

5.1 数据库设计概述

5.1.1 数据库设计的定义和知识要求

数据库设计是指对于一个给定的应用环境,根据用户的需求,在某一具体的数据库管理系统上,构造一个性能良好的数据模式,建立数据库及其应用系统,使之能够有效地存储数据,满足各种用户的信息需求 and 处理需求。

(1) 信息需求。信息需求表示一个单位所需要的数据及其结构,表达了对数据库的内容及结构的要求,也就是静态要求。信息需求定义所设计的数据库将要用到的所有信息,描述实体、属性、联系的性质以及数据之间的联系。

(2) 处理需求。处理需求表示一个单位需要经常进行的数据处理,表达了基于数据库的数据处理要求,也就是动态要求。处理需求定义所设计的数据库将要进行的数据处理,描述操作的优先次序、操作执行的频率和场合以及操作与数据之间的联系。

因此,数据库设计就是把现实世界中的数据,根据各种应用处理的要求,加以合理地组织,使其满足硬件和操作系统的特性;同时,利用已有的 DBMS 建立数据库,使其能够实现应用系统的目标。

数据库设计是一个庞大而且复杂的工程,数据库设计人员应该具备以下知识。

(1) 计算机科学的基础知识和程序设计的方法与技巧。

作为一个数据库的设计人员,首先必须是一个懂得计算机专业的人员,而作为计算机专业的人员,最基本的就是计算机科学的基础知识,其次应该掌握关于程序设计的知识和程序设计的方法与技巧。

(2) 数据库的基础知识和数据库设计技术。

除了具有计算机的基础知识以外,作为数据库的设计人员必须具有数据库的基础知识和数据库设计技巧。

(3) 软件工程的原理和方法。

在数据库领域内,常常把使用数据库的各类系统称为数据库应用系统。数据库应用系统的开发应该遵循软件工程的方法和原理。尤其是大型数据库设计,其开发周期长、耗资大,失败的风险也大,必须把软件工程的原理和方法应用到数据库建设中来。

(4) 应用领域的知识。

应用领域的知识随着应用系统所属的领域不同而不同,如财务管理、仓库管理、人事管理、教务管理等。而且同样是教务管理,大学、中学、小学等不同类型的学校,各不相同,即使

都是高等学校各个学校的管理方式也不相同。因此,数据库设计人员必须深入实际与用户紧密结合,对应用环境、专业业务流程进行详细的调查研究,才能设计出符合具体应用领域和用户要求的数据库应用系统。

5.1.2 数据库设计的内容

数据库设计包括结构设计和行为设计两方面的内容。

1. 数据库的结构设计

数据库的结构设计是指根据给定的应用环境,进行数据库的模式或子模式的设计。它包括数据库的概念设计、逻辑设计和物理设计。数据库模式是各应用程序共享的结构,是静态的、稳定的,一经形成后通常情况下是不容易改变的,所以结构设计又称为静态模型设计。

2. 数据库的行为设计

数据库的行为设计是指确定数据库用户的行为和动作。在数据库系统中,用户的行为和动作指用户对数据库的操作,这些要通过应用程序来实现,所以数据库的行为设计就是应用程序的设计。用户的行为总是使数据库的内容发生变化,所以行为设计是动态的,行为设计又称为动态模型设计。

数据库的结构设计和行为设计是不能分离的,分离会导致数据与程序不易结合,增加数据库设计的复杂性。本章重点介绍数据库的结构设计,对于数据库的行为设计请参考“软件工程”相关书籍。

因为数据库的设计和开发是一个庞大而且复杂的工程,涉及多学科的综合性技术,所以,数据库设计是涉及硬件、软件和管理的技术,这也是数据库设计的另外一个特点。有人讲“三分技术,七分管理,十二分基础数据”是数据库建设的基本规律,这是有一定道理的。

5.1.3 数据库设计方法

数据库设计方法目前可分为4类:直观设计法、规范设计法、计算机辅助设计法和自动化设计法。

1. 直观设计法

直观设计法也叫手工试凑法,它是最早使用的数据库设计方法。这种方法依赖于设计者的经验和技巧,缺乏科学理论和工程原则的支持,设计的质量很难保证,常常是数据库运行一段时间后又发现了各种问题,这样就不得不修改原有设计,增加了系统维护的代价。因此这种方法越来越不适应信息管理发展的需要。

