

数据库设计

本章要点

数据库设计的目标就是根据特定的用户需求及一定的计算机软硬件环境,设计并优化数据库的逻辑结构和物理结构,建立高效、安全的数据库,为数据库应用系统的开发和运行提供良好的平台。

数据库技术是研究如何对数据进行统一、有效地组织、管理和加工处理的计算机技术,该技术已应用于社会方方面面,大到一个国家的信息中心,小到个体小企业,都会利用数据库技术对数据进行有效的管理,以达到提高生产效率和决策水平的目标。一个国家的数据库建设规模(指数据库的个数、种类),数据库的信息量的大小和使用频度已成为衡量这个国家信息化程度高低的重要标志之一。

本章详细介绍了设计一个数据库应用系统需经历的6个阶段,即系统需求分析、概念结构设计、逻辑结构设计、物理结构设计、数据库实施与数据库运行与维护。其中,概念结构设计和逻辑结构设计是本章的重点,也是本章的难点。

5.1 数据库设计概述

5.1.1 数据库设计的任务、内容和特点

1. 数据库设计的任务

数据库设计是指根据用户需求研制数据库结构并应用数据库的过程。具体地说,数据库设计是指对于给定的应用环境,构造最优的数据库模式,建立数据库及其应用系统,使之能有效地存储数据,满足用户的信息要求和处理要求,也就是把现实世界中的数据,根据各种应用处理的要求,加以合理组织,使之能满足硬件和操作系统的特性,利用已有的DBMS来建立能够实现系统目标的数据库。数据库设计的优劣直接影响信息系统的质量和运行效果。因此,设计一个结构优化的数据库是对数据进行有效管理的前提和正确利用信息的保证。



2. 数据库设计的内容

数据库设计内容包括数据库的结构设计和数据库的行为设计两方面。

数据库的结构设计是指根据给定的应用环境,进行数据库的模式设计或子模式的设计。它包括数据库的概念结构设计、逻辑结构设计和物理结构设计,即设计数据库框架或数据库结构。数据库结构是静态的,稳定的,一经形成后通常情况下是不容易也不需要改变的,因此,结构设计又称为静态模式设计。

数据库的行为设计是指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,数据库的行为设计就是操作数据库的应用程序的设计,即设计应用程序、事务处理等,因此,行为设计是动态的,行为设计又称为动态模式设计。

3. 数据库设计的特点

数据库设计既是一项涉及多学科的综合性的技术,又是一项庞大的软件工程项目,具有如下特点。

(1) 数据库建设是硬件、软件和干件(技术和管理的界面)的结合。

(2) 数据库设计应该与应用系统设计相结合,也就是说,要把行为设计和结构设计密切结合起来,是一种“反复探寻,逐步求精的过程”。首先从数据模型开始设计,以数据模型为核心展开,将数据库设计和应用设计相结合,建立一个完整、独立、共享、冗余小和安全、有效的数据库系统。

早期的数据库设计致力于数据模型和建模方法的研究,着重于应用中数据结构特性的设计,而忽视了对数据行为的设计。结构特性设计是指数据库总体概念的设计,所设计的数据库应具有最小数据冗余、能反映不同用户需求、能实现数据充分共享。行为特性是指数据库用户的业务活动,通过应用程序去实现。用户通过应用程序访问和操作数据库,用户的行为是和数据库紧密相关的。显然,数据库结构设计和行为设计必须相互参照进行。

5.1.2 数据库设计方法简述

数据库设计是一项工程技术,需要科学理论和工程方法作为指导,否则,工程的质量很难保证。为了使数据库设计更合理、更有效,人们通过努力探索,提出了各种各样的数据库设计方法,在很长一段时间内,数据库设计主要采用直观设计法。直观设计法也称为手工试凑法,是最早使用的数据库设计方法。这种方法与设计人员的经验和水平有直接的关系,缺乏科学理论和工程原则的支持,设计的质量很难保证,常常是数据库运行了一段时间以后又发现了各种问题,再进行重新修改,增加了维护的代价。因此,它不适应信息管理发展的需要,后来又提出了各种数据库设计方法,这些方法运用了软件工程的思想和方法,提出了数据库设计的规范,这些方法都属于规范设计方法,其中比较著名的有新奥尔良(New Orleans)法。它是目前公认的比较完整和权威的一种规范设计法,它将数据库设计分为4个阶段:需求分析(分析用户的需求)、概念结构设计(信息分析和定

