

06 关系数据理论

- 6.1 问题的提出
- 6.2 规范化
 - 6.2.1 函数依赖
 - 6.2.1.1 函数依赖
 - 6.2.1.2 平凡函数依赖与非平凡函数依赖
 - 6.2.1.3 完全函数依赖与部分函数依赖
 - 6.2.1.4 传递函数依赖
 - 6.2.2 码
 - 6.2.3 范式
 - 6.2.4 2NF
 - 6.2.5 3NF
 - 6.2.6 BCNF
 - 6.2.7 多值依赖
 - 6.2.8 4NF
- 6.4 模式的分解

6.1 问题的提出

举例：有三个属性的工资表

| 姓名 | 级别 | 工资 |
|------|------|------|
| ---- | ---- | ---- |
| A | 1 | 900 |
| B | 2 | 1000 |
| C | 1 | 900 |
| D | 3 | 1200 |

问题：

- 数据冗余度大
 - 浪费存储空间
 - 容易造成数据的不一致
- 插入与删除异常
 - 无法插入某部分信息
 - 删除掉不应删除的信息

解决办法：

- 将表分解为两个模式表达：职工级别（姓名、级别），级别工资（级别、工资）
- 数据量减少、表达能力强、修改方便

关系数据库逻辑设计

- 针对具体问题，如何构造一个适合于它的数据模式

- 数据库逻辑设计的工具——关系数据库的规范化理论

关系模式由五部分组成： **$R(U, D, DOM, F)$** ——一个五元组

- R: 关系名R是符号化的元组语义
- U: 组成该关系的属性名集合，一组属性
- D: 属性组U中属性所来自的域
- DOM: 属性向域的映射
- F: 属性间U上的一组数据依赖
- 一般将关系模式看作三元组—— **$R\langle U, F \rangle$** ，当且仅当U上的一个关系r满足F时，r称为关系模式 $R\langle U, F \rangle$ 的一个关系
- 作为二维表，关系要符合一个最基本条件：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属于**第一范式 1NF**

