

在关系数据库应用系统中数据的规范化探讨

郎 洪 陈勇孝

(郑州工业大学计算中心)

摘 要 本文探讨了在关系数据库应用系统中数据的规范化。

关键词 数据库 应用系统 数据规范化

中图分类号 TP311

随着计算机技术的飞速发展,数据库管理系统越来越广泛地应用于现代化管理中。在针对某一方面的应用开发研制一个关系数据库应用管理系统时,数据库的逻辑设计是一个关键的环节,数据库设计的好坏,直接影响到整个系统的性能。众所周知,在现实世界中,往往存在一些比较复杂的数据结构,若我们完全按照自然规律和人们的习惯来设计数据库时,必然存在大量的重复的字段和数据,这不仅产生了大量的数据冗余,还使得修改困难,容易造成数据不一致的情况。因此在关系数据库的设计中,数据规范化的目的就是要减少数据冗余和保持数据的一致性。其解决的方法就是将复杂的数据结构逐步分解为简单的数据结构。

通常认为,如果一个数据结构没有重复出现的数据项,就称该数据结构是规范化的。数据结构的规范化程度由“范式”(NF)度量。

在关系型数据库中,规范关系有几种范式,下边以一个学生上机登记表为例说明其规范关系:

姓 名	系 别	专 业	班 级	所用机时		日 期	机器号码
				上机时间	下机时间		
赵 祥	计算机	软件工程	94.1	10.21	12.00	95.10.03	31
				10.21	12.00	95.10.05	25
				10.21	12.00	95.10.06	3
				10.21	12.00	95.10.07	35
李晓明	计算机	软件工程	94.1	10.21	12.00	95.10.05	17
				10.21	12.00	95.10.06	22
...

根据第一范式(1NF)的定义知:一个关系模式R 的每个具体关系r 的每个属性值都是

收稿日期:1996-09-10

不可再分的数据单位。在上边的关系模式中,出现了层次结构和重复组的现象,我们知道,在关系模型中不能处理具有层次结构的属性记录和具有重复组的记录,这样的关系模式是非规范化的,按照 1NF 的定义可以把上表的非规范化关系转化为 1NF 关系:

姓名	系别	专业	班级	上机时间	下机时间	日期	机器号码
赵 祥	计算机	软件工程	94.1	10.21	12.00	95.10.03	31
赵 祥	计算机	软件工程	94.1	10.21	12.00	95.10.05	25
赵 祥	计算机	软件工程	94.1	10.21	12.00	95.10.06	3
赵 祥	计算机	软件工程	94.1	10.21	12.00	95.10.07	35
李晓明	计算机	软件工程	94.1	10.21	12.00	95.10.05	17
李晓明	计算机	软件工程	94.1	10.21	12.00	95.10.06	22
...

满足第一范式是设计关系数据库最基本的要求,但是从上表我们可以看出它明显地存在冗余和潜在的更新异常(即插入异常、删除异常和修改异常)问题。其主要原因是上述关系模式的关键字是由姓名、上机时间和日期联合组成的,其系别、专业、班级等只与姓名有关而与其它项无关。我们知道,在一个关系模式中,包含在任何关键字中的属性称为主属性,其它属性称为非主属性。由第二范式(2NF)的定义知:满足第一范式的关系模式 R 且它的所有非主属性都完全函数依赖于任一关键字。按照其定义可把上述的关系模式分解为:

姓名	系别	专业	班级
赵 祥	计算机	软件工程	94.1
李晓明	计算机	软件工程	94.1
...

姓名	上机时间	下机时间	日期	机器号码
赵 祥	10.21	12.00	95.10.03	31
赵 祥	10.21	12.00	95.10.05	25
赵 祥	10.21	12.00	95.10.06	3
赵 祥	10.21	12.00	95.10.07	35
李晓明	10.21	12.00	95.10.05	17
李晓明	10.21	12.00	95.10.06	22
...

在分解之后,其非主属性就变得完全函数依赖于关键字了。

其实,在某些情况下 2NF 也存在冗余高、更新异常等问题,比如我们分析一下第一个分解式,姓名为关键字,由姓名可以决定系别,而由系别决定专业。当然也可以由姓名决定专业,但这是通过系别才能决定,因此这是一种传递关系,即姓名→专业是从姓名→系别,系别→专业(由传递公理)推出的。一个系有多少个学生,则系名就会重复多少次。由第三范式(3NF)的定义知:满足第二范式的关系模式 R 且它的任何一个非主属性都不传递依赖于任一关键字。按照定义将上边的第一个分解式再度分解如下:

姓名	专业	班级
赵 祥	软件工程	94.1
李晓明	软件工程	94.1
...

专业	系别
软件工程	计算机
硬件工程	计算机
...	...

