

第3章

关系数据库设计理论

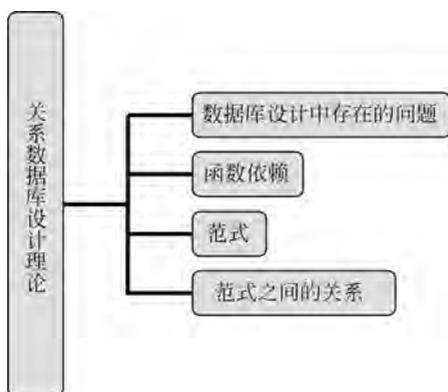


重点难点解析



典型例题

知识结构图



学习目标

- 了解数据库设计中存在的问题
- 了解函数依赖的概念
- 掌握范式的使用方法

导入案例

设计的数据库是否实用、高效？或者是否合理、正确？用户可以依据关系数据库设计理论进行核查。当然，在一般情况下，按照将 E-R 模型转化为关系模式的理论方法进行数据库模式设计是不会出现太大问题的，那么关系数据库设计理论也可以验证或者解释转化原理的必要性和有效性。错误的或者不合理的关系模式必然会在数据库进行增、删、改、查操作时发生种种异常。范式及范式之间的关系是关系模式进行规约和转化的理论基础，是重点学习内容。

3.1 数据库设计中存在的问题



视频讲解

数据库设计是否有据可依？为什么一定要遵循一定的原则？随意安排的一个关系模式到底容易出现哪些问题？例如设计学生关系模式 $S(SNO,$

SNAME,DEPT,HEAD,CNO,G),如图 3.1 所示,这个关系模式是否能满足基本的数据操作?

学号	姓名	系	系主任	课程号	成绩
<u>SNO</u>	<u>SNAME</u>	<u>DEPT</u>	<u>HEAD</u>	<u>CNO</u>	<u>G</u>
S01	杨明	D01	李一	C01	90
S02	李婉	D01	李一	C01	87
S01	杨明	D01	李一	C02	92
S03	刘海	D02	王二	C01	95
S04	安然	D02	王二	C02	78
S05	乐天	D03	赵三	C01	82

图 3.1 学生关系模式

数据操作包括增、删、改、查 4 种,但分析来看,上面的关系模式在进行具体操作时会发生插入异常、删除异常、数据冗余和更新异常等问题。

(1) 插入异常:如果一个系刚成立没有学生,或者有了学生但学生尚未选课,那么就无法将这个系及其负责人的信息插入数据库。

(2) 删除异常:如果某个系的学生全部都毕业了,则删除该系学生及其选修课程的同时把这个系及其负责人的信息也删掉了。

(3) 数据冗余:学生及其所选课程很多,而系主任只有一个,但其却要和学生及其所选课程出现的次数一样多。

(4) 更新异常:如果某个系要更换系主任,就必须修改这个系的学生所选课程的每个元组,修改其中的系主任信息。若有疏忽,就会造成数据的不一致。

发生这些操作异常的原因是把多个实体型用一个关系模式表示,解决办法是将现有关系模式进行分解,如图 3.2 所示。

<u>SNO</u>	<u>CNO</u>	<u>G</u>	<u>SNO</u>	<u>SNAME</u>	<u>DEPT</u>	<u>DEPT</u>	<u>HEAD</u>
S01	C01	90	S01	杨明	D01	D01	李一
S02	C01	87	S02	李婉	D01	D02	王二
S01	C02	92	S03	刘海	D02	D03	赵三
S03	C01	95	S04	安然	D02		
S04	C02	78	S05	乐天	D03		
S05	C01	82					

图 3.2 分解后的关系模式

那么如何设计、验证和修改关系模式?首要的问题是将关系模式中各属性之间的关系分析清楚。一个关系模式中各属性之间的关系可以被分成存在函数依赖和不存在函数依赖两种。用户可以依据不同的依赖关系将关系模式修正,将具体操作中的诸多问题消除。

3.2 函数依赖



视频讲解

一个实体型的诸属性之间具有内在的联系。通过对这些联系进行分析,可以做到一个关系模式只表示一个实体型的信息,从而消除上述问题。

在关系模型中,实体类型属性间这种既相互依赖又相互制约的关系称为数据依赖。

数据依赖是通过关系中属性值的相等与否体现出来的数据间的相互关系,它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。其中最重要的是函数依赖。

