

第五章 关系数据理论

第五章讲解关系数据理论。这是关系数据库的又一个重点。学习本章的目的有两个。一个是理论方面的,本章用更加形式化的关系数据理论来描述和研究关系模型。另一个是实践方面的,关系数据理论是我们进行数据库设计的有力工具。因此,人们也把关系数据理论中的规范化理论称为数据库设计理论,有的书把它放在数据库设计部分介绍以强调它对数据库设计的指导作用。

一、基本知识点

本章讲解关系数据理论,内容理论性较强,分为基本要求部分(《概论》5.1~5.3)和高级部分(《概论》5.4)。前者是计算机大学本科学生应该掌握的内容;后者是研究生应该学习掌握的内容。

需要了解的:什么是一个“不好”的数据库模式;什么是模式的插入异常和删除异常;规范化理论的重要意义。

需要牢固掌握的:关系的形式化定义;数据依赖的基本概念(函数依赖、平凡函数依赖、非平凡的函数依赖、部分函数依赖、完全函数依赖、传递函数依赖的概念,码、候选码、外码的概念和定义,多值依赖的概念);范式的概念;从1NF到4NF的定义;规范化的含义和作用。

需要举一反三的:四个范式的理解与应用,各个级别范式存在的问题(插入异常、删除异常、数据冗余)和解决方法;能够根据应用语义,完整地写出关系模式的数据依赖集合,并能根据数据依赖分析某一个关系模式属于第几范式。

难点:各个级别范式的关系及其证明。

二、习题解答和解析

1. 理解并给出下列术语的定义:

函数依赖、部分函数依赖、完全函数依赖、传递依赖、候选码、主码、外码、全码(All-key)、1NF、2NF、3NF、BCNF、多值依赖、4NF。

解析

解答本题不能仅仅把《概论》上的定义写下来。关键是真正理解和运用这些概念。

答

函数依赖: 设 $R(U)$ 是一个关系模式, U 是 R 的属性集合, X 和 Y 是 U 的子集。对于 $R(U)$ 的任意一个可能的关系 r , 如果 r 中不存在两个元组, 它们在 X 上的属性值相同, 而在 Y 上的属性值不同, 则称“ X 函数确定 Y ”或“ Y 函数依赖于 X ”, 记作 $X \rightarrow Y$ 。

解析

(1) 函数依赖是最基本的一种数据依赖, 也是最重要的一种数据依赖。

(2) 函数依赖是属性之间的一种联系, 体现在属性值是否相等。由上面的定义可以知道, 如果 $X \rightarrow Y$, 则 r 中任意两个元组, 若它们在 X 上的属性值相同, 那么在 Y 上的属性值一定也相同。

(3) 要从属性间实际存在的语义来确定他们之间的函数依赖, 即函数依赖反映了(描述了)现实世界的一种语义。

(4) 函数依赖不是指关系模式 R 在某个时刻的关系(值)满足的约束条件, 而是指 R 任何时刻的一切关系均要满足的约束条件。

答

完全函数依赖、部分函数依赖: 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 $X' \not\rightarrow Y$, 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:

$$X \xrightarrow{P} Y$$

传递依赖: 在 $R(U)$ 中, 如果 $X \rightarrow Y$, $(Y \setminus X) \rightarrow Z$, 则称 Z 对 X 传递函数依赖。

候选码、主码: 设 K 为 $R(U, F)$ 中的属性或属性组合, 若 $K \xrightarrow{F} U$ 则 K 为 R 的候选码 (Candidate key)。若候选码多于一个, 则选定其中的一个为主码 (Primary key)。

解析

1) 这里我们用函数依赖来严格定义码的概念。在第二章中我们只是描述性地定义码(可以复习 2.2.1): 若关系中的某一属性组的值能惟一地标识一个元组, 则称该属性组为候选码 (Candidate key)。

2) 因为码有了严格定义, 在学习了《概论》5.3 数据依赖的公理系统后就可以从 $R(U, F)$ 的函数依赖集 F 出发, 用算法来求候选码。

答

外码: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 (Foreign key),也称外码。

全码:整个属性组是码,称为全码 (All - key)。

答

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项,则 R 1NF。

解析

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。

答

2NF:若关系模式 R 1NF,并且每一个非主属性都完全函数依赖于 R 的码,则 R 2NF。

3NF:关系模式 $R < U, F >$ 中若不存在这样的码 X ,属性组 Y 及非主属性 Z ($Z \setminus Y$)使得 $X \twoheadrightarrow Y, (Y \setminus X) \twoheadrightarrow Z$,成立,则称 $R < U, F >$ 3NF。

BCNF:关系模式 $R < U, F >$ 1NF。若 $X \twoheadrightarrow Y$ 且 $Y \setminus X$ 时 X 必含有码,则 $R < U, F >$ BCNF。

解析

读者要真正理解这些范式的内涵。各种范式之间的联系:5NF \supset 4NF \supset BCNF \supset 3NF \supset 2NF \supset 1NF(《概论》上图 5.2)。能够理解为什么有这种包含关系。

