

- 6.1 问题的提出
- 6.2 规范化
- 6.3 数据依赖的公理系统

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 范式

6.2.1 函数依赖

- 函数依赖
- 平凡函数依赖与非平凡函数依赖
- 完全函数依赖与部分函数依赖
- 传递函数依赖

- 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 和 Y 是 U 的子集。若对于 $R(U)$ 的任意一个可能的关系 r ， r 中不可能存在两个元组在 X 上的属性值相等，而在 Y 上的属性值不等，则称“ X 函数确定 Y ”或“ Y 函数依赖于 X ”，记作 $X \rightarrow Y$ 。

$$X : Y \left\{ \begin{array}{l} 1 : 1 \quad X \rightarrow Y \quad Y \rightarrow X \\ m : 1 \quad X \rightarrow Y \\ m : n \end{array} \right.$$

Student(Sno, Sname, Ssex, Sage, Sdept)

$Sno \rightarrow Sname$ $Sno \rightarrow Ssex$ $Sno \rightarrow Sage$ $Sno \rightarrow Sdept$

假设没有重名的情况

$Sname \rightarrow Sno$ $Sname \rightarrow Ssex$ $Sname \rightarrow Sage$ $Sname \rightarrow Sdept$

若 $X \rightarrow Y$ ，并且 $Y \rightarrow X$ ，则记为 $X \longleftrightarrow Y$ 。 $Sno \longleftrightarrow Sname$

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

- 平凡函数依赖与非平凡函数依赖

$X \rightarrow Y$, 但 $Y \not\subseteq X$ 则称 $X \rightarrow Y$ 是非平凡的函数依赖。

$X \rightarrow Y$, 但 $Y \subseteq X$ 则称 $X \rightarrow Y$ 是平凡的函数依赖。

对于任一关系模式，平凡函数依赖都是必然成立的，它不反映新的语义。
若不特别声明，我们总是讨论非平凡函数依赖。

- 完全函数依赖与部分函数依赖

定义6.2 在 $R(U)$ 中, 如果 $X \rightarrow Y$, 并且对于 X 的任何一个真子集 X' , 都有 $X' \nrightarrow Y$, 则称 Y 对 X 完全函数依赖, 记作 $X \xrightarrow{F} Y$ 。

若 $X \rightarrow Y$, 但 Y 不完全函数依赖于 X , 则称 Y 对 X 部分函数依赖, 记作 $X \xrightarrow{P} Y$

[例] 在关系 $SC(Sno, Cno, Grade)$ 中, 有:

由于: $Sno \nrightarrow Grade$, $Cno \nrightarrow Grade$,

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

- 传递函数依赖

定义6.3 在 $R(U)$ 中, 如果 $X \rightarrow Y (Y \not\subseteq X)$, $Y \not\rightarrow X$, $Y \rightarrow Z$, $Z \not\subseteq Y$, 则称 Z 对 X 传递函数依赖(transitive functional dependency)。

记为： $X \xrightarrow{T} Z$ 。

[例] 在关系 $Std(Sno, Sdept, Mname)$ 中, 有：

$Sno \rightarrow Sdept, Sdept \rightarrow Mname$

$Mname$ 传递函数依赖于 Sno

- 候选码

定义6.4 设 K 为 $R\langle U, F \rangle$ 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)。

如果 U 部分函数依赖于 K , 即 $K \xrightarrow{P} U$, 则 K 称为超码 (Surpkey) 。

候选码是最小的超码, 即 K 的任意一个真子集都不是候选码。

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)。

主属性与非主属性

包含在任何一个候选码中的属性，称为主属性

不包含在任何码中的属性称为非主属性或非码属性

全码：整个属性组是码，称为全码 (All-key)

[例6.3] $R(P,W,A)$ P：演奏者 W：作品 A：听众

一个演奏者可以演奏多个作品

某一作品可被多个演奏者演奏

听众可以欣赏不同演奏者的不同作品

码为(P,W,A)，即All-Key

定义6.5 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**外码** (Foreign key)。

6.2.2 范式

- 范式是符合某一种级别的关系模式的集合

第一范式(1NF)

