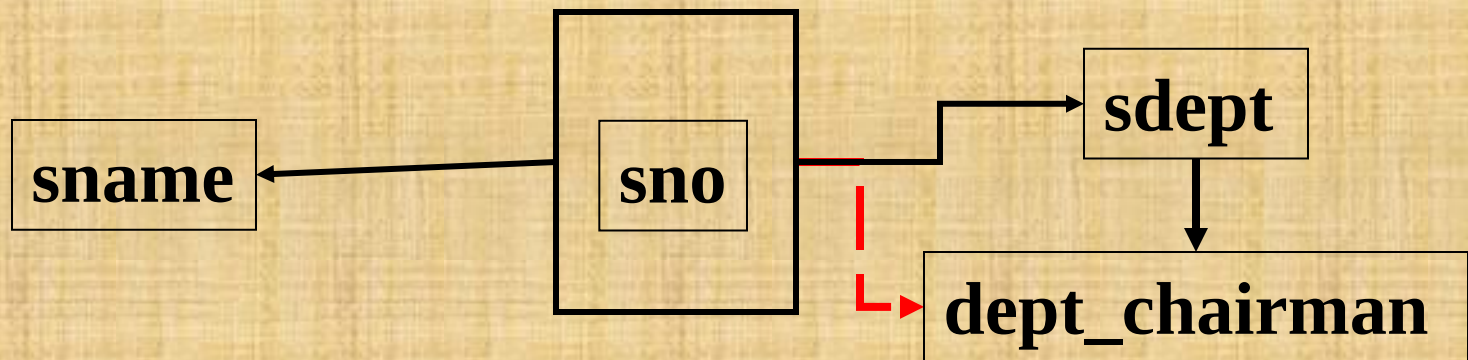
范式——第二范式(2NF)



范式——第三范式(3NF)

- 对于关系R，若 $R \in 2NF$ ，且每个非主属性都不传递依赖于码，则 $R \in 3NF$ 。
- 主属性可以传递依赖于码
student(sno,sname,sdept,dept_chairman)
sno→sname,sdept sdept→dept_chairman
传递依赖： sno→dept_chairman

范式——第三范式(3NF)



范式——Boyce/Codd范式(BCNF)

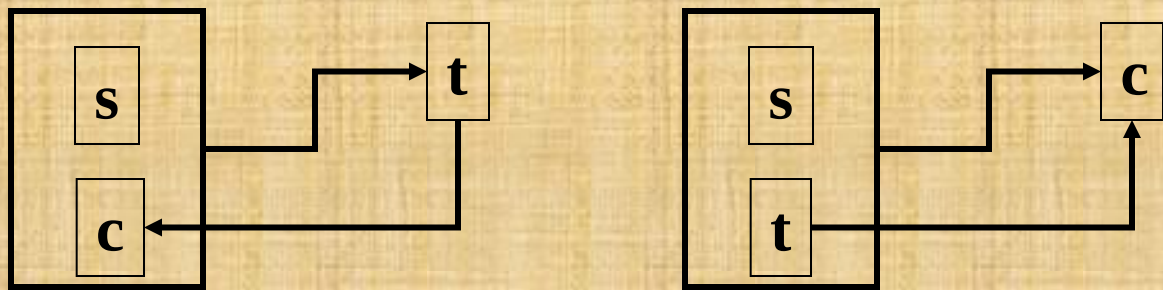
- 对于关系R，若 $R \in 1NF$ ，且对于所有非平凡的函数依赖(右边属性不在左边)，其决定因素是候选码，则 $R \in BCNF$ 。

sct(s,c,t) //学生，课程，教师

$t \rightarrow c$ //每位教师只上一门课

$\underline{s}, \underline{c} \rightarrow t$ $\underline{s}, \underline{t} \rightarrow c$ //每门课有若干位教师

范式——Boyce/Codd范式(BCNF)



$sct \in 3NF$, $sct \notin BCNF$

范式——Boyce/Codd范式(BCNF)

- 最高范式

- BCNF 是基于函数依赖的最高范式
- 但不是数据库模式设计的最高范式

范式优化

- 分解

- $\{A_1, A_2, \dots, A_n\} = \{B_1, B_2, \dots, B_n\} \cup \{C_1, C_2, \dots, C_n\}$