分析函数依赖关系可以改造性能较差的关系模式集合。

函数依赖极为普遍地存在于现实生活中。考查关系模式 $S(SNO, SNAME, DEPT, HEAD, CNO, G)$, 由于一个 SNO 只对应一个学生, 而一个学生只能在一个系中学习, 所以当 SNO 的值确定后, SNAME 和 DEPT 也被唯一确定了。就像自变量 x 确定后, 相应的 $f(x)$ 也被确定了一样。通常说 SNO 函数决定 (SNAME, DEPT), 而 (SNAME, DEPT) 函数依赖于 SNO。

1. 关于函数的定义

(1) 函数依赖: 设 $R(U)$ 是属性集 U 上的关系模式, $X, Y \subseteq U$, r 是 $R(U)$ 上的任意一个关系, 如果对任意两个元组 $t, s \in r$, 若 $t[X] = s[X]$, 则 $t[Y] = s[Y]$, 那么称“ X 函数决定 Y ”或“ Y 函数依赖于 X ”, 记作 $X \rightarrow Y$, 称 X 为决定因素或决定属性集。

例如 $SNO \rightarrow SNAME, (SNO, CNO) \rightarrow G$ 。

函数依赖是不随时间变化的。若关系 R 具有函数依赖 $X \rightarrow Y$, 那么虽然关系 R 的值随时间变化, 从而 $R[X, Y]$ 也会发生变化, 但 $R[X, Y]$ 在任一特定时刻仍保持为一个函数。

函数依赖与属性间的联系类型有关。当 X, Y 之间是“一对一”联系时, 则存在函数依赖 $X \rightarrow Y$ 和 $Y \rightarrow X$; 当 X, Y 之间是“多对一”联系时, 则存在函数依赖 $X \rightarrow Y$; 当 X, Y 之间是“多对多”联系时, 则不存在函数依赖。

函数依赖不是指关系模式 R 的某个或某些元组满足的约束条件, 而是指 R 的一切元组均要满足的约束条件。

函数依赖是现实世界中属性间关系的客观存在和数据库设计者的人为强制相结合的产物。

(2) 平凡函数依赖: 如果 $X \rightarrow Y$, 但 $Y \subseteq X$, 则称其为非平凡函数依赖, 否则称其为平凡函数依赖。

例如 $(SNO, SNAME) \rightarrow SNAME$ 是平凡函数依赖。

(3) 部分函数依赖: 在 $R(U)$ 中, 如果 $X \rightarrow Y$, 且对于任意 X 的真子集 X' 都有 $X' \not\rightarrow Y$, 则称 Y 对 X 完全函数依赖, 记作 $X \xrightarrow{f} Y$, 否则称 Y 对 X 部分函数依赖, 记作 $X \xrightarrow{p} Y$ 。

$$(SNO, CNO) \xrightarrow{f} G$$

$$(SNO, CNO) \xrightarrow{p} SNAME$$

(4) 传递函数依赖: 在 $R(U)$ 中, 如果 $X \rightarrow Y, Y \rightarrow Z$, 且 X 不包含 $Y, Y \not\rightarrow X$, 则称 Z 对 X 传递函数依赖。

【例 3.1】 $SNO \rightarrow DEPT, DEPT \rightarrow HEAD$, HEAD 对 SNO 传递函数依赖。

2. 关于键的定义

(1) 超键: 设 K 为 $R(U, F)$ 的属性或属性组, 若 $K \rightarrow U$, 则称 K 为 R 的超键。

【例 3.2】 $SNO \rightarrow U, (SNO, SNAME) \rightarrow U$ 。

(2) 候选键: 设 K 为 $R(U, F)$ 的属性或属性组, 若 K 满足以下条件, 则称 K 为 R 的一个候选键。

条件 1: $K \rightarrow U$ 。

条件 2: 不存在 K 的真子集 Z 使得 $Z \rightarrow U$ 成立。

或者: 设 K 为 $R(U, F)$ 的超键, 若 $K \xrightarrow{f} U$, 则称 K 为 R 的候选键。

例如 $SNO \rightarrow U$ 。

(3) 主键: 若 $R(U, F)$ 有多个候选键, 则可以从中选择一个作为 R 的主键。

主键是唯一确定一个实体的最少属性的集合。

例如, S 关系模式中的 SNO ; SC 关系模式中的 (SNO, CNO) 。

(4) 键属性: 包含在任何一个候选键中的属性, 称为键属性。

(5) 非键属性: 不包含在任何一个候选键中的属性, 称为非键属性。

(6) 全键: 关系模式的键由整个属性组构成。

例如关系模式 $S(\underline{SNO}, SNAME, DEPT, HEAD, \underline{CNO}, G)$, 主键为 (SNO, CNO) 。