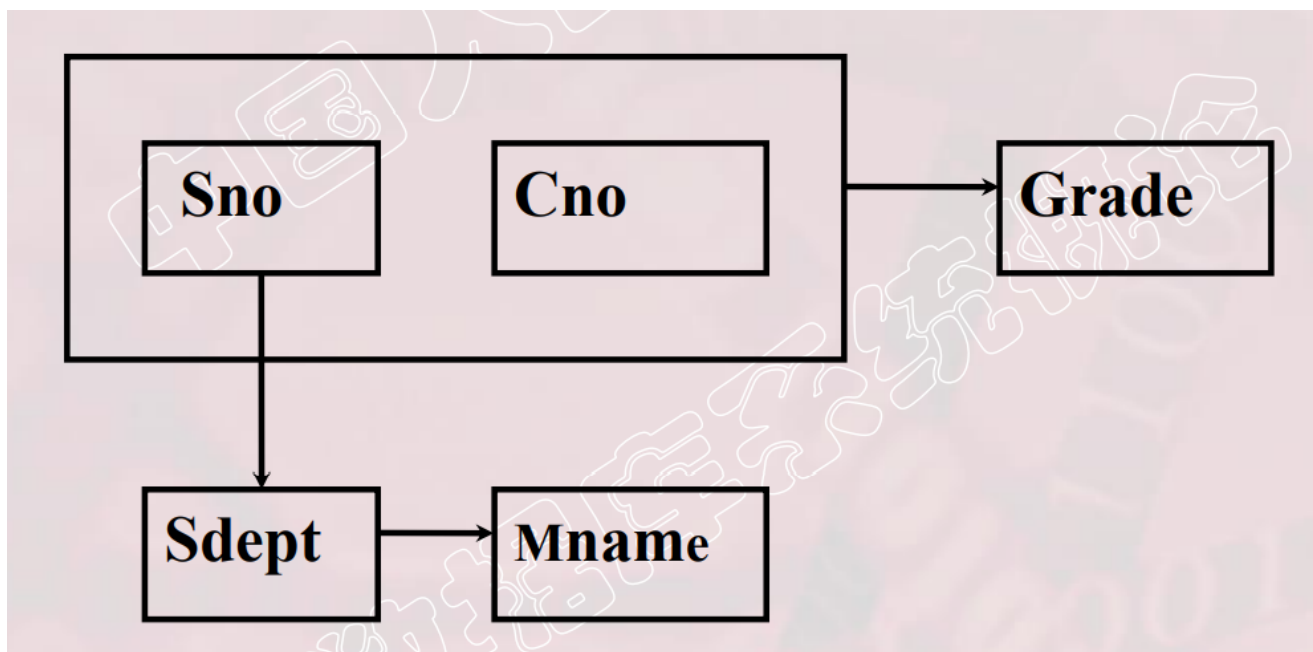
数据依赖

- 定义属性**值**间的相互关联：主要体现于**值的相等与否**
- 数据库模式设计的关键
- 是一个关系内部属性与属性之间的约束关系
- 是现实世界属性间相互联系的抽象
- 是数据内在的性质
- 是语义的体现
- **主要类型**：函数依赖、多值依赖

函数依赖普遍存在于现实生活中

- 描述一个学生关系，可以有学号、姓名、系名等属性
 - 一个学号只对应一个学生，一个学生只在一个系中学习
 - “学号”值确定以后，学生的姓名及所在系的值就被唯一确定
- **$Sname=f(Sno), Sdept=f(Sno)$**
 - Sno函数决定Sname、Sdept
 - $Sno \rightarrow Sname, Sno \rightarrow Sdept$

数据依赖对关系模式的影响



问题:

- 数据冗余太大: 浪费存储空间, Mname重复出现
- 更新异常: 数据冗余, 更新数据时维护数据完整性代价很大, 修改Mname过程繁琐
- 插入异常: 若系中没有学生, 则Mname信息无法存入数据库
- 删除异常: 删除学生的同时也会删除系主任的信息

原因: 由存在于模式中的 **某些数据依赖** 引起的

解决方法: 通过 **分解** 关系模式 **用规范化理论改造关系模式** 来消除其中不合适的数据依赖

分解 Student 关系模式:

- S (Sno, Sdept, Sno → Sdept)
- SC (Sno, Cname, Grade, (Sno, Cno) → Grade)
- DEPT (Sdept, Mname, Sdept → Mname)

6.2 规范化

规范化理论 正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

6.2.1 函数依赖

6.2.1.1 函数依赖

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 $X \rightarrow Y$

函数依赖说明

- **所有关系实例** 均要满足
- **语义范畴** 的概念, 只能根据数据的语义来确定一个函数依赖

- 数据库设计者可以对现实世界作强制的规定

6.2.1.2 平凡函数依赖与非平凡函数依赖

在关系模式 $R(U)$ 中，对于 U 的子集 X 和 Y

如果 $X \rightarrow Y$ ，但 Y 不是 X 的子集，则称 $X \rightarrow Y$ 是 非平凡的函数依赖

若 $X \rightarrow Y$ ，但 Y 是 X 的子集，则称 $X \rightarrow Y$ 是 平凡的函数依赖

若 $X \rightarrow Y$ ，则 X 称为这个函数依赖的 决定属性组，也称为决定因素

6.2.1.3 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 $X \xrightarrow{P} Y$

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

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

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

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

6.2.1.4 传递函数依赖

在 $R(U)$ 中，如果 $X \rightarrow Y (Y \not\subseteq X)$ ， $Y \nrightarrow X$ ， $Y \rightarrow Z$ ， $Z \not\subseteq Y$ ，则称 Z 对 X 传递函数依赖

如果 $Y \rightarrow X$ ，即 $X \leftarrow Y$ ，则 Z 直接依赖于 X ，而不是传递函数依赖

举例：在关系 $Std(Sno, Sdept, Mname)$ 中，有： $Sno \rightarrow Sdept$ ， $Sdept \rightarrow Mname$ ，
 $Mname$ 传递函数依赖于 Sno

6.2.2 码

若关系中的某一属性组的值能唯一地标识一个元组，则称该属性组为候选码

定义： 设 K 为 $R\langle U, F \rangle$ 中的属性或属性组合，若 U 完全依赖于 K ，则称 K 为 R 的候选码，若候选码多于一个，则选定其中的一个作为主码