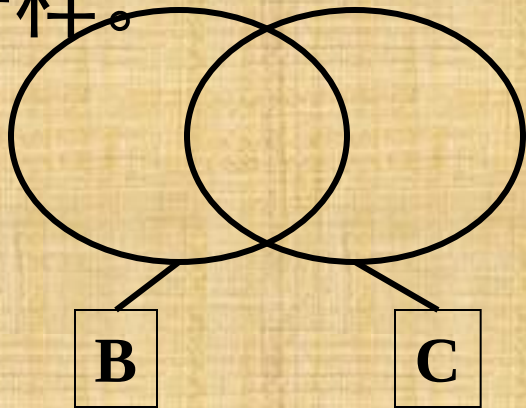
- 分解的要求

- 无损连接：分解后两个关系做连接运算后，元组个数和原来一模一样。

- 保持依赖

$$\underline{F^+ = (F_b \cup F_c)^+}$$

- 范式要求



无损连接的含义

A	B	C
a1	b1	c1
a1	b2	c2
a2	b1	c3



A	B
a1	b1
a1	b2
a2	b1

A	C
a1	c1
a1	c2
a2	c3

A	B	C
a1	b1	c1
a1	b2	c2
a2	b1	c3

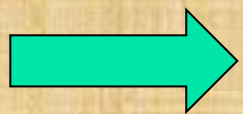


B	C
b1	c1
b2	c2
b1	c3

A	C
a1	c1
a1	c2
a2	c3

范式分解方法

1. 找出一个不符合范式要求的FD ($X \rightarrow Y$)，并使其右端的属性尽可能地多(求出 X^+)
2. 创建新的模式，包含上述FD的所有（左右）属性(X^+ 的所有属性)
3. 创建另一模式，包含前述FD中的决定因素及剩余的所有属性($X \cup (U - X^+)$)
4. 反复1~3，直至符合范式要求



分解后的关系可以无损连接，函数依赖关系不一定保持。

范式优化——练习

Sno	Sname	Sage	ssex	sdept	cno	cname	credit	grade
95001	张三	25	M	CS	C01	数据库	3	95
95001	张三	25	M	CS	C02	网络技术	3	80
95001	张三	25	M	CS	C03	英语	4	76
95002	李四	23	F	MA	C01	数据库	3	80
95002	李四	23	F	MA	C03	英语	4	80

sno→sname, sage, ssex,sdept

cno→cname, grade

sno,cno→grade

范式优化——分解

- 分解并不都是保持依赖关系，在不能实现依赖保持时，宁可降低范式的要求

例：分解SCT(S,C,T) FD为 $T \rightarrow C$; $SC \rightarrow T$; $ST \rightarrow C$

候选码SC/ST

分解为R1(T,C)及R2(T,S)无法保持FD

只能以3NF的形式

分解算法（总结）

- 既保持无损连接，又需要保持函数依赖，分解后的模式可以达到3NF，不一定满足BCNF。算法见书上算法6.3（合成法）
- 保持无损连接，一定可以达到BCNF。

课堂练习: EXAMPLE

关系 $R(ABCD)$, 具有FD: $C \rightarrow D; C \rightarrow A; B \rightarrow C$

请问:

- 1) 找出 R 的候选码;
- 2) 指出 R 符合的最高范式;
- 3) 如果 R 不符合BCNF, 则将其分解为具有依赖保持的BCNF

EXAMPLE: $R(ABCD), \text{ FD: } C \rightarrow D; C \rightarrow A; B \rightarrow C$

- 1) $B^+ = BCDA, C^+ = CDA$, 因此 **B是候选码**
- 2) 主属性是B, 其余均为非主属性, 它们均完全依赖于码; D, A均通过C传递依赖码(B); $C \rightarrow D; C \rightarrow A$ 函数依赖于非码属性, 不是BCNF; **仅满足2NF**
- 3) $\because C \rightarrow D; C \rightarrow A \therefore C^+ = CDA$
 $\therefore$ 分解为 $R_1(CDA), R_2(CB)$
 $\because F_1 \cup F_2 = F \therefore$ **依赖保持**

EXAMPLE

对于关系 $R(A,B,C)$, 有 $A \rightarrow B, B \rightarrow C, A \rightarrow C$

若将其分解为 (A,B) 和 (B,C)

无损连接?