【例 3.3】 指出关系模式 $S(\underline{SNO}, SNAME, DEPT, HEAD, \underline{CNO}, G)$ 中的函数依赖。

$(SNO, CNO) \xrightarrow{f} G$

$SNO \rightarrow SNAME, (SNO, CNO) \xrightarrow{p} SNAME$

$SNO \rightarrow DEPT, (SNO, CNO) \xrightarrow{p} DEPT$

$DEPT \rightarrow HEAD, (SNO, CNO) \xrightarrow{p} HEAD$

3.3 范 式



视频讲解

范式是对关系的不同数据依赖程度的要求。如果一个关系满足某个范式所指定的约束集, 则称它属于某个特定的范式。

范式(简称 NF)从低级到高级依次可分为 1NF、2NF、3NF、BCNF、4NF、5NF 乃至更高。范式之间的包含关系如图 3.3 所示。

下面将结合实例对范式及其相关概念进行解释。

1. 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级范式的关系模式的集合, 这一过程称为规范化。

2. 1NF

关系中的每一个分量必须是原子的, 不可再分。即不能以集合、序列等作为属性值。

【例 3.4】 图 3.4 所示的关系是否满足第一范式? 为什么? 如何解决?

图 3.4(a)所示的关系不满足 1NF, 因为不满足原子性, 分解成图 3.4(b)后满足 1NF。

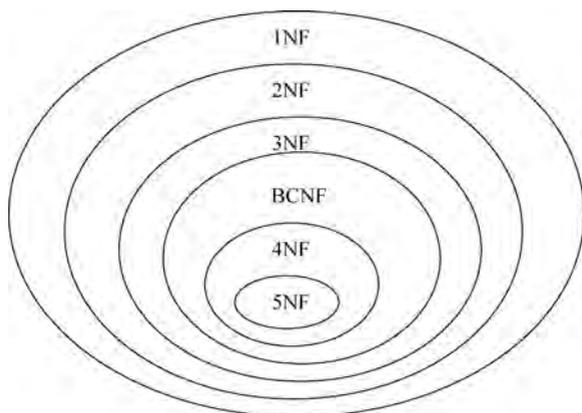


图 3.3 范式之间的包含关系

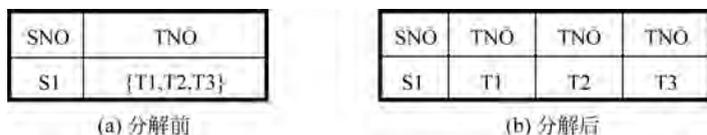


图 3.4 1NF 分解实例

3. 2NF

若 $R \in 1NF$ ，且每个非键属性完全依赖于键，则称 $R \in 2NF$ 。

将 1NF 的关系模式规范化为 2NF 的关系模式，其方法是消除 1NF 的关系模式中非键属性对键的部分依赖。

思考：如果关系 R 的所有属性都是 R 的键属性，或者 R 的所有候选键都只含一个属性，那么 R 是否属于第二范式？

答：属于 2NF。因为 R 的所有候选键都只含一个属性，所以满足 1NF；但不存在非键属性，所以满足 2NF 的定义。

规范化到 2NF 的必要性见例 3.5。

【例 3.5】 分析关系模式 $S(\underline{SNO}, SNAME, DEPT, HEAD, \underline{CNO}, G)$ 在实际应用中有什么问题？为什么？如何解决？

答：不良特性如下。

插入异常：如果学生没有选课，关于他的个人信息及所在系的信息就无法插入。

删除异常：如果删除学生的选课信息，则有关他的个人信息及所在系的信息也随之删除了。

更新异常：如果学生转系，若他选修了 k 门课，则需要修改 k 次。

数据冗余：如果一个学生选修了 k 门课，则有关他的所在系的信息重复。

原因： $S \notin 2NF$ ，因为 $(SNO, CNO) \xrightarrow{p} SNAME$ ， $(SNO, CNO) \xrightarrow{p} DEPT$ 。