答

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。

4NF:关系模式 $R < U, F >$ 1NF,如果对于 R 的每个非平凡多值依赖 $X \twoheadrightarrow Y (Y \setminus X)$, X 都含有码,则称 $R < U, F >$ 4NF。

解析

对于多值依赖的定义有多种。《概论》上定义 5.9 后面又给出了一种等价的定义。习题中的第 4 题是另一种等价的定义。可以对比不同的定义来理解多值依赖,选择自己容易理解的一种定义来掌握多值依赖概念。

2 建立一个关于系、学生、班级、学会等诸信息的关系数据库。

描述学生的属性有:学号、姓名、出生年月、系名、班号、宿舍区。

描述班级的属性有:班号、专业名、系名、人数、入校年份。

描述系的属性有:系名、系号、系办公室地点、人数。

描述学会的属性有:学会名、成立年份、地点、人数。

有关语义如下：一个系有若干专业，每个专业每年只招一个班，每个班有若干学生。一个系的学生住在同一宿舍区。每个学生可参加若干学会，每个学会若干学生。学生参加某学会有一入会年份。

请给出关系模式，写出每个关系模式的极小函数依赖集，指出是否存在传递函数依赖，对于函数依赖左部是多属性的情况讨论函数依赖是完全函数依赖，还是部分函数依赖。

指出各关系的候选码、外部码，有没有全码存在？

答

关系模式：学生 S(S # ,SN,SB,DN,C # ,SA)

班级 C(C # ,CS,DN,CNUM,CDATE)

系 D(D # ,DN,DA,DNUM)

学会 P(PN,DATE1,PA,PNUM)

学生 - 学会 SP(S # ,PN,DATE2)

其中，S # —学号，SN—姓名，SB—出生年月，SA—宿舍区

C # —班号，CS—专业名，CNUM—班级人数，CDATE—入校年份

D # —系号，DN—系名，DA—系办公室地点，DNUM—系人数

PN—学会名，DATE1—成立年月，PA—地点，PNUM—学会人数，

DATE2—入会年份

每个关系模式的极小函数依赖集：

S:S # SN,S # SB,S # C # ,C # DN,DN SA

C:C # CS,C # CNUM,C # CDATE,CS DN,(CS,CDATE) C #

/ * 因为每个专业每年只招一个班 */

D:D # DN,DN D # ,D # DA,D # DNUM

/ * 按照实际情况，系名和系号是一一对应的 */

P:PN DATE1,PN PA,PN PNUM

SP:(S # ,PN) DATE2

S 中存在传递函数依赖：S # DN,S # SA,C # SA

/ * 因为 S # C # ,C # DN,DN SA */

C 中存在传递函数依赖：C # DN

/ * 因为 C # CS,CS DN */

(S # ,PN) DATE2 和 (CS,CDATE) C # 均为 SP 中的函数依赖，是完全函数依赖。

数依赖。

关系	候选码	外部码	全码
S	S #	C # ,DN	无
C	C # ,(CS,CDATE)	DN	无

D	D # 和 DN	无	无
P	PN	无	无
SP	(S # , PN)	S # , PN	无

解析

读者应该根据题目中给出的有关语义写出关系模式中的数据依赖,有些依赖可以按照实际情况写出,也许题目中并没有明显指出。例如,按照实际情况,系名和系号是一一对应的,因此有 $D \# \rightarrow DN, DN \rightarrow D \#$ 。

3 试由 Armstrong 公理系统推导出下面三条推理规则:

- (1) 合并规则:若 $X \rightarrow Z, X \rightarrow Y$,则有 $X \rightarrow YZ$
- (2) 伪传递规则:由 $X \rightarrow Y, WY \rightarrow Z$ 有 $XW \rightarrow Z$
- (3) 分解规则: $X \rightarrow Y, Z \rightarrow Y$,有 $X \rightarrow Z$

证明

- (1) 已知 $X \rightarrow Z$,由增广律知 $XY \rightarrow YZ$,又因为 $X \rightarrow Y$,可得 $XX \rightarrow XY \rightarrow YZ$,最后根据传递律得 $X \rightarrow YZ$ 。
- (2) 已知 $X \rightarrow Y$,据增广律得 $XW \rightarrow WY$,因为 $WY \rightarrow Z$,所以 $XW \rightarrow WY \rightarrow Z$,通过传递律可知 $XW \rightarrow Z$ 。
- (3) 已知 $Z \rightarrow Y$,根据自反律知 $Y \rightarrow Z$,又因为 $X \rightarrow Y$,所以由传递律可得 $X \rightarrow Z$ 。

4 关于多值依赖的另一种定义是:

给定一个关系模式 $R(X, Y, Z)$,其中 X, Y, Z 可以是属性或属性组合。

设 $x \in X, y \in Y, z \in Z$, xz 在 R 中的像集为:

$$Y_{xz} = \{r.Y \mid r.X = x \wedge r.Z = z \wedge r \in R\}$$

定义 $R(X, Y, Z)$ 当且仅当 $Y_{xz} = Y_{x'z}$ 对于每一组 (x, z, z) 都成立,则 Y 对 X 多值依赖,记作 $X \twoheadrightarrow Y$ 。这里,允许 Z 为空集,在 Z 为空集时,称为平凡的多值依赖。

请证明这里的定义和《概论》5.2.7 节中定义 5.9 是等价的。

证明

设 $Y_{xz} = Y_{x'z}$ 对于每一组 (x, z, z) 都成立,现证其能推出定义 5.9 的条件:

设 s, t 是关系 r 中的两个元组, $s[X] = t[X]$,由新定义的条件可知对于每一个 z 值,都对应相同的一组 y 值。这样一来,对相同的 x 值,交换 y 值后所得的元组仍然属于关系 r ,即定义 5.9 的条件成立;

如果定义 5.9 的条件成立,则对相同的 x 值,交换 y 值后所得的元组仍然属于关系 r ,由于任意性及其对称性,可知每个 z 值对应相同的一组 y 值,所以 $Y_{xz} = Y_{x'z}$ 对于每一组 (x, z, z) 都成立。

综上可知,新定义和定义 5.9 的条件是等价的,所以新定义和定义 5.9 是等

价的。

5 试举出 3 个多值依赖的实例。

答

(1) 关系模式 $MSC(M, S, C)$ 中, M 表示专业, S 表示学生, C 表示该专业的必修课。假设每个专业有多个学生, 有一组必修课。设同专业内所有学生选修的必修课相同, 实例关系如下。按照语义对于 M 的每一个值 M_i , S 有一个完整的集合与之对应而不问 C 取何值, 所以 $M \twoheadrightarrow S$ 。由于 C 与 S 的完全对称性, 必然有 $M \twoheadrightarrow C$ 成立。

M	S	C
M1	S1	C1
M1	S1	C2
M1	S2	C1
M1	S2	C2
.....

(2) 关系模式 $ISA(I, S, A)$ 中, I 表示学生兴趣小组, S 表示学生, A 表示某兴趣小组的活动项目。假设每个兴趣小组有多个学生, 有若干活动项目。每个学生必须参加所在兴趣小组的所有活动项目, 每个活动项目要求该兴趣小组的所有学生参加。

按照语义有 $I \twoheadrightarrow S, I \twoheadrightarrow A$ 成立。

(3) 关系模式 $RDP(R, D, P)$ 中, R 表示医院的病房, D 表示责任医务人员, P 表示病人。假设每个病房住有多个病人, 有多个责任医务人员负责医治和护理该病房的所有病人。按照语义有 $R \twoheadrightarrow D, R \twoheadrightarrow P$ 成立。

* 6 试证明《概论》上给出的关于 FD 和 MVD 公理系统的 A_4, A_6 和 A_8 。

证明

A_4 : 若 $X \twoheadrightarrow Y, V \twoheadrightarrow W \mid U$, 则 $XW \twoheadrightarrow YV$

设 $Z = U - X - Y$

已知 $X \twoheadrightarrow Y$, 设 r 是 R 上的任一关系, $s, t \subseteq r$, 且 $t[X] = s[X]$, 则存在元组 $p, q \in r$, 使 $p[X] = q[X] = t[X]$, 而 $p[Y] = t[Y], p[Z] = s[Z], q[Y] = s[Y], q[Z] = t[Z]$ 。

设 $t[XW] = s[XW]$, 我们以上构造的元组 p 和 q , 是某部分属性在 s 和 t 上翻转而成, 所以 $p[W] = q[W]$, 可知 $p[XW] = q[XW]$, 同理 $p[YV] = t[YV]$ (由 $V \twoheadrightarrow W$ 知 $t[V] = s[V]$), $q[YV] = s[YV], p[U - YV - XW] = s[U - YV - XW]$ (因为 $U - YV - XW \twoheadrightarrow Z$), $q[U - YV - XW] = t[U - YV - XW]$ 。所以 $XW \twoheadrightarrow YV$ 。

A_6 : 若 $X \twoheadrightarrow Y, Y \twoheadrightarrow Z$ 则 $X \twoheadrightarrow Z - Y$

由 $Y \twoheadrightarrow Z$ 容易证得 $Y \twoheadrightarrow Z - Y$ 。

设 $R_1 = U - X - Y$, $R_2 = U - Y - Z$, $R_3 = U - X - Z + Y$ 。

已知 $X \twoheadrightarrow Y$, 设 r 是 R 上的任一关系, $s, t \subseteq r$, 且 $t[X] = s[X]$, 则存在元组 $p, q \in r$, 使 $p[X] = q[X] = t[X]$, 而 $p[Y] = t[Y]$, $p[R_1] = s[R_1]$, $q[Y] = s[Y]$, $q[R_1] = t[R_1]$ 。