对于一个简单的程序设计过程来说,这样的方法具有周期短、效率高、操作简便、易于实现等优点。但是对于数据库设计,尤其是大型数据库系统的设计,由于其信息结构复杂、应用环境多样、应用需求全面等系统化综合性的要求,通常需要若干人的共同努力、相互协调,综合多种知识才能完成,因此,在具有丰富经验和设计技巧的前提下,还应该以严格的科学理论和软件工程设计原则为依托,完成数据库设计的全过程。

2. 规范设计法

规范设计法是将数据库设计分为若干阶段,明确规定各阶段的任务,采用自顶向下、分层实现、逐步求精的设计原则,结合数据库理论和软件工程设计方法,实现设计过程的每一

细节,最终完成整个设计任务。

1978年10月,来自三十多个国家的数据库专家在美国新奥尔良市专门讨论了数据库设计问题,他们运用软件工程的思想和方法,提出了数据库设计的规范,这就是著名的新奥尔良法,它是目前公认的比较完整和权威的一种规范设计法。新奥尔良法将数据库设计分成需求分析(分析用户需求)、概念设计(信息分析和定义)、逻辑设计(设计实现)和物理设计(物理数据库设计)。此后,S. B. Yao等提出了数据库设计的6个步骤:需求分析、模式构成、模式汇总、模式重构、模式分析和物理数据库设计,从而逐渐形成了数据库规范化设计方法。

目前,常用的各种数据库设计方法都属于规范设计法,即都是运用软件工程的思想和方法,根据数据库设计的特点,提出了各种设计原则与设计规程。常用的规范化设计方法主要有基于E-R模型的数据库设计方法、基于3NF的数据库设计方法和基于视图概念的数据库设计方法等。

1) 基于E-R模型的数据库设计方法

基于E-R模型的数据库设计方法是由P. P. S. Chen于1976年提出的数据库设计方法,其基本思想是在需求分析的基础上,用E-R图构造一个反映现实世界实体之间联系的企业模式,然后再将此企业模式转换为基于某一特定的DBMS的概念模式。

2) 基于3NF的数据库设计方法

基于3NF的数据库设计方法是一种结构化设计方法,其基本思想是在需求分析的基础上,确定数据库模式中的全部属性和属性间的依赖关系,将它们组织在一个单一的关系模式中,然后分析模式中不符合3NF的约束条件,将其进行投影分解,规范成若干3NF关系模式的集合。

其具体设计步骤分为以下5个阶段。

(1) 设计企业模式,利用规范化得到的3NF关系模式画出企业模式。

(2) 设计数据库的概念模式,把企业模式转换为DBMS所能接受的概念模式,并根据概念模式导出各个应用的外模式。

(3) 设计数据库的物理模式(存储模式)。

(4) 对物理模式进行评价。

(5) 实现数据库。

3) 基于视图的数据库设计方法

此方法先从分析各个应用的数据着手,其基本思想是为每个应用建立自己的视图,然后把这些视图汇总起来合并成整个数据库的概念模式。合并过程中要解决以下问题。

(1) 消除命名冲突。

(2) 消除冗余的实体和联系。

(3) 进行模式重构,在消除了命名冲突和冗余后,需要对整个汇总模式进行调整,使其满足全部完整性约束条件。

除了以上三种方法外,规范化设计方法还有实体分析法、属性分析法和基于抽象语义的设计方法等。规范设计法从本质上来说仍然是手工设计方法,其基本思想是过程迭代和逐步求精。

3. 计算机辅助设计法

计算机辅助设计法是指在数据库设计的某些过程中模拟某一规范化设计的方法,并以人的知识或经验为主导,通过人机交互方式实现设计中的某些部分。目前许多计算机辅助软件工程工具可以自动或辅助设计人员完成数据库设计过程中的很多任务。

4. 自动化设计法

自动化设计法是缩短数据库设计周期、加快数据库设计速度的一种方法。这种方法往往是直接用户,特别是非专业人员在数据库设计专业知识不太熟悉的情况下,较好地完成任务的一种捷径。例如,设计人员只要熟悉某种 MIS 辅助设计软件的使用,通过人机会话,输入原始数据和有关要求,无须人工干预,就可以由计算机系统自动生成数据库结构及相应的应用程序。由于该设计方法基于某一 MIS 辅助设计系统,从而受限于某种 DBMS,使得最终产生的数据库及其软件系统带有一定的局限性。此外,一个好的数据库模型往往需要设计者与用户反复商讨,是在用户的参与及合作下所形成的一个最终结果,设计者的经验及对应用部门的熟悉程度,在很大程度上是数据库设计质量的关键。因此,相对于其他设计方法而言,自动化设计法并不是一种理想的设计手段。下面围绕规范化设计法,深入分析和介绍其详细设计过程。