解决办法：规范化。

步骤：

(1) 找出关系模式中所有的完全函数依赖。

$(SNO, CNO) \rightarrow G, SNO \rightarrow SNAME, SNO \rightarrow DEPT, DEPT \rightarrow HEAD$ 。

(2) 将满足完全函数依赖的属性关系划分到一个关系模式中。

$SC(\underline{SNO}, \underline{CNO}, G) \in 2NF$

$S_SD(\underline{SNO}, SNAME, DEPT, HEAD) \in 2NF$

4. 传递函数依赖

在关系模式 $R(U)$ 中, 如果 $X \rightarrow Y, Y \rightarrow Z$, 并且 X 不包含 $Y, Y \not\rightarrow X$, 则称 Z 对 X 传递函数依赖。

5. 3NF

在关系模式 $R(U, F)$ 中, 若不存在这样的键 X 、属性组 Y 及非键属性 $Z (Z \not\subseteq Y)$, 使得 $X \rightarrow Y, Y \rightarrow Z, Y \not\rightarrow X$ 成立, 则称 $R \in 3NF$ 。

将 2NF 的关系模式规范化为 3NF 的关系模式, 其方法是消除 2NF 的关系模式中非键属性对键的传递依赖。

2NF 规范化到 3NF 的必要性见例 3.6。

【例 3.6】 关系模式 $S_SD(SNO, SNAME, DEPT, HEAD)$ 在实际应用中有什么问题? 为什么? 如何解决?

答: 不良特性如下。

插入异常: 如果系中没有学生, 则有关系的信息就无法插入。

删除异常: 如果学生全部毕业了, 则在删除学生信息的同时有关系的信息也随之删除了。

更新异常: 如果学生转系, 不仅要修改 DEPT, 还要修改 HEAD, 如果换系主任, 则该系的每个学生元组都要做相应修改。

数据冗余: 每个学生都存储了所在系的系主任的信息。

原因: $S_SD \notin 3NF$, 因为 $SNO \rightarrow DEPT, DEPT \rightarrow HEAD$ 。

解决办法: 规范化。

步骤:

(1) 找出传递依赖关系。

(2) 在原关系模式中去除传递依赖主属性的属性, 并将起传递作用的属性和其决定的属性组成新的关系模式。

将 S_SD 分解为 $STUDENT(\underline{SNO}, SNAME, DEPT), DEPT(\underline{DEPT}, HEAD)$ 。

6. BCNF 范式

若关系模式 $R(U, F) \in 1NF$, 如果对于 R 的每个函数依赖 $X \rightarrow Y$, 且 $Y \not\subseteq X$ 时, X 必含有键, 则 $R(U, F) \in BCNF$ 。

由 BCNF 的定义可以看到, 每个 BCNF 的关系模式都具有以下 3 个性质:

- (1) 所有非键属性都完全函数依赖于每个候选键。
- (2) 所有键属性都完全函数依赖于每个不包含它的候选键。
- (3) 没有任何属性完全函数依赖于非键的任何一组属性。

也就是说, 如果关系模式 R 的每一个决定因素都包含键, 则 R 属于 BCNF 范式(不存在非键决定因素)。

【例 3.7】 有 $STJ(S, T, J)$, S 表示学生, T 表示教师, J 表示课程。每个教师只教一门

课,每门课由若干教师教,某一学生选定某门课就确定了一个固定的教师,因此具有函数依赖 $T \rightarrow J, (S, J) \rightarrow T, (S, T), (S, J)$ 为候选键。请问 STJ 在实际应用中有何问题? 为什么? 如何解决? 满足 BCNF 范式吗?

答: 不良特性如下。

插入异常: 如果没有学生选修某个教师的任课,则该教师担任课程的信息就无法插入。

删除异常: 删除学生选课信息,会删除掉教师的任课信息。

更新异常: 如果教师所教的课程有所改动,则所有选修该教师课程的学生元组都要做改动。

数据冗余: 每位学生都存储了有关教师所教的课程的信息。

原因: 键属性对键的不良依赖。如 $STJ \notin BCNF$, 因为 $T \rightarrow J$, 而 T 不含有键。

改造: 将 STJ 分解为 $(S, T), (T, J)$ 。

【例 3.8】 考虑两个关系,分析它们是否满足 3NF 和 BCNF 范式。

(1) 关系模式 $S(S\#, SNAME, SADD, SAGE)$, 限制 SNAME 唯一;

