关系型数据库 逻辑设计 高阶范式 (9)

计算机科学 1994Vol. 21№. 3

40-41 TP311.13

关系数据库中保证高阶范式的简单条件

高 铃 编译

A

摘 要 本文对关系模式的高阶范式提出了新的简易实用的判别条件。文中的一些结果借助函数依赖为数据库设计者提供了实现高阶范式的简单、充分条件。

关键词 逻辑设计、范式，Boyce-Codd 范式、第四范式、第五范式或投影连接范式

一、引言

对于一般较简单的数据库逻辑设计，在函数依赖范畴内，只要关系模式达到 3NF 或 BCNF，即可避免数据冗余和插入、删除的异常。但在关系模式存在多值依赖的情况下，即使达到 BCNF，上述异常仍不可避免。Fagin 先后采用多值依赖定义了第四范式(4NF)，用连接依赖定义了投影连接范式(PJ/NF)，PJ/NF 又称为第五范式(5NF)。但在一般数据库教科书中，对高阶范式(4NF，5NF)的表述往往有些问题，或者过份地考虑形式体系，给出高阶范式的精确定义，但很少考虑实用的途径；或者完全相反，对高阶范式作了很不精确的描述，而很少有实用价值。这些问题在于：4NF 和 PJ/NF 是用多值依赖和连接依赖定义的，较之函数依赖更难于理解。

本文的目的是给出一个用函数依赖定义的易于理解的条件，该条件在很大一类情况下成立且足以保证实现高阶范式。文中已证明：如果关系模式是 3NF 且每个码都是简码，那么它是 PJ/NF。进而证明：若关系模式是 BCNF 且某个码是简码，则它是 4NF(但不一定是 PJ/NF)。这些结果便于数据库设计者理解，且为他们提供了实用的数据库设计指南。它对数据库教学也十分有益，即使学生没有多值依赖和连接依赖的知识，也能理解达到 4NF 和 5NF 的简单条件。

二、符号、定义和规则

为了说明方便，我们引用下列符号、定义、规则和算法等。

通常以 U 表示关系模式中全部属性的集合。若 X、Y 皆为属性集，则用 XY 表示 $X\cup Y$。

含单个属性的码称为简码。函数依赖简记为 FD，多值依赖记为 MVD。

定义 2.1 码依赖(KD) 设 K 是模式的码，FD $K\rightarrow U$ 称为模式的码依赖(KD)。

定义 2.2 连接依赖(JD) 连接依赖是形如$\bowtie\{X_1,\cdots,X_n\}$的语句，其中 X_1、X_2、…、X_n 是属性集。如果 $R=\bowtie\{R[X_1],\cdots,R[X_n]\}$，则说 JD$\bowtie\{X_1,X_2,\cdots,X_n\}$对关系 R 成立。

JD$\bowtie\{X_1,X_2\}$等价于 MVD $X_1\cap X_2\rightarrow\rightarrow X_2$。

定义 2.3 Zaniolo 的第三范式 一个关系模式 R(U，F)中，若对任一非平凡 FD $X\rightarrow A$，其中 A 是单个属性，或 X 是超码，或 A 是码属性，则 R(U，F)是 3NF。

定义 2.4 第四范式(4NF) 一个关系模式是 4NF，当且仅当模式的每个 MVD 是模式的码依赖(KD)集的逻辑导出。

定义 2.5 投影连接范式(PJ/NF)或 5NF 关系模式是 PJ/NF，如果模式中的每个 JD 是该模式的 KD 集的逻辑导出。

规则 2.6 FD 和 MVD 的推理规则 若 $X\rightarrow\rightarrow Z$ 且 $Y\rightarrow Z$，其中 Y 和 Z 不相交，则 $X\rightarrow Z$。

算法 2.7 确定 JD 是否是 KD 集的逻辑导出。

算法输入：JD$\bowtie\{X_1,\cdots,X_n\}$和 KD 集$\{K_1\rightarrow U,\cdots,K_s\rightarrow U\}$。对初始集 $\phi=\{X_1,\cdots,X_n\}$应用下列规则，直到不可再用为止：若$K_i\subseteq Y\cap Z$，对某个码$K_i(1\leqslant i\leqslant s)$和 ϕ 的成员 Y 和 Z 成立，则将 Y 和 Z 用它们的并取代(此时 ϕ 中的成员数减 1)。设$\{Y_1,\cdots,Y_r\}$是执行此算法的最后结果。如果某个 Y_i 等于属性集 U，则$\{X_1,\cdots,X_n\}$严格地是 KD 集$\{K_1\rightarrow U,\cdots,K_s\rightarrow U\}$的逻辑导出。

三、PJ/NF 的简单判别条件

下面证明保证达到 PJ/NF 的一个简单判别条

件。

定理 3.1 假设关系模式 R(U,F)∈3NF,且其中每个码是简码,则 R(U,F)∈PJ/NF。

首先证明下列引理:

引理 3.2 设 R(U,F)∈3NF,且每个码皆为简码,则 R(U,F)∈BCNF。

证明:设 X→A 是 R(U,F)的非平凡 FD,其中,A 是简单属性。依假设 R∈3NF,欲证 R∈BCNF,只须证 X 为超码即可。由 R∈3NF,知或 X 为超码或 A 为码属性。若 A 是码属性,由每个码皆为简码的假设知 A 本身是码。又因 X→A 是模式的 FD,所以 X 必为超码。由此,在任何情况下,X 都是超码。□