对元组 t, p , 已知 $t[Y] = p[Y]$, $t[X] = p[X]$, 由 $Y \twoheadrightarrow Z - Y$ 知: 存在元组 $m \in r$, 使 $m[Z - Y] = p[Z - Y]$, $m[R_2] = t[R_2]$ 。因为 $(Z - Y) \twoheadrightarrow (R_1)$, 又 $p[R_1] = s[R_1]$, 所以 $m[Z - Y] = s[Z - Y]$ 。因为元组 p 和 s 在除属性 Y 之外的属性上值相等, 所以 $m[R_2] = t[R_2]$, 另外元组 m 是由元组 t 和 p 交换某些属性上的值而产生的, 而 t 和 p 在属性 X 上值相等, 显然 $m[X] = t[X]$, 所以 $m[U - (Z - Y)] = t[U - (Z - Y)]$, 即 $m[R_3] = t[R_3]$ 。

对元组 s, q , 同理可知 $s[Y] = q[Y]$, 存在元组 n , 使 $n[Z - Y] = t[Z - Y]$, 即 $n[R_3] = s[R_3]$ 。

综上所述, 对 $t, s \subseteq r$, $t[X] = s[X]$, 存在元组 $m, n \in r$, 使 $m[X] = n[X] = t[X]$, 而 $m[Z - Y] = s[Z - Y]$, $m[R_3] = t[R_3]$, $n[Z - Y] = t[Z - Y]$, $n[R_3] = s[R_3]$ 。

A8: 若 $X \twoheadrightarrow Y$, $W \twoheadrightarrow Z$, $W \twoheadrightarrow Y = \emptyset$, $Z \twoheadrightarrow Y$, 则 $X \twoheadrightarrow Z$ 。

设 r 是 R 上的任一关系, 对任意 $s, t \subseteq r$, 若 $t[X] = s[X]$, 设 $R_1 = U - X - Y$, 则根据 $X \twoheadrightarrow Y$ 知: 存在元组 $p, q \in r$, 使 $p[X] = q[X] = t[X]$, 而 $p[Y] = t[Y]$, $p[R_1] = s[R_1]$, $q[Y] = s[Y]$, $q[R_1] = t[R_1]$ 。因为 $W \twoheadrightarrow Y = \emptyset$, 所以 $s[W] = p[W]$, 又 $W \twoheadrightarrow Z$, 所以 $s[Z] = p[Z]$; 因为 $Z \twoheadrightarrow Y$, 且 $p[Y] = t[Y]$, 所以 $p[Z] = t[Z]$; 所以可得 $t[Z] = s[Z]$, 即 $X \twoheadrightarrow Z$ 。

* 7. 设关系模式为 $R(U, F)$, X, Y 为属性集, $X, Y \subseteq U$ 。

证明: (1) $X \subseteq X_F^+$

$$(2) (X_F^+)_F^+ = X_F^+$$

$$(3) \text{若 } X \twoheadrightarrow Y \text{ 则 } X_F^+ \twoheadrightarrow Y_F^+$$

$$(4) U_F^+ = U$$

解析

1. 要证明 $(X_F^+)_F^+ = X_F^+$, 只要证明 $X_F^+ \subseteq (X_F^+)_F^+$, 并且 $(X_F^+)_F^+ \subseteq X_F^+$ 。

而 $X_F^+ \subseteq (X_F^+)_F^+$ 是显然的, 因此只要证明 $(X_F^+)_F^+ \subseteq X_F^+$ 。

2. 这里的证明要用集合论的基本知识, 读者应该复习一下有关集合论中的有关概念和证明方法。

证明

(1) 因为 $X \subseteq X$, 所以 $X \subseteq X_F^+$ (根据 X_F^+ 的定义)。

(2) 下面求证 $(X_F^+)_F^+ = X_F^+$ 。

任意 $A \in (X_F^+)_F^+$, (由题意知) 存在 $B \in X_F^+$, 使 $B \rightarrow A$ 能由 F 根据 Armstrong 公理导出, 而从 $B \in X_F^+$ 可知 $X \rightarrow B$ 能由 F 根据 Armstrong 公理导出, 根据公理中的传递律可知 $X \rightarrow A$ 能由 F 根据 Armstrong 公理导出, 所以 $A \in X_F^+$, 因此 $(X_F^+)_F^+ \subseteq X_F^+$ 。

所以 $(X_F^+)_F^+ = X_F^+$ 。

(3) 对任意 $A \in X_F^+$, 可知 $X \rightarrow A$ 能由 F 根据 Armstrong 公理导出, 因为 $X \rightarrow Y$, 由自反律可以得 $Y \rightarrow X$, 由传递律得 $Y \rightarrow A$, 所以 $A \in Y_F^+$ 。 $X_F^+ \subseteq Y_F^+$ 得证。

(4) 下面证明 $U_F^+ = U$, 即证 U 由 F 据 Armstrong 公理推出的集合仍属于 U 。
解析

要证明 $U_F^+ = U$, 只要证明 $U \subseteq U_F^+$ 并且 $U_F^+ \subseteq U$ 。 $U \subseteq U_F^+$ 是显然的。