5.1.4 数据库设计的基本步骤

按照规范设计方法,考虑数据库及其应用系统开发全过程,并仿照软件生存周期,将数据库设计分为需求分析、概念结构设计、逻辑结构设计、物理结构设计、数据库实施以及数据库运行和维护 6 个阶段。数据库设计过程可以用图 5-1 表示。

1. 需求分析

需求分析是对具体应用环境的业务流程和用户提出的各种要求加以调查研究和分析,并和用户共同对各种原始数据加以综合、整理的过程,是形成最终设计目标的首要阶段,也是整个数据库设计过程中最困难的阶段。该阶段任务的完成,将为以后各阶段任务打下坚实的基础。因此,对用户的各种需求及数据,能否做出准确无误、充分完备的分析,并在此基础上形成最终目标,是整个数据库设计成败的关键。

2. 概念结构设计

概念结构设计是对用户信息需求所进行的进一步抽象和归纳,结果为数据库概念结构,通常用 E-R 模型来表示。

数据库的概念结构与 DBMS 和相关软硬件无关,它是对现实世界中具体数据的抽象,实现了从现实世界到信息世界的转换过程。概念结构设计是数据库设计的一个重要环节,是数据库的逻辑结构设计和物理结构设计的基础。

3. 逻辑结构设计

概念结构设计的结果是得到一个与 DBMS 无关的概念模式,而逻辑结构设计就是将概念模式转换为与选用的具体 DBMS 所支持的数据模型相符合的逻辑结构。所以,在逻辑结构设计阶段选择什么样的数据模型和哪一个具体 DBMS 尤为重要,它是能否满足用户各种要求的关键。