—Yes 依赖保持?

--- Yes $F_{ABC} = F_{AB} \text{ Union } F_{BC}$ ———Include $A \rightarrow C$

多值依赖

- BCNF: 仍有异常

<u>name</u>	<u>street</u>	<u>city</u>	<u>title</u>	<u>year</u>
C.Fisher	123 maple St.	Hollywood	Star Wars	1977
C.Fisher	5 Locust Ln.	Mailbu	Star Wars	1977
C.Fisher	123 maple St.	Hollywood	Empire Strikes Back	1980
C.Fisher	5 Locust Ln.	Mailbu	Empire Strikes Back	1980
C.Fisher	123 maple St.	Hollywood	Return of the Jedi	1983
C.Fisher	5 Locust Ln.	Mailbu	Return of the Jedi	1983

多值依赖:

$x \twoheadrightarrow Z$, 有多个 z , z 与 y 无关

	X	Y	Z
t	$a_1 \dots a_i$	$a_{i+1} \dots a_j$	$a_{j+1} \dots a_n$
u	$a_1 \dots a_i$	$b_{i+1} \dots b_j$	$b_{j+1} \dots b_n$
v	$a_1 \dots a_i$	$a_{i+1} \dots a_j$	$b_{j+1} \dots b_n$
w	$a_1 \dots a_i$	$b_{i+1} \dots b_j$	$a_{j+1} \dots a_n$

多值依赖

- 多值依赖 (MVD)
 - 设 $R(U)$ 是属性集 U 上的一个关系。 X 、 Y 、 Z 是 U 的子集，且 $Z=U-X-Y$ 。关系 $R(U)$ 中多值依赖 $X \twoheadrightarrow Y$ 成立，当且仅当对 $R(U)$ 的任一元组 r ，给定一对 (x,z) 值，有一组 y 的值，这组值仅仅取决于 x 值，而与 z 值无关。
 - 对给定 X 的值，则 Y 的取值与 Z 的取值无关

多值依赖实例

name $\twoheadrightarrow$ **title, year**

<u>name</u>	<u>street</u>	<u>city</u>	<u>title</u>	<u>year</u>
C.Fisher	123 maple St.	Hollywood	Star Wars	1977
C.Fisher	123 maple St.	Hollywood	Empire Strikes Back	1980
C.Fisher	123 maple St.	Hollywood	Return of the Jedi	1983
C.Fisher	5 Locust Ln.	Mailbu	Star Wars	1977
C.Fisher	5 Locust Ln.	Mailbu	Empire Strikes Back	1980
C.Fisher	5 Locust Ln.	Mailbu	Return of the Jedi	1983

多值依赖

- 平凡的多值依赖

- 如果X包含Y或 $X \cup Y = U$ ，则 $X \twoheadrightarrow Y$ 是平凡的多值依赖

- 性质

- If $X \twoheadrightarrow Y$, and $Y \twoheadrightarrow Z$ then $X \twoheadrightarrow Z - Y$
- If $X \twoheadrightarrow Y$, and $Z = U - X - Y$ then $X \twoheadrightarrow Z$
- If $X \rightarrow Y$ then $X \twoheadrightarrow Y$

多值依赖——第四范式

- 对于关系R，若 $R \in 1NF$ ，且所有非平凡的多值依赖，其决定因素是候选码，则 $R \in 4NF$ 。
- 基于数据依赖的最高范式
- 分解要求
 - 无损连接
 - 保持依赖
 - 4NF

多值依赖——第四范式

<u>name</u>	<u>street</u>	<u>city</u>	<u>title</u>	<u>year</u>
C.Fisher	123 maple St.	Hollywood	Star Wars	1977
C.Fisher	123 maple St.	Hollywood	Empire Strikes Back	1980
C.Fisher	123 maple St.	Hollywood	Return of the Jedi	1983
C.Fisher	5 Locust Ln.	Mailbu	Star Wars	1977
C.Fisher	5 Locust Ln.	Mailbu	Empire Strikes Back	1980
C.Fisher	5 Locust Ln.	Mailbu	Return of the Jedi	1983