义)、逻辑结构设计(设计的实现)和物理结构设计(物理数据库设计)。其后,S.B.Yao 等又将数据库设计分为 5 个步骤。大多数设计方法都起源于新奥尔良法,并在设计的每个阶段采用一些辅助方法来具体实现,下面简单介绍几种比较有影响的设计方法。

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的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与实体之间联系的企业模式,然后再将此企业模式转换成基于某一特定的 DBMS 的概念模式。

E-R 方法的基本步骤是:①确定实体类型;②确定实体联系;③画出 E-R 图;④确定属性;⑤将 E-R 图转换成某个 DBMS 可接受的逻辑数据模型;⑥设计记录格式。

2. 基于 3NF 的数据库设计方法

基于 3NF 的数据库设计方法的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与属性之间的依赖关系,将它们组织在一个单一的关系模式中,然后再将其投影分解,消除其中不符合 3NF 的约束条件,将其规范成若干 3NF 关系模式的集合。

3. 计算机辅助数据库设计方法

计算机辅助数据库设计是数据库设计趋向自动化的一个重要方面,其设计的基本思想不是要把人从数据库设计中赶走,而是提供一个交互式过程,一方面充分利用计算机的速度快、容量大和自动化程度高的特点,完成比较规则、重复性大的设计工作;另一方面又充分发挥设计者的技术和经验,做出一些重大的决策,人机结合,互相渗透,帮助设计者更好地进行数据库设计。常见的辅助设计工具有 ORACLE Designer、Sybase PowerDesigner、Microsoft Office Visio 等。

计算机辅助数据库设计主要分为需求分析、概念结构设计、逻辑结构设计、物理结构设计几个步骤。设计中,哪些可在计算机辅助下进行?能否实现全自动化设计呢?这是计算机辅助数据库设计需要研究的课题。

当然,除了介绍的几种方法以外,还有基于视图的数据库设计方法,基于视图的数据库设计方法是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,然后再把这些视图汇总起来,合并成整个数据库的概念模式。这里不再详细介绍。

5.1.3 数据库设计的步骤

按照规范化的设计方法,以及数据库应用系统开发过程,数据库的设计过程可分为以下 6 个设计阶段(见图 5.1):系统需求分析、概念结构设计、逻辑结构设计、物理结构设计、数据库实施、数据库运行与维护。

数据库设计中,前两个阶段是面向用户的应用要求,面向具体的问题,中间两个阶段是面向数据库管理系统,最后两个阶段是面向具体的实现方法。前四个阶段可统称为“分析和设计阶段”,后面两个阶段统称为“实现和运行阶段”。