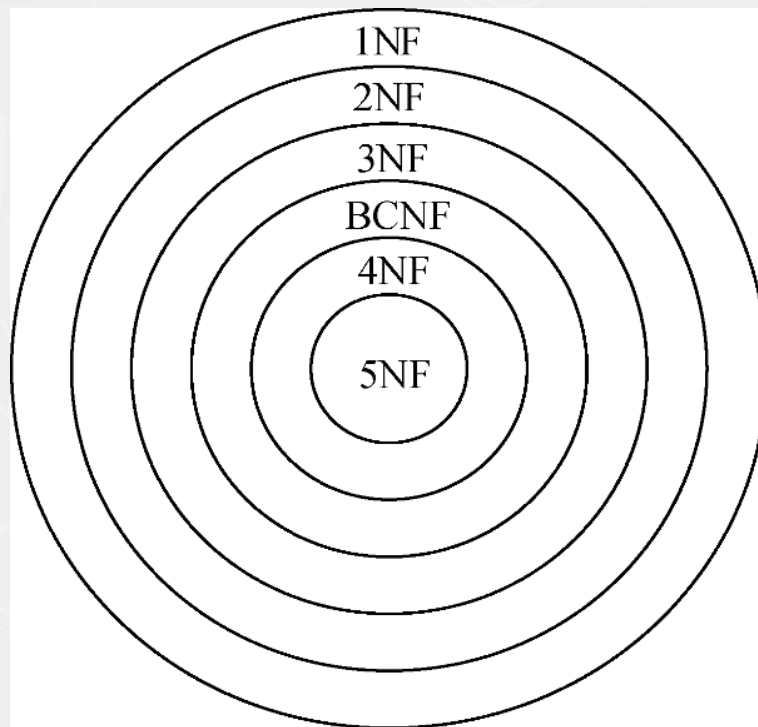
第二范式(2NF)

第三范式(3NF)

BC范式(BCNF)

第四范式(4NF)

第五范式(5NF)



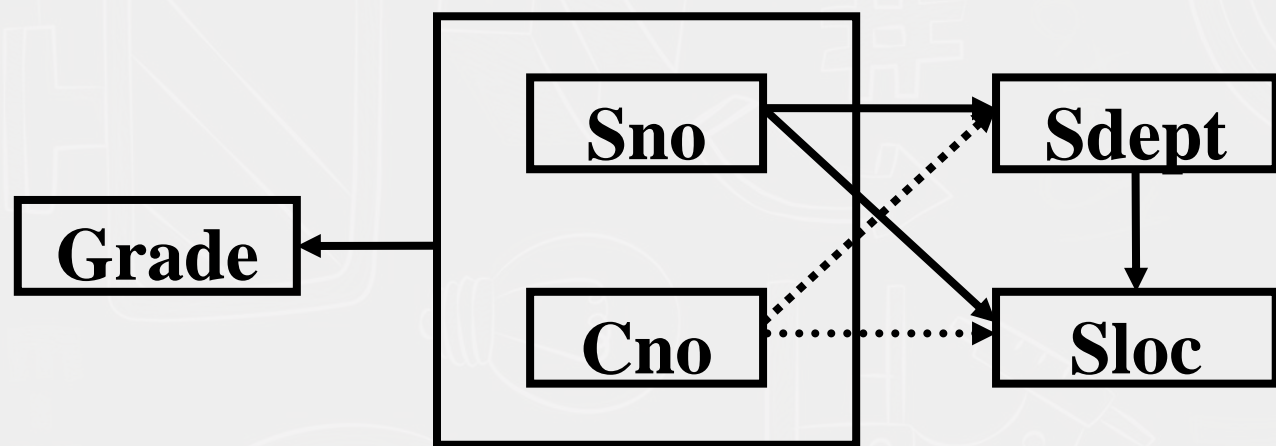
■ 关系要符合一个最基本的条件：每个分量必须是不可分开的数据项。

满足了这个条件的关系模式就属于第一范式

(1NF)，即 $R \in 1NF$

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任何一个候选码，则 $R \in 2NF$

[例6.4] S-L-C(Sno,Sdept,Sloc,Cno,Grade), Sloc为学生的住处，并且每个系的学生住在同一个地方。S-L-C的码为(Sno,Cno)。



非主属性Sdept、Sloc并不完全依赖于码

关系模式S-L-C不属于2NF

❖ 一个关系模式不属于2NF，会产生以下问题：

■ 插入异常

- 如果插入一个新学生，但该生未选课，即该生无Cno，由于插入元组时，必须给定码值，因此插入失败。

■ 删除异常

- 如果S4只选了一门课C3，现在他不再选这门课，则删除C3后，整个元组的其他信息也被删除了。

■ 修改复杂

- 如果一个学生选多门课，则Sdept, Sloc被存储了多次。如果该生转系，则需要修改所有相关的Sdept和Sloc，造成修改的复杂化。

❖ 出现这种问题的原因

■ 例子中有两类非主属性：

- 一类如Grade，它对码完全函数依赖
- 另一类如Sdept、Sloc，它们对码不是完全函数依赖

❖ 解决方法：

■ 用投影分解把关系模式S-L-C分解成两个关系模式

- SC(Sno,Cno,Grade)
- S-L(Sno,Sdept,Sloc)

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

- SC没有传递依赖, 因此 $SC \in 3NF$

- S-L中 $Sno \rightarrow Sdept$ ($Sdept \not\rightarrow Sno$), $Sdept \rightarrow Sloc$,
可得 $Sno \rightarrow Sloc$ 。