(2) 关系模式 $SS(S\#, SNAME, C\#, G)$, 限制 SNAME 唯一。

答: 在 $S(S\#, SNAME, SADD, SAGE)$ 中, 键为 $S\#$ 和 SNAME, 且除此以外无其他决定因素, 是 3NF 范式和 BCNF 范式。

在 $SS(S\#, SNAME, C\#, G)$ 中, 键为 $(S\#, C\#)$ 和 $(SNAME, C\#)$, 非主属性 G 不传递任何候选键, 所以 SS 是 3NF 范式, 但它不是 BCNF 范式。因为 $S\# \rightarrow SNAME$, $S\#$ 不是 SS 的候选键。

一个关系数据库模式中的关系都属于 BCNF, 则在函数依赖的范畴内已实现了彻底的分离, 消除了插入、删除和修改的异常。3NF 的“不彻底”性表现在当关系模式具有多个候选键, 且这些候选键具有公共属性时, 可能存在主属性对键的部分依赖和传递依赖。

关系模式的属性之间除了函数依赖以外, 还存在多值依赖关系。

【例 3.9】 有关系模式 $TEACH(C\#, P\#, B\#)$, 一门课程由多个教师担任, 一门课程使用相同的一套参考书, 如图 3.5 所示。它是否属于 BCNF 范式? 在实际应用中有何问题?

C#	P#	B#
物理	{张明, 张平}	{普通物理学, 光学原理}
化学	{张明, 王微}	{无机化学, 有机化学}

(a) 原始数据

C#	P#	B#
物理	张明	普通物理学
物理	张明	光学原理
物理	张平	普通物理学
物理	张平	光学原理
化学	张明	无机化学
化学	张明	有机化学
化学	王微	无机化学
化学	王微	有机化学

(b) 规范到 BCNF 的关系模式

图 3.5 TEACH 关系模式

答：它的键是(C#,P#,B#),是全键,没有函数依赖关系,所以属于 BCNF 范式。其不良特性如下。

插入异常：当某门课程增加一个教师时,该门课程有多少本参考书就必须插入多少个元组；同样,当某门课程需要增加一本参考书时,它有多少个教师就必须插入多少个元组。

删除异常：当删除一门课程的某个教师或者某本参考书时,需要删除多个元组。

更新异常：当一门课程的教师或参考书作出改变时,需要修改多个元组。

数据冗余：同一门课的教师与参考书的信息被反复存储多次。

7. 多值依赖

(1) 描述型定义：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, 当且仅当在 $R(U)$ 的任一关系 r 中, 对于给定的 X 属性值, 都有一组 Y 的值与之对应, 而与其他属性 Z 值无关。

例如在关系模式 TEACH 中, 对(物理, 普通物理学)有一组 $P\#$ 值(张明, 张平), 对(物理, 光学原理)也有一组 $P\#$ 值(张明, 张平), 这组值仅取决于 $C\#$ 的取值, 而与 $B\#$ 的取值无关。因此, $P\#$ 多值依赖于 $C\#$, 记作 $C\# \twoheadrightarrow P\#$, 同样有 $C\# \twoheadrightarrow B\#$ 。

(2) 形式化定义：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：

$$w[X] = s[X] = v[X] = t[X]$$

$$w[Y] = t[Y], v[Y] = s[Y]$$

$$w[Z] = s[Z], v[Z] = t[Z]$$

则称 Y 多值依赖于 X , 记作 $X \twoheadrightarrow Y$ 。

若 $(C\#, P\#, B\#)$ 满足 $C\# \twoheadrightarrow P\#$, 含有元组 $t=(\text{物理}, \text{张明}, \text{普通物理学}), s=(\text{物理}, \text{张平}, \text{光学原理})$, 则也一定含有元组 $w=(\text{物理}, \text{张明}, \text{光学原理}), v=(\text{物理}, \text{张平}, \text{普通物理学})$ 。

多值依赖有如下性质：

(1) 多值依赖具有对称性, 即若 $X \twoheadrightarrow Y$, 则 $X \twoheadrightarrow Z$, 其中 $Z=U-X-Y$ 。