主属性和非主属性

- 包含在任何候选码中的属性，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

定义： 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

主码与外码一起提供了表示关系间联系的手段

6.2.3 范式

范式是符合某一种级别的关系模式的集合

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

$1NF > 2NF > 3NF > BCNF \dots > 5NF$

某一关系模式 R 为第 n 范式，记为 $R \in nNF$

规范化： 一个低一级范式的关系模式，通过模式分解可以转换为若干个高一级范式的关系模式的集合

6.2.4 2NF

1NF：

- 如果一个关系模式 R 的所有属性都是不可分的基本数据，则 $R \in 1NF$
- 第一范式是对关系模式的最起码的要求，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

举例：

- 关系模式 $S-L-C(Sno, Sdept, Sloc, Cno, Grade)$
- 函数依赖： $(Sno, Cno) \rightarrow F Grade; Sno \rightarrow Sdept, (Sno, Cno) \rightarrow P Sdept; Sno \rightarrow Sloc, (Sno, Cno) \rightarrow P Sloc; Sdept \rightarrow Sloc$
- (Sno, Cno) 为 $S-L-C$ 的码，满足第一范式，非主属性 $Sdept$ 和 $Sloc$ 部分函数依赖于码 (Sno, Cno)
- 这不是一个好的关系模式
 - 插入异常：未选课的学生无法插入
 - 删除异常：删除选课信息后可能将整个元组全部删除
 - 数据冗余
 - 修改复杂：如果一个学生选了多门课，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
- 原因：有两类非主属性，一类对码完全函数依赖，另一类对码部分函数依赖

- **解决方法：** 将S-L-C分解为两个关系模式，以消除部分函数依赖
 - SC(Sno, Cno, Grade)
 - S-L(Sno, Sdept, Sloc)
 - 这样非主属性对码都是完全函数依赖

2NF

- 若 $R \in 1NF$ ，且 每一个非主属性完全函数依赖于码，则 $R \in 2NF$
- 若一个数据库模式是二范式，则其每一个关系模式都是二范式
- 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，**但并不能完全消除各种异常情况**

6.2.5 3NF

定义： 设关系模式 $R<U, F> \in 1NF$ ，若R中不存在这样的码X、属性组Y及非主属性Z（Z不是Y的子集），使得 $X \rightarrow Y$ ， $Y \rightarrow Z$ 成立， $Y \not\rightarrow X$ ，则称 $R<U, F> \in 3NF$

若 $R \in 3NF$ ，则每一个非主属性 **既不部分依赖于码也不传递依赖于码**

解决方法： 采用投影分解法，将S-L分解为两个关系模式，以消除传递函数依赖

6.2.6 BCNF

定义： 关系模式 $R<U, F> \in 1NF$ ，若 $X \rightarrow Y$ 且 Y 不是 X 的子集时，X必含有码，则 $R<U, F> \in BCNF$

等价于 每一个决定属性因素都包含码

若 $R \in BCNF$

- 所有非主属性对每一个码都是完全函数依赖
- 所有的主属性对每一个不包含它的码，也是完全函数依赖
- 没有任何属性完全函数依赖于非码的任何一组属性

若 $R \in 3NF$ 且 R 只有一个候选码，则 $R \in BCNF$

3NF 与 BCNF 是以函数依赖为基础的关系模式规范化程度的测度

6.2.7 多值依赖

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

等价于： 在 $R(U)$ 的任一关系r中，如果存在元组t, s使得 $t[X] = s[X]$ ，那么 **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， $X \twoheadrightarrow Y$

平凡多值依赖和非平凡多值依赖

- 若 $X \twoheadrightarrow Y$ ，而 $Z = \emptyset$ ，则称 $X \twoheadrightarrow Y$ 为 **平凡的多值依赖**
- 否则为 **非平凡的多值依赖**

多值依赖

- **对称性**：若 $X \twoheadrightarrow Y$ ，则 $X \twoheadrightarrow Z$ ，其中 $Z = U - X - Y$
- **传递性**：若 $X \twoheadrightarrow Y$ ， $Y \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Z - Y$
- **函数依赖是多值依赖的特殊情况**：若 $X \rightarrow Y$ ，则 $X \twoheadrightarrow Y$

多值依赖与函数依赖的区别

- **多值依赖的有效性和属性集的范围有关**
 - 若 $X \twoheadrightarrow Y$ 在 U 上成立，则在 W ($XY \subset W \subset U$) 上一定成立；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
 - 原因：多值依赖的定义中不仅涉及属性组 X 和 Y ，而且涉及 U 中其余属性 Z
 - 函数依赖的有效性仅决定于 X ， Y 这两个属性集的值
- **若函数依赖 $X \rightarrow Y$ 在 $R(U)$ 上成立，则对于任何 $Y' \subset Y$ 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**