在逻辑结构设计阶段还有一个很重要的工作就是模式优化,该工作主要以用规范化理论为指导,目的是能够合理存放数据集合。逻辑结构设计阶段的模式优化,已成为影响数据

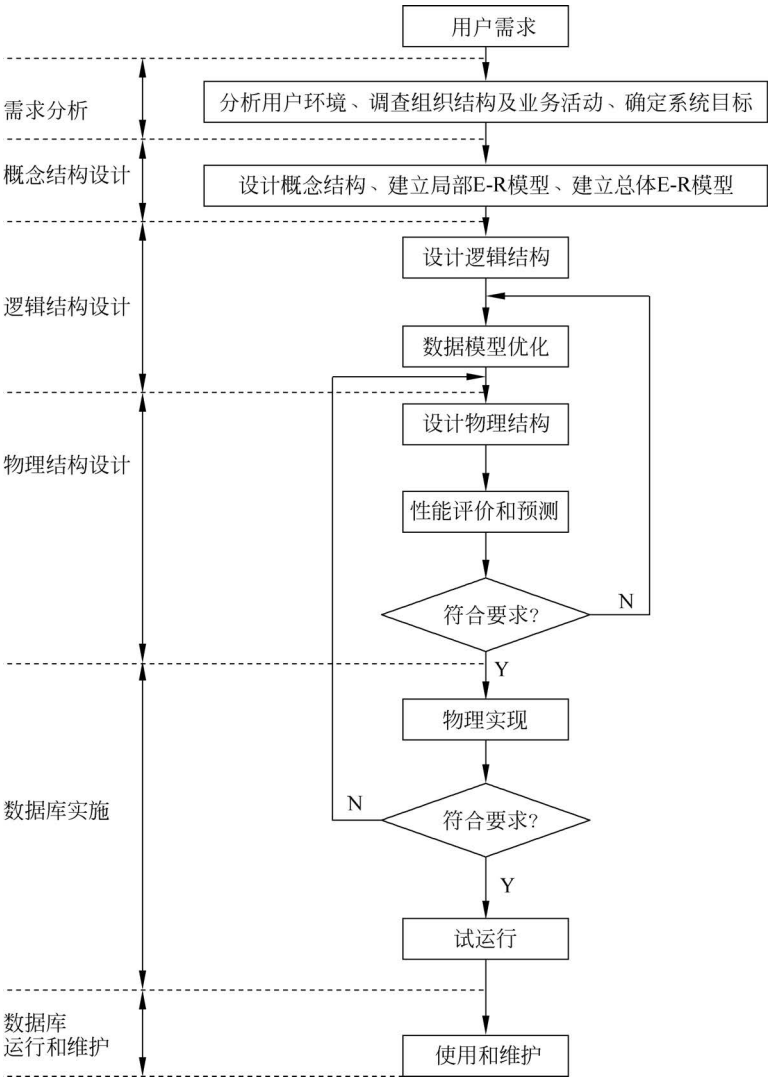


图 5-1 数据库设计步骤

库设计质量的一项重要工作。

4. 物理结构设计

数据库物理设计是将逻辑结构设计阶段所产生的逻辑数据模型,转换为某一计算机系统所支持的数据库物理结构的实现过程。

数据库的物理结构主要指数据库的存储记录格式、存储记录安排和存取方法,完全依赖于给定的硬件环境、具体的 DBMS 和操作系统。

存储记录格式的设计包括记录的组成,数据项的类型、长度,以及逻辑记录到存储记录的映射。存储记录的安排是指可以把经常同时被访问的数据组合在一起。存取方法的设计主要是指存取路径,存取路径分为主存取路径与辅存取路径,前者用于主码检索,后者用于辅助键检索。

除此之外,物理结构设计还要进行完整性和安全性考虑,设计者应在完整性、安全性、有

效性和效率方面进行分析,做出权衡。

完成物理结构设计后,对该物理结构做出相应的性能评价,若评价结果符合原设计要求,则进一步实现该物理结构;否则,对该物理结构做出相应的修改,若属于最初设计问题所导致的物理结构的缺陷,必须返回到概念设计阶段修改其概念数据模型或重新建立概念数据模型,如此反复,直至评价结果最终满足原设计要求为止。

5. 数据库实施

数据库实施阶段,即数据库调试、试运行阶段。一旦数据库物理结构形成,就可以用已选定的 DBMS 定义、描述相应的数据库结构,装入数据库数据,以生成完整的数据库,编制有关应用程序,进行联机调试并转入试运行,同时进行时间、空间等性能分析,若不符合要求,则需调整物理结构、修改应用程序,直至高效、稳定、正确地运行该数据库系统为止。

6. 数据库运行和维护

数据库实施阶段结束,标志着数据库系统投入正常运行的开始。严格地说,数据库运行和维护不属于数据库设计的范畴,早期的新奥尔良法明确规定数据库设计的 4 个阶段,不包括数据库运行和维护内容。随着人们对数据库设计的深刻了解和设计水平的不断提高,已经充分认识到数据库运行和维护工作与数据库设计的紧密联系。数据库设计是一种动态和不断完善的运行过程,运行和维护阶段开始,并不意味着设计过程的结束,哪怕只有细微的结构改变,也许就会引起对物理结构的调整、修改,甚至物理结构的完全改变,因此数据库运行和维护阶段是保证数据库日常活动的一个重要阶段。

5.2 需求分析

5.2.1 需求分析的任务

需求分析的任务是通过详细调查现实世界要处理的对象(组织、部门、企业等),充分了解原系统(手工系统或计算机系统)工作概况,明确用户的各种需求,然后在此基础上确定新系统的功能。新系统必须充分考虑今后可能的扩充和改变,不能仅按当前应用需求来设计数据库。

需求分析的重点是调查、收集与分析用户在数据管理中的信息要求、处理要求、安全性与完整性要求。

1. 信息要求

信息要求是指用户需要从数据库中获得信息的内容与性质。由用户的信息要求可以导出数据要求,即在数据库中需要存储哪些数据。

2. 处理要求

处理要求是指用户要求完成什么处理功能,如对处理的响应时间有什么要求、处理方式是批处理还是联机处理等。新系统的功能必须能够满足用户的信息要求、处理要求。

3. 安全性与完整性要求

数据库的安全性就是保护数据库,以防止因用户非法使用而造成的数据泄露、更改或者破坏。而完整性则保证数据库中数据的正确性和相容性。安全性和完整性同样是保障数据库应用系统正确、高效运行的重要内容。