证明

自反律: $Y \rightarrow U$, $U \rightarrow Y$ 为 F 所蕴含, 显然 U 由 F 据 Armstrong 公理的自反律推出的 Y 仍属于 U ;

增广律: $U \rightarrow Y$ 为 F 所蕴含, 且 $Z \rightarrow U$, 则 $UZ \rightarrow YZ$ 为 F 所蕴含, $YZ \rightarrow U$;

传递律: $U \rightarrow Y$ 和 $Y \rightarrow Z$ 都为 F 所蕴含, 则 $U \rightarrow Z$ 为 F 所蕴含, $Z \rightarrow U$ 。

* 8 设关系模式为 $R(U, F)$, 若 $X_F^+ = X$, 则称 X 相对于 F 是饱和的。定义饱和集 $F = \{X \mid X = X_F^+\}$, 试证明 $F = \{X_F^+ \mid X \rightarrow U\}$ 。

解析

(1) 读者首先要理解饱和集定义的含义: 饱和集 F 是这样一个集合, 集合中的元素 X 都要满足条件 $X = X_F^+$

(2) 现在要证明 $F = \{X_F^+ \mid X \rightarrow U\}$, 即要证明 F 中的元素是由 U 中所有属性 X 的 X_F^+ 组成。弄清这两点, 下面的证明就容易了。

(3) 要证明 $F = \{X_F^+ \mid X \rightarrow U\}$, 根据集合论, 只要证明 $F \subseteq \{X_F^+ \mid X \rightarrow U\}$, 并且 $\{X_F^+ \mid X \rightarrow U\} \subseteq F$ 。

证明

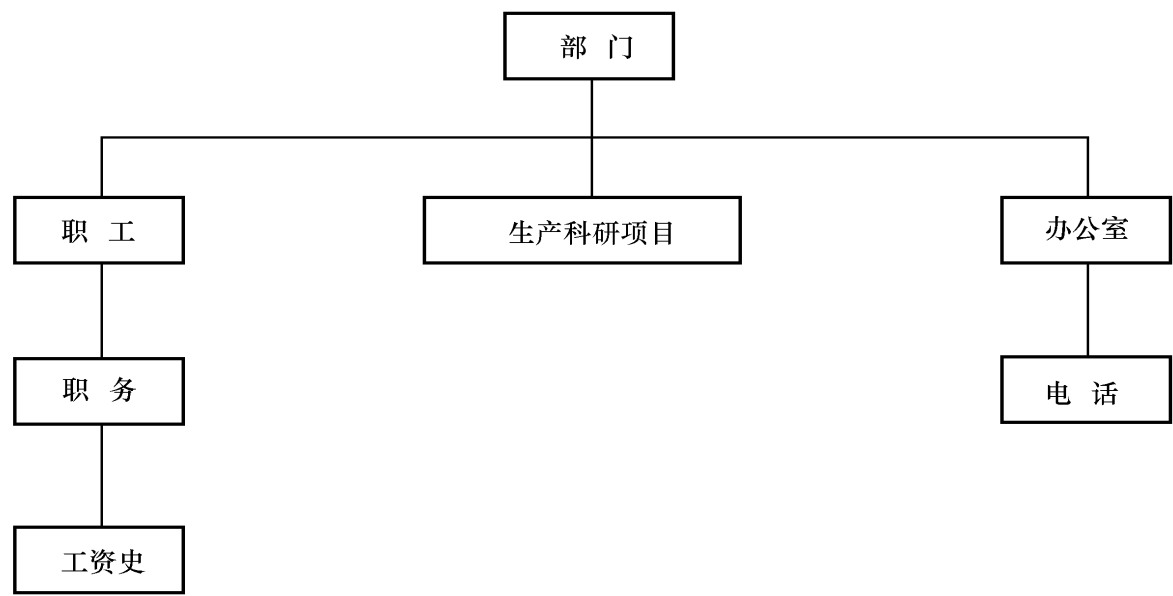
(1) 证 $F \subseteq \{X_F^+ \mid X \rightarrow U\}$ 。

对任意 $A \in F$, 由已知条件得 $A = A_F^+$, 因为 $A \rightarrow U$, $A = A_F^+$, 所以 $A \in \{X_F^+ \mid X \rightarrow U\}$ 。

(2) 证 $\{X_F^+ \mid X \rightarrow U\} \subseteq F$ 。

对任意 $A \in \{X_F^+ \mid A \rightarrow U\}$, 因为 $(A_F^+)_F^+ = A_F^+$ (见习题 7), 令 $B = A_F^+$, 有 $B_F^+ = B$ 所以 $B \in F$ 即 $A_F^+ \in F$, $A \in F$ 得证。