数据库设计之前,首先必须选择参加设计的人员,包括系统分析人员,数据库设计人



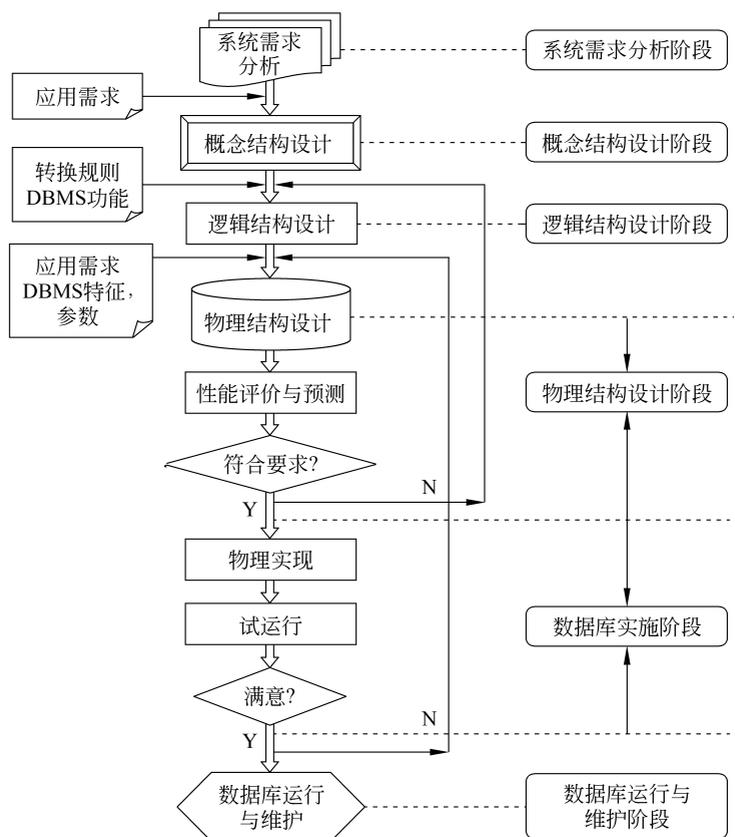


图 5.1 数据库设计步骤

员和程序员、用户和数据库管理员。系统分析和数据库设计人员是数据库设计的核心人员,他们将自始至终参加数据库的设计,他们的水平决定了数据库系统的质量。用户和数据库管理员在数据库设计中也是举足轻重的人物,他们主要参加需求分析和数据库的运行维护,他们的积极参与不但能加速数据库的设计,而且也是决定数据库设计是否成功的重要因素。程序员是在系统实施阶段参与进来,分别负责编制程序和准备软硬件环境。

如果所设计的数据库应用系统比较复杂,还应该考虑是否需要使用数据库设计工具和 CASE 工具,以提高数据库设计的质量,并减少设计工作量。

以下是数据库设计 6 个步骤的具体介绍。

1. 系统需求分析阶段

系统需求分析是指准确了解和分析用户的需求,这是最困难、最费时、最复杂的一步,但也是最重要的一步。它决定了以后各步设计的速度和质量。需求分析做得不好,可能会导致整个数据库设计返工重做。

2. 概念结构设计阶段

概念结构设计是指对用户的需求进行综合、归纳与抽象,形成一个独立于具体

DBMS 的概念模型,是整个数据库设计的关键。

3. 逻辑结构设计阶段

逻辑结构设计是指将概念模型转换成某个 DBMS 所支持的数据模型,并对其进行优化。

4. 物理结构设计阶段

物理结构设计是指为逻辑数据模型选取一个最适合应用环境的物理结构(包括存储结构和存取方法)。

5. 数据库实施阶段

数据库实施是指建立数据库,编制与调试应用程序,组织数据入库,并进行试运行。

6. 数据库运行与维护阶段

数据库运行与维护是指对数据库系统实际运行使用,并实时进行评价、调整与修缮。

可以看出,设计一个数据库不可能一蹴而就,它往往是上述各个阶段的不断反复。以上 6 个阶段是从数据库应用系统设计和开发的全过程来考察数据库设计的问题。因此,它既是数据库也是应用系统的设计过程。在设计过程中,努力使数据库设计和系统其他部分的设计紧密结合,把数据和处理的需求收集,分析,抽象,设计和实现在各个阶段同时进行、相互参照、相互补充,以完善数据和处理两方面的设计。按照这个原则,数据库各个阶段的设计可用图 5.2 描述。