引理 3.3 设 J_1 是 JD,J_2 是由 J_1 通过将其中的两个分量用它们的并取代而得的连接依赖,则 J_1 逻辑蕴含 J_2。

引理 3.4 假设 R(U,F)∈BCNF,且每个码皆为简码,则 R(U,F)∈PJ/NF。

证明:(反证法)若 R(U,F)∉PJ/NF,则有 JD⋈$\{X_1,\cdots,X_m\}$不是 KD 集的逻辑导出。

设$\{Y_1,\cdots,Y_r\}$是应用算法 2.7 于 JD⋈$\{X_1,\cdots,X_m\}$的最后结果。按反证法假设⋈$\{X_1,\cdots,X_m\}$不是模式 KD 集的逻辑导出,由算法 2.7 知,无一 Y_i 等于 U。施用算法 2.7 的过程,亦满足引理 3.3 的前提,由此知 JD⋈$\{X_1,\cdots,X_m\}$逻辑蕴含 JD⋈$\{Y_1,\cdots,Y_r\}$。

由假设,模式中每个码皆为简码,每个关系模式至少有一个码。由此,模式有一个简码,设为 A。由于 Y_i 的并是全属性集,所以 A 恰在某个 Y_i 中。但 A 不能同时在 Y_i 和 Y_j 中,其中 $i\neq j$,因为 A 是码,它若同在 Y_i,Y_j 中,则由算法 2.7 知,Y_i 和 Y_j 要被 $Y_i\cup Y_j$ 取代。

不失一般性,设 Y_1 就是含 A 的 Y_i,定义 $W=Y_2\cup\cdots\cup Y_r$,即 W 是除 Y_1 外的所有 Y_i 的并。因为没有任何 Y_i 等于 U,所以 $Y_1\neq U$。重复使用引理 3.3,我们得到 JD⋈$\{Y_1,W\}$是 JD⋈$\{X_1,\cdots,X_m\}$的逻辑导出。如前所述,JD⋈$\{Y_1,W\}$等价于 MVD $Y_1\cap W\to\to W$,故 MVD $Y_1\cap W\to\to W$ 是此模式的 MVD。

因为 $Y_1\cap W\to\to W$ 是模式的 MVD,A→W 是模式的 FD,其中 A 是码,A 和 W 不相交(A 在 Y_1 中,$W=Y_2\cup\cdots\cup Y_r$,A∉W),所以,由推理规则 2.6 得知,$Y_1\cap W\to W$ 是此模式的 FD。我们证明:$Y_1\cap W\to W$ 是非平凡的。反证之,若不然,W 是 $Y_1\cap W$ 的子集,则 W 必是 Y_1 的子集,而实际上并非如此,故 FD $Y_1\cap W\to W$ 是非平凡的,且它是模式的 FD。该模式属 BCNF,所以 $Y_1\cap W$ 是超码。由假设,模式的每个码是简码,故 $Y_1\cap W$ 含某个简码 B。因为 $B\in Y_1\cap W\subseteq W$,$W=Y_2\cup\cdots\cup Y_r$,我们知道 $B\in Y_j$ 对某个 j 成立($2\leq j\leq r$),又知 $B\in Y_1\cap W\subseteq Y_1$。这样 B 既是 Y_1 的成员,又是 Y_j 的成员,但这是不可能的,由于 B 是码,按算法 2.6,Y_1 和 Y_j 必定被 $Y_1\cup Y_j$ 取代。引出矛盾,由此引理 3.4 得证。□

由引理 3.2 和引理 3.4 我们证得定理 3.1。

四、假设某个码是简码

以上关于每个码是简码的假设是相当强的。在这节我们证明:在每个属于 BCNF 的关系模式中某个码是简码,则它属于 4NF。我们还通过反例说明,满足上述条件的关系模式是 4NF,但不是 PJ/NF。

首先证明下述引理。

引理 4.1 假设关系模式是 BCNF,但不是 4NF,V→→W 是模式的 MVD,而 V 不是超码,设 W′是不在 V 或 W 中的属性的集合,则此模式的每个码包含 W 的成员和 W′的成员。

证明:不失一般性地假设 V 和 W 是不相交的,使每个属性恰在 V、W 或 W′三者之一中。若引理的假设成立而结论不成立,则设 K 是模式的码,而不含 W 和 W′的成员。由于 W 和 W′是对称的,故设 K 不含 W 的成员,即 K 和 W 不相交亦不失一般性。因 K 是码,所以 K→W 是此模式的 FD。现已知 V→→W 是此模式的 MVD,K→W 是模式的 FD 且 K 与 W 不相交,由规则 2.6 得 V→W 是模式的 FD。因为此模式是 BCNF,W 不是 V 的子集,所以 V 是超码,与题设矛盾,故引理为真。□

定理 4.2 假设关系模式是 BCNF,其中某个码是简码,则关系模式是 4NF。

证明:(反证法)设关系模式是 BCNF,但不是 4NF:证明该模式无简码。因为关系模式不是 4NF,它有一非平凡的 MVD V→→W,其中 V 不是超码。设 W′=U−V−W,由于 MVD V→→W 是非平凡的,故 W 和 W′非空。由引理 4.1,模式的每个码含 W 和 W′的成员,又 W 和 W′非空且不相交,由此推得该模式无简码。导出矛盾,结论得证。□

下列反例说明在定理 4.2 中,不能用 PJ/NF 取代 4NF。

设模式有属性集 U={A,B,C,D},这个模式的依赖集由 KD={A→U,BC−V}导出,它具有 JD⋈{ABC,BD,CD},该模式是 BCNF,且有一个简码 A,故它属 4NF。应用成员算法,我们发现,JD⋈{ABC,BD,CD}不是 KD{A→U,BC−U}的逻辑导出,故此关系模式不是 PJ/NF。

主要参考文献:ACM TODS,Vol. 17 No. 3,1992