9 《概论》 上图 5 12 表示一个公司各部门的层次结构,重画如下。



对每个部门,数据库中包含部门号(惟一的)D #、预算费(BUGET)以及此部门领导人员的职工号E # (惟一的)信息。

对每一个部门,还存有关于此部门的全部职工、生产与科研项目以及办公室的信息。

职工信息包括:职工号、他所参加的生产与科研项目号(J #)、他所在办公室的电话号码(PHONE #)。

生产科研项目包含:项目号(惟一的)、预算费。

办公室信息包含办公室房间号(惟一的)、面积。

对每个职工,数据库中有他曾担任过的职务以及担任某一职务时的工资历史。

对每个办公室包含此办公室中全部电话号码的信息。

请给出你认为合理的数据依赖,把这个层次结构转换成一组规范化的关系。

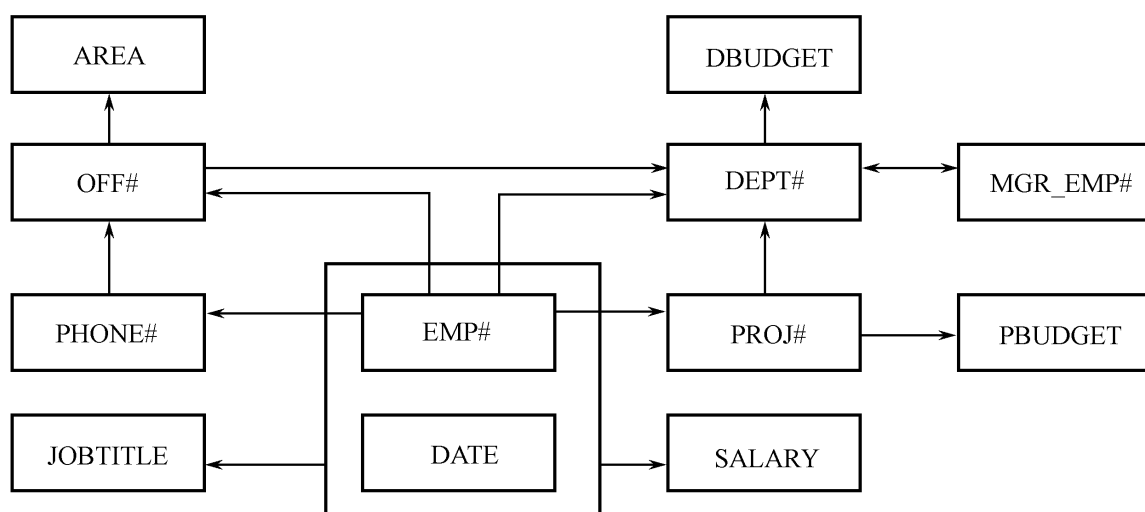
提示:此题可分步完成,第一步先转换成一组 1NF 的关系,然后逐步转换为 2NF,3NF,BCNF。

答

(1) 首先画出一些重要的函数依赖,所有这些函数依赖都是根据习题的文字说明和语义假设导出。

语义假设如下:

- 1) 一个职工不能同时成为多个部门的领导人;
- 2) 一个职工不能同时在多个部门就职;
- 3) 一个职工不能同时参加多个生产项目;
- 4) 一个职工不能同时在两个不同的办公室办公;
- 5) 一个职工不能同时拥有两部或两部以上的电话;



- 6) 一个生产项目不能同时分配给多个部门;
- 7) 一个办公室不能同时分配给多个部门;
- 8) 部门号、职工号、项目号、办公室号码及电话号码是全局惟一的。

(2) 先按照图 5 12 设计一组关系模式,它们都是属于 1NF 的。

DEPT (DEPT # , DBUDGET, MGR _ EMP #)

PRIMARY KEY(DEPT #)

DEPT # 和 MGR _ EMP # 都是侯选码,把 DEPT # 作为主码。

F = { DEPT # DBUDGET, DEPT # MGR _ EMP # , MGR _ EMP #
DEPT # }

EMP1 (EMP # , DEPT # , PROJ # , OFF # , PHONE #)

PRIMARY KEY (EMP #)

F = { EMP # DEPT # , EMP # PROJ # , EMP # OFF # , EMP #
PHONE # , PHONE # OFF # , OFF # DEPT # , PROJ #
DEPT # }

JOB(EMP # , JOBTITLE)

PRIMARY KEY (EMP # , JOBTITLE)

F = { EMP # , JOBTITLE EMP # , EMP # , JOBTITLE JOBTITLE }

SALHIST (EMP # , JOBTITLE, DATE, SALARY)

PRIMARY KEY (EMP # , DATE)

F = { EMP # , DATE JOBTITLE, EMP # , DATE SALARY }

PROJ(PROJ # , DEPT # , PBUDGET)

PRIMARY KEY (PROJ #)

$F = \{ PROJ\# \quad DEPT\#, PROJ\# \quad PBUDGET \}$

OFFICE (OFF # , DEPT # , AREA)

PRIMARY KEY (OFF #)

$F = \{ OFF\# \quad DEPT\#, OFF\# \quad AREA \}$

PHONE (PHONE # , OFF #)