设计各阶段	设计描述	
	数 据	处 理
系统需求分析	数据字典,全系统中数据项、数据流、数据存储的描述	数据流图和判定表(或判定树)、数据字典中处理过程的描述
概念结构设计	概念模型(E-R 图) 数据字典	系统说明书,包括: (1) 新系统要求、方案和概图; (2) 反映新系统信息的数据流图
逻辑结构设计	某种数据模型 关系模型	系统结构图 模块结构图
物理结构设计	存储安排 存取方法选择 存取路径建立	模块设计 IPO 表
数据库实施	编写模式 装入数据 数据库试运行	程序编码 编译连接 测试
数据库运行与维护	性能测试、转储/恢复数据库、重组和重构	新旧系统转换,运行,维护(修正性、适应性、改善性维护)

图 5.2 数据库各个设计阶段的描述

在图 5.2 中有关处理特性的描述中,采用的设计方法和工具属于软件工程和管理信息系统等课程中的内容,本书不再讨论,这里重点介绍数据特性的设计描述以及在结构特性中参照处理特性设计以完善数据模型设计的问题。

按照这样的设计过程,经历这些阶段能形成数据库的各级模式,如图 5.3 所示。在系统需求分析阶段,综合各用户的应用需求;在概念结构设计阶段形成独立于机器特点,独立于各 DBMS 产品的概念模型,在本书中就是 E-R 图;在逻辑结构设计阶段将 E-R 图转换成具体的数据库产品支持的数据模型,如关系模型中的关系模式;然后根据用户处理的要求、安全性完整性要求等,在基本表的基础上再建立必要的视图(可认为是外模式或子模式);在物理结构设计阶段,根据 DBMS 特点和处理性能等的需要,进行物理结构设计(如存储安排、建立索引等),形成数据库内模式;在数据库实施阶段开发设计人员基于外模式,进行系统功能模块的编码与调试;设计成功的话就进入系统的数据库运行与维护阶段。

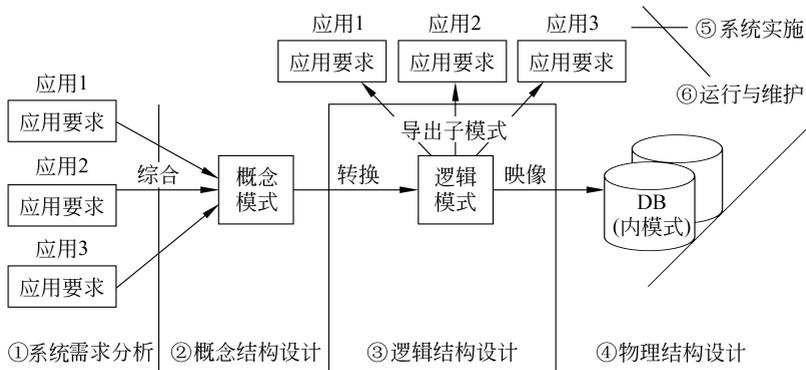


图 5.3 数据库设计过程与数据库各级模式

下面就以图 5.1 所示的规范化六步骤来进行介绍。

5.2 系统需求分析

需求分析简单的说是分析用户的要求,需求分析是设计数据库的起点,需求分析的结果是否准确地反映了用户的实际需求,将直接影响到后面的各阶段的设计,并影响到设计结果是否合理与实用。也就是说,如果这一步走得不对,获取的信息或分析结果就有误,那么后面的各步设计即使再优化也只能前功尽弃。因此,必须高度重视系统的需求分析。

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(如组织、部门、企业等),通过充分对原系统的工作概况的了解,明确用户的各种需求(如数据需求、完整性约束条件、事物处理和安全性要求等),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变化,不能只是仅仅按当前应用需求来设计数据库及其功能要求。