(2) 函数依赖是多值依赖的特例, 即若 $X \rightarrow Y$, 则 $X \twoheadrightarrow Y$ 。

(3) 若 $X \twoheadrightarrow Y, U-X-Y=\varphi$, 则称 $X \twoheadrightarrow Y$ 为平凡的多值依赖。

函数依赖与多值依赖的区别如下：

$X \rightarrow Y$ 的有效性仅决定于 X, Y 属性集上的值, 它在任何属性集 $W(XY \subseteq W \subseteq U)$ 上都成立。

若 $X \rightarrow Y$ 在 $R(U)$ 上成立, 则对于任何 $Y' \subseteq Y$, 均有 $X \rightarrow Y'$ 成立。

$X \twoheadrightarrow Y$ 的有效性与属性集范围有关。

$X \twoheadrightarrow Y$ 在属性集 $W(XY \subseteq W \subseteq U)$ 上成立, 但在 U 上不一定成立。

$X \twoheadrightarrow Y$ 在 U 上成立 $\Rightarrow X \twoheadrightarrow Y$ 在属性集 $W(XY \subseteq W \subseteq U)$ 上成立。

若在 $R(U)$ 上, $X \twoheadrightarrow Y$ 在属性集 $W(XY \subseteq W \subseteq U)$ 上成立, 则称 $X \twoheadrightarrow Y$ 为 $R(U)$ 的嵌入式多值依赖。

若 $X \twoheadrightarrow Y$ 在 $R(U)$ 上成立, 则不能断言对于 $Y' \subseteq Y$ 是否有 $X \twoheadrightarrow Y'$ 成立。

8. 4NF

关系模式 $R(U, F) \in 1NF$, 如果对于 R 到每个非平凡的多值依赖 $X \twoheadrightarrow Y (Y \not\subseteq X)$, X 都含有键, 则称 $R \in 4NF$ 。

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。因为根据定义, 对于每一个非平凡的多值依赖 $X \twoheadrightarrow Y$, X 都含有候选键, 于是就有 $X \rightarrow Y$, 所以 4NF 允许的非平凡的多值依赖实际上是函数依赖。

例如关系模式 $CPB, C\# \twoheadrightarrow P\#, C\# \twoheadrightarrow B\#$, 键为 $(C\#, P\#, B\#)$, 所以 $CPB \notin 4NF$ 。如果一门课 C_i 有 m 个教师、 n 本参考书, 则关系中分量为 C_i 的元组共有 $m \times n$ 个, 数据冗余非常大。

改造: 将 CPB 分解为 $CP(C\#, P\#), CB(C\#, B\#)$, 在分解后的关系中分量为 C_i 的元组共有 $m+n$ 个, 如图 3.6 所示。

CP		CB	
C#	P#	C#	B#
物理	张明	物理	普通物理学
物理	张平	物理	光学原理
化学	张明	化学	无机化学

图 3.6 4NF 分解实例

3.4 范式之间的关系

在 3.3 节中, 对于范式的定义, 并未全部说明此范式满足上一级范式, 本节将证明范式之间的包含关系是对的, 即 $4NF \subset BCNF \subset 3NF \subset 2NF \subset 1NF$ 。

证明 3.1: $3NF \subset 2NF$

证明: 反证法。

若 $R \in 3NF$, 但 $R \notin 2NF$, 则按 2NF 的定义, 一定有非键属性部分依赖于键。

设 X 为 R 的键, 则存在 X 的真子集 X' , 以及非键属性 $Z (Z \not\subseteq X')$, 使得 $X' \rightarrow Z$ 。

于是在 R 中存在键 X 、属性组 X' 以及非键属性 $Z (Z \not\subseteq X')$, 使得 $X \rightarrow X', X' \twoheadrightarrow Z, (X' \rightarrow X), X \rightarrow Z$ 成立, 这与 $R \in 3NF$ 矛盾, 所以 $R \in 2NF$ 。

证明 3.2: $BCNF \subset 3NF$

证明: 反证法。

若 $R \in BCNF$, 但 $R \notin 3NF$, 则按 3NF 的定义, 一定有非键属性对键的传递依赖, 于是存在 R 的键 X 、属性组 Y 以及非主属性 $Z (Z \not\subseteq Y)$, 使得 $X \rightarrow Y, Y \rightarrow Z, Y \twoheadrightarrow X$ 成立。

