

第六章 关系数据理论

6.1 关系模式及范式

- 关系模式由五部分组成，是一个五元组：

$$R = (U, D, DOM, F)$$

- 关系名 R 是符号化的元组语义
 - U 为一组属性
 - D 为属性组 U 中的属性所来自的域
 - DOM 为属性到域的映射
 - F 为属性组 U 上的一组数据依赖
- 由于 D 、 DOM 与模式设计关系不大，因此可以把关系模式看作一个三元组： $R < U, F >$
 - 当且仅当 U 上的一个关系 r 满足 F 时， r 称为关系模式 $R < U, F >$ 的一个关系
 - 作为二维表，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）
- 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
 - 通过属性间值的相等与否体现出来的数据间相关联系
- 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
- 数据依赖的主要类型
 - 函数依赖（Functional Dependency，简记为 FD）
 - 多值依赖（Multi-Valued Dependency，简记为 MVD）
- 第一范式的问题有
 - 数据冗余
 - 更新异常
 - 插入异常
 - 删除异常

6.2 规范化

6.2.1 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 $X \rightarrow Y$ ， X 称为这个函数依赖的决定因素

- 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
- 函数依赖是语义范畴的概念。只能根据数据的语义来确定函数依赖
- 数据库设计者可以对现实世界作强制的规定

在关系模式 $R(U)$ 中, 对于 U 的子集 X 和 Y ,

- $X \rightarrow Y$, 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。因此若不特别声明, 我们总是讨论非平凡函数依赖。

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 $X \xrightarrow{P} Y$

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, 记为:

$X \xrightarrow{\text{传递}} Z$

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

6.2.2 码

设 K 为 $R < U, F >$ 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

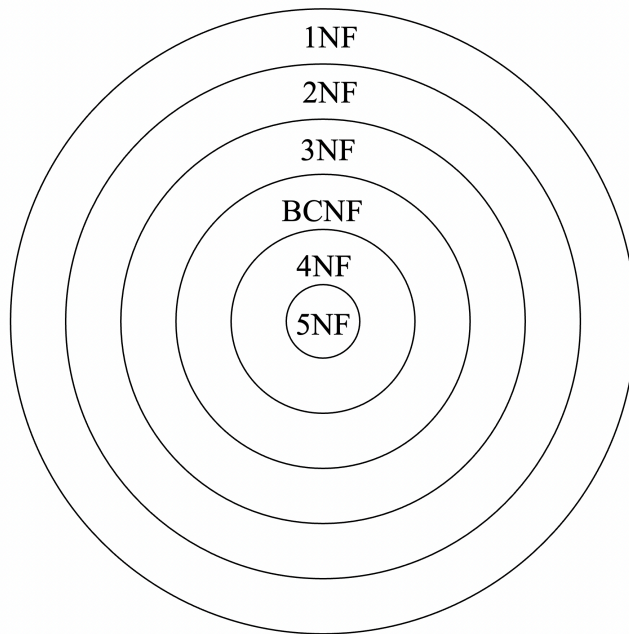
- 如果 U 部分函数依赖于 K , 即 $K \xrightarrow{P} U$, 则 K 称为超码
 - 候选码的超集是超码, 候选码的真子集一定不是超码
- 若关系模式 R 有多个候选码, 则选定其中的一个做为主码
- 主属性与非主属性
 - 包含在任何一个候选码中的属性, 称为主属性
 - 不包含在任何码中的属性称为非主属性或非码属性
- 全码: 整个属性组是码, 称为全码

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 的外部码也称外码

主码与外部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
- 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, 这种过程就叫规范化



6.2.4 2NF

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 $R \in 2NF$

6.2.5 3NF

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 $Z (Z \not\subseteq Y)$ ，使得 $X \rightarrow Y$ ， $Y \rightarrow Z$ 成立， $Y \nrightarrow X$ 不成立，则称 $R \langle U, F \rangle \in 3NF$

6.2.6 BCNF

- BCNF 由 Boyce 和 Codd 提出，比 3NF 更进了一步。通常认为 BCNF 是修正的第三范式，有时也称为扩充的第三范式
- 设关系模式 $R \langle U, F \rangle \in 1NF$ ，若 $X \rightarrow Y$ 且 $Y \not\subseteq X$ 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

6.2.7 多值依赖

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

- 平凡多值依赖和非平凡的多值依赖
 - 若 $X \twoheadrightarrow Y$ ，而 $Z = \emptyset$ ，即 Z 为空，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

多值依赖的另一个等价的定义：在 $R(U)$ 的任一关系 r 中，如果存在元组 t, s 使得 $t[X] = s[X]$ ，那么就必然存在元组 $w, v \in r$ ，（ w, v 可以与 s, t 相同），使得 $w[X] = v[X] = t[X]$ ，而 $w[Y] = t[Y], w[Z] = s[Z], v[Y] = s[Y], v[Z] = t[Z]$ （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， $Z = U - X - Y$ ）

多值依赖的性质：

- 多值依赖具有对称性。即若 $X \twoheadrightarrow Y$ ，则 $X \twoheadrightarrow Z$ ，其中 $Z = U - X - Y$
- 多值依赖具有传递性。即若 $X \twoheadrightarrow Y, Y \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Z - Y$
- 函数依赖是多值依赖的特殊情况。即若 $X \rightarrow Y$ ，则 $X \twoheadrightarrow Y$
- 若 $X \twoheadrightarrow Y, X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow YZ$
- 若 $X \twoheadrightarrow Y, X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Y \cap Z$
- 若 $X \twoheadrightarrow Y, X \twoheadrightarrow Z$ ，则 $X \twoheadrightarrow Y - Z, X \twoheadrightarrow Z - Y$

多值依赖与函数依赖的区别

- 多值依赖的有效性与属性集的范围有关，函数依赖 $X \rightarrow Y$ 的有效性仅决定于 X, Y 这两个属性集的值
 - 若 $X \twoheadrightarrow Y$ 在 U 上成立，则在 $W(XY \subseteq W \subseteq U)$ 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
 - 一般地，在 $R(U)$ 上若有 $X \twoheadrightarrow Y$ 在 $W(W \subset U)$ 上成立，则称 $X \twoheadrightarrow Y$ 为 $R(U)$ 的嵌入型多值依赖
- 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

关系模式 $R < U, F > \in 1NF$ ，如果对于 R 的每个非平凡多值依赖 $X \twoheadrightarrow Y (Y \not\subseteq X)$ ， X 都含有码，则 $R < U, F > \in 4NF$

4NF 就是限制关系模式的属性之间不允许有非平凡且非函数依赖的多值依赖。4NF 所允许的非平凡多值依赖实际上是函数依赖