数据库需求分析的任务主要包括“数据或信息”和“处理”两方面。

(1) 信息要求:指用户需要从数据库中获得信息的内容与性质。由信息要求可以导出各种数据要求。

(2) 处理要求:指用户有什么处理要求(如响应时间、处理方式等),最终要实现什么处理功能。

具体而言,需求分析阶段的任务包括以下几方面。

1. 调查、收集、分析用户需求,确定系统边界

进行需求分析首先是调查清楚用户的实际需求,与用户达成共识。以确定这个目标的功能域和数据域。具体做法如下。

(1) 调查组织机构情况。包括了解该组织的部门组成情况、各部门的职责等,为分析信息流程做准备。

(2) 调查各部门的业务活动情况,包括了解各部门输入和使用什么数据,如何加工处理这些数据?输出什么信息?输出到什么部门?输出结果的格式是什么?这是调查的重点。

(3) 在熟悉业务的基础上,明确用户对新系统的各种要求,如信息要求,处理要求,完全性和完整性要求。因为,用户可能缺少计算机方面的知识,不知道计算机能做什么,不能做什么,从而不能准确地表达自己的需求。另外,数据库设计人员不熟悉用户的专业知识,不易理解用户的真正需求,甚至误解用户的需求,因此,设计人员必须不断与用户深入交流,才能完全掌握用户的真正需求。

(4) 确定系统边界。即确定哪些活动由计算机和将来由计算机来完成,哪些只能由人工来完成。由计算机完成的功能是新系统应该实现的功能。

2. 编写系统需求分析说明书

系统需求分析说明书也称为系统需求规范说明书,是系统分析阶段的最后工作,是对需求分析阶段的一个总结,编写系统需求分析说明书是一个不断反复、逐步完善的过程。系统需求分析说明书一般应包括如下内容。

- (1) 系统概况,包括系统的目标、范围、背景、历史和现状等;
- (2) 系统的原理和技术;
- (3) 系统总体结构和子系统结构说明;
- (4) 系统总体功能和子系统功能说明;
- (5) 系统数据处理概述、工程项目体制和设计阶段划分;
- (6) 系统方案及技术、经济、实施方案可行性等。

完成系统需求分析说明书后,在项目单位的主持下要组织有关技术专家评审说明书内容,这也是对整个需求分析阶段结果的再审查。审核通过后由项目方和开发方领导签字认同。

随系统需求分析说明书可提供以下附件。

- (1) 系统的软硬件支持环境的选择及规格要求(所选择的数据库管理系统、操作系

统、计算机型号及其网络环境等)。

(2) 组织机构图、组织之间联系图和各机构功能业务一览表图。

(3) 数据流程图、功能模块图和数据字典等图表。

系统需求分析说明书及其附件内容,一经双方确认,它们就是设计者和用户方的权威性文献,是今后各阶段设计与工作的依据,也是评判设计者是否完成项目的依据。

5.2.2 需求分析的方法

调查了解了用户的需求以后,还需要进一步分析和表达用户的需求,用于需求分析的方法有很多种,主要的方法有自顶向下和自底向上两种,其中自顶向下的结构化分析方法(Structured Analysis, SA)是一种简单实用的方法。SA方法是从最上层的系统组织入手,采用自顶向下、逐层分解的方法分析系统。

SA方法把每个系统都抽象成图 5.4 的形式。图 5.4 只是给出了最高层次抽象的系统概貌,要反映更详细的内容,可将处理功能分解为若干子系统,每个子系统还可以继续分解,直到把系统工作过程表示清楚为止。在处理功能逐步分解的同时,它们所用的数据也逐级分解,形成有若干层次的数据流图。

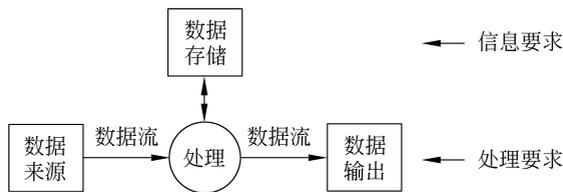


图 5.4 系统最高层数据抽象图

数据流图表达了数据和处理过程的关系。在 SA 方法中,处理过程的处理逻辑常常借助判定表和判定树来描述。系统中的数据则借助数据字典(DD)来描述。

下面介绍一下数据字典和数据流图。

1. 数据字典

数据流图表达了数据和处理的关系,数据字典则是系统中各类数据描述的集合,是各类数据结构和属性的清单。它与数据流图互为解释,数据字典贯穿于数据库需求分析直到数据库运行的全过程,在不同的阶段其内容形式和用途各有区别,在需求分析阶段,它通常包含以下 5 部分内容。