多值依赖——第四范式

$R(\text{name}, \text{title}, \text{year}, \text{street}, \text{city})$

函数依赖：没有

多值依赖： $\text{name} \twoheadrightarrow \text{title}, \text{year}$ $\text{name} \twoheadrightarrow \text{street}, \text{city}$

码：All Key 满足 BCNF 不满足 4NF

R关系分解为：

$R_1(\text{name}, \text{title}, \text{year})$

$R_2(\text{name}, \text{street}, \text{city})$

码：All Key

满足4NF，为什么？

多值依赖——范式间的联系

- 高层范式包含低层范式
- 分解特性

特性	3NF	BCNF	4NF
消除 FD 引起的冗余	大部分	是	是
消除 MVD 引起的冗余	否	否	是
保持 FD	是	可能	可能
保持 MVD	可能	可能	可能

范式的使用

- 作为数据库模式的优化工具
 - 依据范式来设计模式
 - 用范式来解决异常的问题
 - 主题最小、主题单一化
- 反范式的设计
 - 过多的连接操作
 - 适当增加冗余，以减少连接

课堂练习

R(A,B,C,D) with

a) FD's $AB \rightarrow C$, $C \rightarrow D$, $D \rightarrow A$

b) $AB \rightarrow C$, $BC \rightarrow D$, $CD \rightarrow A$, $AD \rightarrow B$

1. 哪些函数依赖关系违背了BCNF.
2. 分解关系，让所有模式符合BCNF范式
3. 哪些函数依赖关系违背了3NF
4. 分解关系，满足3NF.

解题思路：

$R(A,B,C,D)$ with FD's $AB \rightarrow C, C \rightarrow D, D \rightarrow A$

- 第一步：根据函数依赖关系，求出该关系的码
- 第二步：根据BCNF定义，查看哪些函数依赖了BCNF。
- 第三步：分解该关系。
- 第四步：验证分解后的关系是否满足BCNF范式。
- 第五步：如果不满足，再回到第一步。

码是AB, BC, BD,

$C \rightarrow A, C \rightarrow D,$
 $D \rightarrow A, AC \rightarrow D,$
 $CD \rightarrow A$

$R_1(AC), R_2(BC), R_3(CD)$
Or $R_1(CD), R_2(BC), R_3(AD)$
已满足BCNF

该关系已满足
第三范式

R (A,B,C,D)

$AB \rightarrow C, BC \rightarrow D, CD \rightarrow A, AD \rightarrow B$

哪些函数依赖关系违背了BCNF.

分解关系，让所有模式符合BCNF范式

哪些函数依赖关系违背了3NF

分解关系，满足3NF.

- **答案：该关系模式已经满足BCNF范式和第3范式**

多值依赖练习

- a) $R(A,B,C,D)$ with MVD's $A \twoheadrightarrow B$, $A \twoheadrightarrow C$
- b) $R(A,B,C,D)$ with MVD $AB \twoheadrightarrow C$, and FD $B \rightarrow D$
- 找出所有违背4NF的依赖关系
 - 分解，所有模式达到4NF

A解题思路与答案

$R(A,B,C,D)$ with MVD's $A \twoheadrightarrow B$, $A \twoheadrightarrow C$

- 第一步：求出关系的码：ABCD
- 第二步得出所有的多值依赖关系

$A \twoheadrightarrow CD$, $A \twoheadrightarrow BD$

- 根据4NF定义，违背4NF的依赖关系是：

$A \twoheadrightarrow B$, $A \twoheadrightarrow C$, $A \twoheadrightarrow CD$, $A \twoheadrightarrow BD$

- 分解关系

$R_1(AB)$ $R_2(ACD) \rightarrow R_1(AB)$ $R_{21}(AC)$ $R_{22}(AD)$

$R(A,B,C,D)$ with MVD $AB \twoheadrightarrow C$, and FD $B \rightarrow D$

找出所有违背4NF的依赖关系，分解所有模式达到4NF。

B答案

1) 违背第4NF范式的依赖关系是：

■ $AB \twoheadrightarrow C, B \rightarrow D, \underline{AB \twoheadrightarrow D}, B \twoheadrightarrow AC.$

2) 分解关系为： $R_1(A,B,C)$ $R_2(B,D)$ 满足第4范式。