6.2.8 4NF

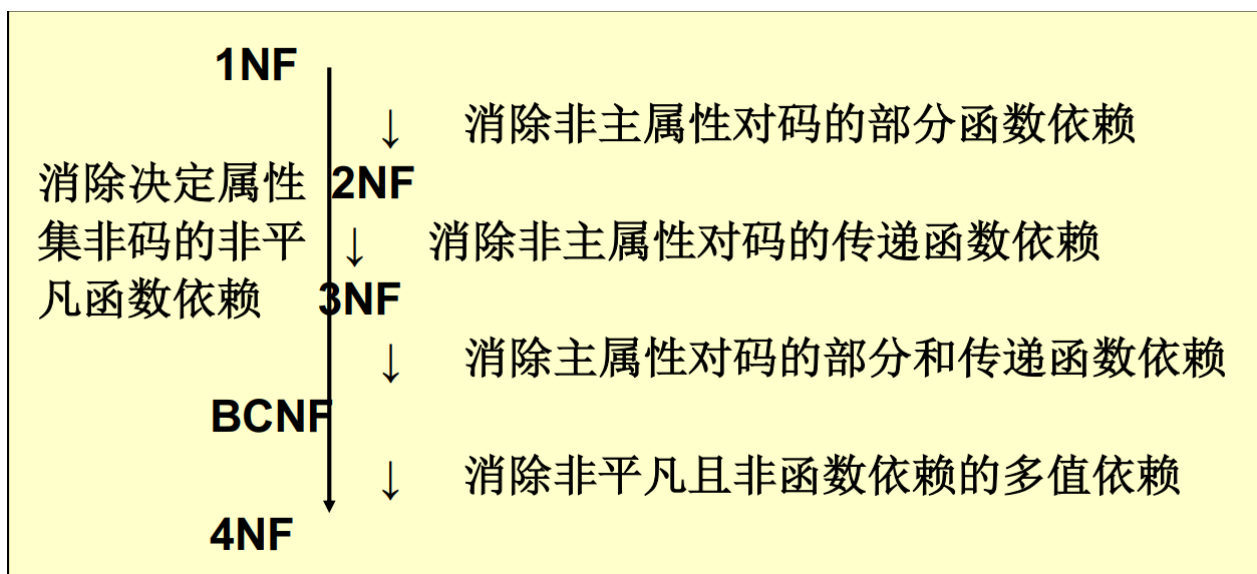
定义：关系模式 $R \langle U, F \rangle \in 1NF$ ，如果对于 R 的每个非平凡多值依赖 $X \twoheadrightarrow Y$ ， X 都含有码，则 $R \in 4NF$

如果 $R \in 4NF$ ，则 $R \in BCNF$

- 不允许有非平凡且非函数依赖的多值依赖
- 允许的非平凡多值依赖是函数依赖
- 允许的多值依赖是平凡的多值依赖

规范化

- **目的**：尽量消除插入、删除异常，修改复杂，数据冗余
- **基本思想**：逐步消除数据依赖中不合适的部分，**实质**：概念的单一化
- 不是规范化程度越高的关系模式就越好



6.4 模式的分解

- 把低一级的关系模式分解为若干个高一级的关系模式的方法不是唯一的
- 只有能够保证分解后的关系模式与原关系模式等价，分解方法才有意义

模式分解的标准 三种模式分解等价的定义

1. 分解具有无损连接性

- 分解后不能丢失信息，即分解后的关系可以通过自然连接恢复为原来的关系
- 若一个关系模式被分解为若干个关系模式，而其自然连接的结果相等，则称R的这个分解具有无损连接性
- 只有具有无损连接性的分解才能保证不丢失信息

2. 分解要保持函数依赖

- 若一个关系模式被分解为若干个关系模式，则其所蕴含的函数依赖一定也由分解得到的函数依赖所蕴含，则称R的这个分解是保持函数依赖的

3. 分解既要保持函数依赖，又要具有无损连接性

- 分解具有无损连接性和分解保持函数依赖是两个互相独立的标准
- 具有无损连接性 —— 保证不丢失信息
- 保持了函数依赖 —— 减轻或解决各种异常情况