作了上述分解之后冗余和更新异常基本上得到解决。

3NF 的另一种形式称为 Boyce-Codd 范式(BCNF),其限制条件比 3NF 更强,3NF 消去非关键字依赖便成为 BCNF,BCNF 消除多值依赖就成为第四范式(4NF),此外还有第五范式等。范式越高,条件要求就越严,规范化的程度就越高。

在实际应用中,并非规范化的程度越高效果就会越好,也不是想规范化到多少范式就能规范化到多少范式。比如,分解时既保持函数依赖关系又保持无损连接性,那就只能达到 3NF;而如果只要求保持无损连接性,一定可达到 BCNF。我们应该根据具体情况全面地考虑数据库的规范化设计,规范化可以减少数据冗余和保持数据的一致性,但是规范化的程度越高,要增加的文件的个数就越多,当需要对多个文件进行复杂的连接操作时,不仅加大了系统的复杂性,而且也增加了 I/O 操作,使数据库的性能下降,影响了处理速度。

下边我们在 486(DX2-66,HD-270MB)兼容计算机上的中文 FOXBASE 关系数据库系统下就上边例子中的三种范式作一比较(假定有 1000 名学生):

在第一范式时,只需要一个数据库文件即可存放学生上机登记的信息,其所有操作也全靠这一个库即可。但是,正如上边所述,它存在着大量的数据冗余和潜在的更新异常,在对每一个上机者进行登记操作时,容易造成数据不一致性。

在第二范式时,需要二个数据库文件才能存放学生上机登记的信息,其某些操作也至少需要这二个库。它消除了多余的数据相关性,解决了潜在的更新异常问题,能保持数据的一致性。但是要连接形成所需的象第一范式那样的数据格式时,需要化费大量的机器时间,诸如上机登记库中有 1027 个记录,连接形成时间大约需要一分多钟(库中包含的记录越多,所需时间会大大增加),而对上机登记库中的所有记录的某些字段逐个记录地进行替换最多也不过需要一秒钟左右。此外仍存在数据冗余。

在第三范式时,需三个数据库文件才能存放学生上机登记的信息,其某些操作也至少需要这三个库。它除了具有第二范式的优点外,还消除了数据传递的相关性,解决了数据冗余问题。但是也和第二范式一样,随着文件的增多,当需要连接操作时,需要化费大量的机器时间。

综上所述,在实际针对某一问题设计关系模式时,分解到第三范式(3NF)已足够了,而且不一定非要分解到 3NF,这要视实际问题全面衡量,综合考虑。在计算机系统中,时间和空间本身就是一对矛盾,有时需要牺牲空间,获得系统的处理速度而换来时间,而有时需要牺牲时间而换取空间。

在开发研制一个关系数据库应用系统时,数据库的设计是至关重要的,数据库应用系统实质上是对数据信息的管理,数据信息的录入、存储、修改、查询、插入、删除和统计等这些基本的处理功能与其数据库的关系模式有着直接的关系,在数据库的结构确定之后,才能根据系统的要求开发编制各处理功能模块。因此,对一个数据库的关系模式来说,到底分解到第几范式为好,也就是说数据规范化到什么程度为最佳,要根据提出的具体问题的要求,不仅要考虑到软件环境,还要考虑应用系统的功能和开发的难易性,此外还应该考虑具体的硬件环境,必要时搞一点折中。

总之,在数据库结构的逻辑设计中,数据的规范化应以所提出的具体问题的要求为依据,以范式的理论为指导,结合自己的实际经验,不断地实践,不断地探索,注意有关环境的变化,就一定能够设计出切合实际的、规范化的、较佳的数据库结构来,以满足应用系统的要求。

(下转 118 页)

4.3 掺用活化胶粉的丁苯胶与掺用未活化胶粉的丁苯胶的物机性能相比,除扯断伸长率有所下降外,拉伸强度、撕裂强度、定伸应力、耐老化性能、抗屈挠能力、耐磨性均有显著提高。

参 考 文 献

1 Rubber Word. 188 (5). PP36—9 (1983)
2 彭震. 河南化工. 1 (1993)
3 柯新文等. 再生胶工业. 1 (1992)

A study of activation of rubber powder

Ding XueJia Shen Changyu Chen Jingpo

(Research Institute of mold Zhengzhou University of Techonlogy)

Abstract Activation material A and B through heating can change activation of rubber powder. SBR including activation of rubber powder have better mechanical properties except that tension set is lower than SBR including no activation of rubber powder.

Keywords SBR Rubber powder Activation

上接 114 页

参 考 文 献

1 G. J. DATE An Introduction to Database Systems (Third Edition)
2 冯玉才. 数据库系统基础 1986 华中工学院出版社
3 郎洪、杨国英. 自学汉字 FOXBASE+2.10 数据库系统教程. 1994. 学苑出版社

Inquiry Into Data Standardize To Relational
Database Application System

Lang Hong Chen Yongxiao
(Zhengzhou University of Technology)

Abstract The data standardize to the relational data base application system are in-quirde into.

Keywords Data base Application system Data standardize.