PRIMARY KEY (PHONE #)

$F = \{ PHONE\# \quad OFF\# \}$

(3) 现在来分析一下这 7 个关系模式, 发现: SALHIST (EMP # , DATE, JOBTITLE, SALARY) 的属性包含了 JOB (EMP # , JOBTITLE) 的属性, 所以 JOB (EMP # , JOBTITLE) 可以消去。

EMP1 中 OFF # 和 DEPT # 都传递函数依赖于主码 (EMP #)。OFF # 通过 PHONE # , DEPT # 通过 PROJ # 或 OFF # (然后通过 PHONE #) 传递依赖于 { EMP # }, 所以可以把 EMP1 (EMP # , DEPT # , PROJ # , OFF # , PHONE #) 分解成下面 4 个 3NF 的关系模式:

EMP (EMP # , PROJ # , PHONE #)

PRIMARY KEY (EMP #)

X (PHONE # , OFF #)

PRIMARY KEY (PHONE #)

Y (PROJ # , DEPT #)

PRIMARY KEY (PROJ #)

Z (OFF # , DEPT #)

PRIMARY KEY (OFF #)

然而, X 就是 PHONE, Y 是 PROJ 的投影, Z 是 OFFICE 的投影, 所以 X、Y、Z 都可以消去。

最后可以得到下面 6 个关系模式, 所有这些关系模式都是属于 3NF 的, 进一步发现他们也是 BCNF 的。

DEPT (DEPT # , DBUDGET, MGR _ EMP #)

PRIMARY KEY (MGR _ EMP #)

EMP (EMP # , PROJ # , PHONE #)

PRIMARY KEY (EMP #)

SALHIST (EMP # , DATE, JOBTITLE, SALARY)

PRIMARY KEY (EMP, DATE)

PROJ (PROJ # , DEPT # , PBUDGET)

PRIMARY KEY (PROJ#)
OFFICE (OFF# , DEPT# , AREA)
PRIMARY KEY (OFF#)
PHONE (PHONE# , OFF#)
PRIMARY KEY (PHONE#)

10 在一个订货系统的数据库中,存有顾客、货物和订货单的信息。

每个顾客包含顾客号 CUST# (惟一的)、收货地址 ADDRESS(一个顾客可有几个地址)、赊购限额 CREDLIM、余额 BAL 以及折扣 DISCOUNT。

每个订货单 ORD# 包含顾客号 CUST#、收货地址 ADDRESS、订货日期 DATE、订货细则 LINE# (每个订货单有若干条), 每条订货细则内容为货物号 ITEM# 以及订货数量 QTYORD。

每种货物包含货物号 ITEM# (惟一的)、制造厂商 PLANT#、每个厂商的实际存货量 QTYOH、规定的最低存货量 DANGER 和货物描述 DESCN。

由于处理上的要求,每个订货单 ORD# 的每一订货细则 LINE# 中还应有一个未发货量 QTYOUT(此值初始时为订货数量,随着发货将减为零)。

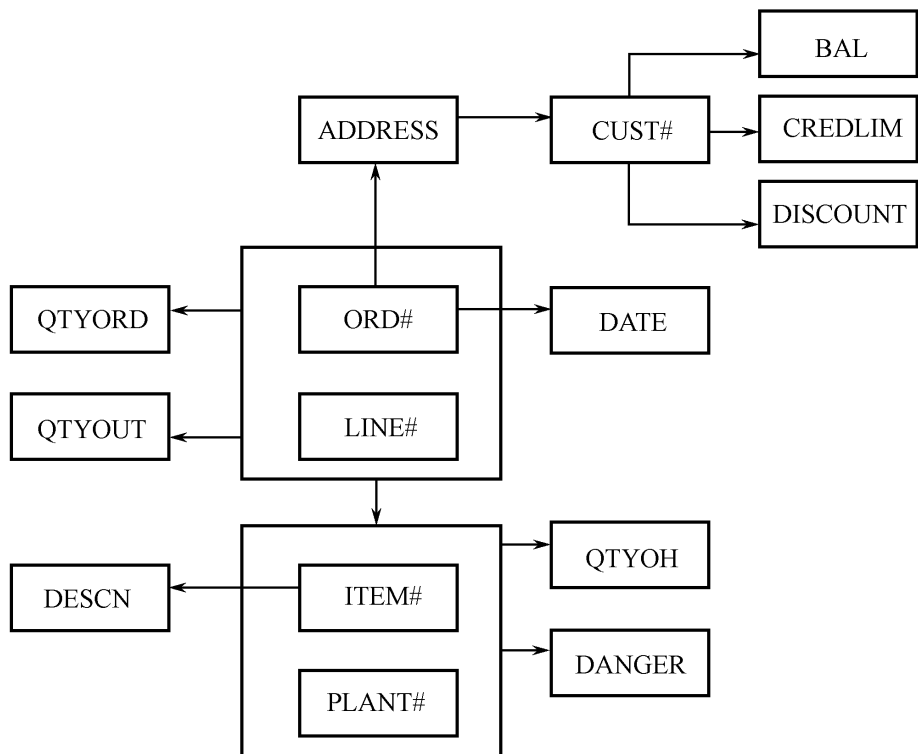
为这些数据设计一个数据库,如第 9 题那样,首先给出合理的数据依赖。

答

其语义假设如下:

- (1) 任何两个顾客的收货地址都不相同;
- (2) 每一个订单都有一个惟一的订单号码。
- (3) 每个订单的订单细则在这个订单里有一个惟一的编号。

函数依赖图如下:



相应的 BCNF 关系模式如下:

CUST (CUST # , BAL, CREDLIM, DISCOUNT)

PRIMARY KEY (CUST #)

SHIPTO (ADDRESS, CUST #)

PRIMARY KEY (ADDRESS)

ORDHEAD (ORD # , ADDRESS, DATE)

PRIMARY KEY (ORD #)

ORDLINE (ORD # , LINE # , ITEM # , QTYORD, QTYOUT)

PRIMARY KEY (ORD # , LINE #)

ITEM (ITEM # , DESCN)

PRIMARY KEY (ITEM #)

IP (ITEM # , PLANT # , QTYOH, DANGER)

PRIMARY KEY (ITEM # , PLANT #)

11. 设在第 10 题中实际上只有很少量的顾客(例如 1 %), 却有多多个发货地址, 由于这些少数的而又不能忽视的情形使得不能按一般的方式来处理问题。你能发现第 10 题答案中的问题吗? 能设法改进吗?

答

如果 99 % 的顾客只有一个收货地址, 则把地址放在与 CUST 不同的关系模式中, 在实际处理订货过程时的效率是很低的。因此我们可以对这个问题进行改进。对于每个顾客, 指定一个合法收货地址作为主地址, 则对于 99 % 的顾客, 该地址就是他的惟一地址。

关系模式 CUST 的定义修改如下:

CUST (CUST # , ADDRESS, BAL, CREDLIM, DISCOUNT)

PRIMARY KEY (CUST #)

99 % 的只有一个收货地址的顾客, 则 CUST # (顾客号) ADDRESS(地址)。

对于其他 1 % 的顾客, 建立关系模式 SECOND (代替原来的关系模式 SHIPTO):

SECOND(ADDRESS, CUST #)

PRIMARY KEY (ADDRESS)

这样, CUST 存放主地址, 而 SECOND 中存放所有的第二地址(和相应的顾客号), 这两个关系变量都是属于 BCNF 的。

该方法具有如下优点:

(1) 对于 99 % 的顾客的处理变得简单(当然更有效)了;

(2) 如果输入订单时把收货地址省略了, 则可以用主地址作为默认地址。

总的说来, 把特殊情况分离开来是个有效的方法, 它可以充分利用两者的优

点,既达到简化处理的目的,又使设计的关系模式达到 BCNF。

12 下面的结论哪些是正确的,哪些是错误的?对于错误的结论请给出理由或给出一个反例说明之。

答

(1) 任何一个二目关系都是属于 3NF 的。

(2) 任何一个二目关系都是属于 BCNF 的。

(3) 任何一个二目关系都是属于 4NF 的。

$R(X, Y)$ 如果 $X \twoheadrightarrow Y$ 即 X, Y 之间存在平凡的多值依赖, R 属于 4NF。

(4) 当且仅当函数依赖 $A \rightarrow B$ 在 R 上成立,关系 $R(A, B, C)$ 等于其投影 $R_1(A, B)$ 和 $R_2(A, C)$ 的连接。×

当 $A \rightarrow B$ 在 R 上成立,关系 $R(A, B, C)$ 等于其投影 $R_1(A, B)$ 和 $R_2(A, C)$ 的连接。反之则不然。

正确的应该是:

当且仅当多值依赖 $A \twoheadrightarrow B$ 在 R 上成立,关系 $R(A, B, C)$ 等于其投影 $R_1(A, B)$ 和 $R_2(A, C)$ 的连接。(参见《概论》上定理 5.6)

(5) 若 $R.A \twoheadrightarrow R.B, R.B \twoheadrightarrow R.C$, 则 $R.A \twoheadrightarrow R.C$

(6) 若 $R.A \twoheadrightarrow R.B, R.A \twoheadrightarrow R.C$, 则 $R.A \twoheadrightarrow R.(B, C)$

(7) 若 $R.B \twoheadrightarrow R.A, R.C \twoheadrightarrow R.A$, 则 $R.(B, C) \twoheadrightarrow R.A$

(8) 若 $R.(B, C) \twoheadrightarrow R.A$, 则 $R.B \twoheadrightarrow R.A, R.C \twoheadrightarrow R.A$ ×

反例:关系模式 $SC(S\#, C\#, G), (S\#, C\#) \twoheadrightarrow G$, 但是 $S\# \not\rightarrow G, C\# \not\rightarrow G$ 。