- 解决的办法是将S-L分解成

- $S-D(Sno, Sdept) \in 3NF$

- $D-L(Sdept, Sloc) \in 3NF$

- ❖ BCNF (Boyce Codd Normal Form) :修正的第三范式, 有时也称为扩充的第三范式。
- ❖ 定义6.8 设关系模式 $R\langle U, F \rangle \in 1NF$, 若 $X \rightarrow Y$ 且 $Y \notin X$ 时 X 必含有码, 则 $R\langle U, F \rangle \in BCNF$ 。
- ❖ 换言之, 在关系模式 $R\langle U, F \rangle$ 中, 如果每一个决定属性集都包含候选码, 则 $R \in BCNF$ 。

❖ BCNF的关系模式所具有的性质

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

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。

[例6.5]考察关系模式 $C(Cno, Cname, Pcno)$

- 它只有一个码 Cno ，没有任何属性对 Cno 部分依赖或传递依赖，所以 $C \in 3NF$ 。
- 同时 C 中 Cno 是唯一的决定因素，所以 $C \in BCNF$ 。

[例6.6] 关系模式 $S(Sno, Sname, Sdept, Sage)$,

- 假定 $Sname$ 也具有唯一性, 那么 S 就有两个码, 这两个码都由单个属性组成, 彼此不相交。
- 其他属性不存在对码的传递依赖与部分依赖, 所以 $S \in 3NF$ 。
- 同时 S 中除 Sno , $Sname$ 外没有其他决定因素, 所以 S 也属于BCNF。

[例6.7] 关系模式SJP(S,J,P)中，S是学生，J表示课程，P表示名次。每一个学生选修每门课程的成绩有一定的名次，每门课程中每一名次只有一个学生（即没有并列名次）。

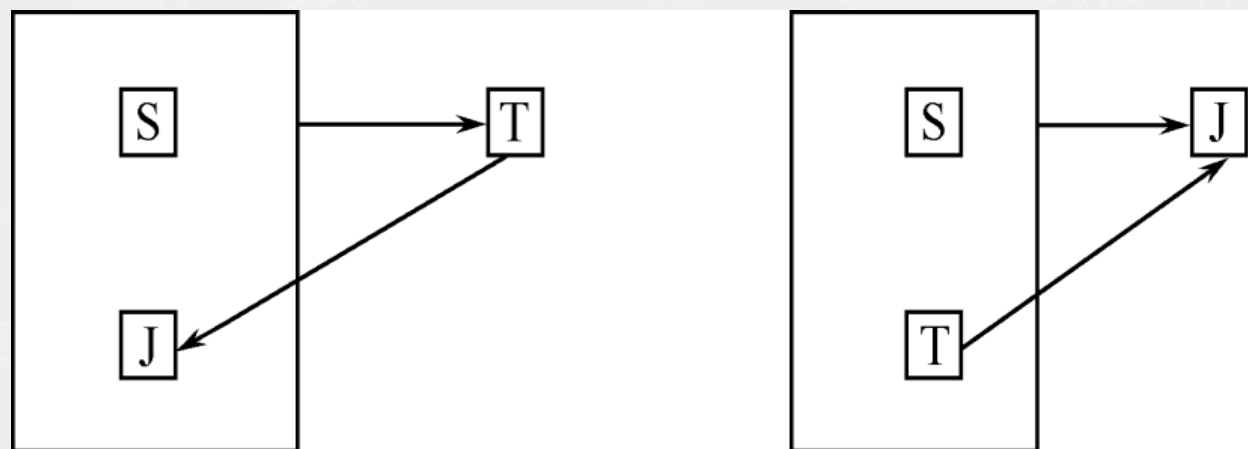
- 由语义可得到函数依赖： $(S,J) \rightarrow P$ ； $(J,P) \rightarrow S$
- (S,J) 与 (J,P) 都可以作为候选码。
- 关系模式中没有属性对码传递依赖或部分依赖，所以 $SJP \in 3NF$ 。
- 除 (S,J) 与 (J,P) 以外没有其他决定因素，所以 $SJP \in BCNF$ 。

❖[例6.8] 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。

■ 由语义可得到函数依赖： $(S,J) \rightarrow T$ ； $(S,T) \rightarrow J$ ； $T \rightarrow J$

■ 因为没有任何非主属性对码传递依赖或部分依赖，
 $STJ \in 3NF$ 。

■ 因为T是决定因素， 而T不包含码， 所以 $STJ \in BCNF$



- 3NF和BCNF是在函数依赖的条件下对模式分解所能达到的分离程度的测度。
 - 一个模式中的关系模式如果都属于BCNF，那么在函数依赖范畴内，它已实现了彻底的分离，已消除了插入和删除的异常。
 - 3NF的“不彻底”性表现在可能存在主属性对码的部分依赖和传递依赖。

1NF



消除非主属性对码的部分函数依赖

2NF



消除非主属性对码的传递函数依赖

3NF



消除主属性对码的部分和传递函数依赖

BCNF