由 $Y \twoheadrightarrow X$, 按 BCNF 的定义, Y 含有键, 于是 $Y \rightarrow X$ 成立, 这与 $Y \twoheadrightarrow X$ 矛盾, 所以 $R \in 3NF$ 。

小 结

平凡函数依赖: 如果 $X \rightarrow Y$, 但 $Y \subseteq X$, 则称其为非平凡函数依赖, 否则称为平凡函数依赖。例如 $(SNO, SNAME) \rightarrow SNAME$ 是平凡函数依赖。

部分函数依赖：在 $R(U)$ 中, 如果 $X \rightarrow Y$, 且对于任意 X 的真子集 X' 都有 $X' \not\rightarrow Y$, 则称 Y 对 X 完全函数依赖, 记作 $X \xrightarrow{f} Y$, 否则称 Y 对 X 部分函数依赖, 记作 $X \xrightarrow{p} Y$ 。例如 $(SNO, CNO) \xrightarrow{f} G, (SNO, CNO) \xrightarrow{p} SNAME$ 。

传递函数依赖：在 $R(U)$ 中, 如果 $X \rightarrow Y, Y \rightarrow Z$, 且 X 不包含 $Y, Y \not\rightarrow X$, 则称 Z 对 X 传递函数依赖。

1NF 的主要判断依据是每个属性是否满足原子性。

2NF 的主要判断依据是所有非键属性必须完全函数依赖于键。

3NF 的主要判断依据是属性之间不存在传递函数依赖。

BCNF 的主要判断依据是起决定作用的属性必须包含键。

BCNF 之前的所有范式适用于规范两个实体及其之间的关系。4NF 是用来规范 3 个实体之间的关系的, 做法一般是将模式分解为两两实体之间的关系。

范式之间的关系是 $4NF \subset BCNF \subset 3NF \subset 2NF \subset 1NF$ 。

范式是检查数据库设计是否合理的标准, 函数依赖是理解范式的重要概念。各级别的范式规范是解释数据库设计所遵循的理论的重要依据。

课 后 题

一、选择题

- 关系模式中数据依赖问题的存在可能会导致库中数据插入异常, 这是指()。
 - 插入了不该插入的数据
 - 数据插入后导致数据库处于不一致状态
 - 该插入的数据不能实现插入
 - 以上都不对
- 关系模式中的候选键()。
 - 有且仅有一个
 - 必然有多个
 - 可以有一个或多个
 - 以上都不对
- 在规范化的关系模式中, 所有属性都必须是()。
 - 相互关联的
 - 互不相关的
 - 不可分解的
 - 长度可变的
- 设关系模式 R 属于第一范式, 若在 R 中消除了部分函数依赖, 则 R 至少属于()。
 - 第一范式
 - 第二范式
 - 第三范式
 - 第四范式
- 若关系模式 R 中的属性都是主属性, 则 R 至少属于()。
 - 第三范式
 - BC 范式
 - 第四范式
 - 第五范式

二、填空题

- 一个不好的关系模式会存在_____异常、_____异常和_____等弊端。
- 设 $X \rightarrow Y$ 为 R 上的一个函数依赖, 若_____, 则称 Y 完全函数依赖于 X 。

3. 包含 R 中所有属性的候选键称_____。不在任何候选键中的属性称_____。
4. Armstrong 公理系统是_____的和_____的。
5. 第三范式是基于_____依赖的范式,第四范式是基于_____依赖的范式。

三、简答题

1. 解释下列术语的含义: 函数依赖、平凡函数依赖、非平凡函数依赖、部分函数依赖、完全函数依赖、传递函数依赖、范式、无损连接性、依赖保持性。

2. 给出 2NF、3NF、BCNF 的形式化定义,并说明它们之间的区别和联系。

3. 试证明全码的关系必是 3NF,也必是 BCNF。

4. 要建立关于系、学生、班级、研究会等信息的一个关系数据库,规定一个系有若干专业,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个宿舍区;每个学生可参加若干研究会,每个研究会有若干学生;学生参加某研究会,有一个入会年份。

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。

描述系的属性有系号、系名、系办公室地点、人数。

描述研究会的属性有研究会名、成立年份、地点、人数。

试给出上述数据库的关系模式;写出每个关系的最小依赖集(即基本的函数依赖集,不是导出的函数依赖);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,讨论其函数依赖是完全函数依赖还是部分函数依赖,指出各关系的候选键、外部键。