1) 数据项

数据项是不可再分的数据单位,对数据项的描述包括以下内容。

数据项描述={数据项名,数据项含义说明,别名,数据类型,长度,取值范围,取值含义,与其他数据项的逻辑关系,数据项之间的联系}。其中,取值范围、与其他数据项的逻辑关系定义了数据的完整性约束条件。

2) 数据结构反映了数据之间的组合关系

数据结构描述={数据结构名,含义说明,组成:{数据项或数据结构}}

3) 数据流

数据流是数据结构在系统内传输的路径。

数据流描述={数据流名,说明,数据流来源,数据流去向,组成:{数据结构},平均流量,高峰期流量}。

- 数据流来源是说明该数据流来自哪个过程。
- 数据流去向是说明该数据流将到哪个过程去。
- 平均流量是指在单位时间(每天、每周、每月等)里的传输次数。
- 高峰期流量则是指在高峰时期的数据流量。

4) 数据存储

数据存储是数据结构停留或保存的地方,也是数据流的来源和去向之一。

数据存储描述={数据存储名,说明,编号,流入的数据流,流出的数据流,组成:{数据结构},数据量,存取方式}

- 流入的数据流指出数据来源。
- 流出的数据流指出数据去向。
- 数据量:每次存取多少数据,每天(或每小时、每周等)存取几次等信息。
- 存取方法:批处理/联机处理;检索/更新;顺序检索/随机检索。

5) 处理过程

处理过程的具体处理逻辑一般用判定表或判定树来描述。数据字典中只需要描述处理过程的说明性信息。

处理过程描述={处理过程名,说明,输入:{数据流},输出:{数据流},处理:{简要说明}}

其中简要说明主要说明该处理过程的功能及处理要求。

- 功能要求:该处理过程用来做什么。
- 处理要求:处理频度要求(如单位时间里处理多少事务,多少数据量);响应时间要求等。

处理要求是后面物理结构设计的输入及性能评价的标准。

最终形成的数据流图和数据字典为“系统需求分析说明书”的主要内容,这是下一步进行概念结构设计的基础。

2. 数据流图

数据流图(Data Flow Diagram,DFD)表达了数据与处理的关系。

数据流图中的基本元素如下。

(1) 圆圈表示处理,输入数据在此进行变换产生输出数据。其中注明处理的名称。

(2) 矩形描述一个输入源点或输出汇点。其中注明源点或汇点的名称。

(3) 命名的箭头描述一个数据流。内容包括被加工的数据及其流向,流线上要注明数据名称,箭头代表数据流动方向。

(4) 向右开口的矩形框  表示文件和数据存储,要在其内标明相应的具体名称。

一个简单的系统可用一张数据流图来表示。当系统比较复杂时,为了便于理解,控制其复杂性,可以采用分层描述的方法,一般用第一层描述系统的全貌,第二层分别描述各子系统的结构。如果系统结构还比较复杂,那么可以继续细化,直到表达清楚为止,在处理功能逐步分解的同时,它们所用的数据也逐级分解,形成若干层次的数据流图。数据流图表达了数据和处理过程的关系。

5.3 概念结构设计

5.3.1 概念结构设计的必要性

将需求分析得到的用户需求抽象为信息结构(即概念模型)的过程就是概念结构设计,它是整个数据库设计的关键。概念结构设计以用户能理解的形式表达信息为目标,这种表达与数据库系统的具体细节无关,它所涉及的数据独立于 DBMS 和计算机硬件,可以在任何 DBMS 和计算机硬件系统中实现。

在进行功能数据库设计时,如果将现实世界中的客观对象直接转换为机器世界中的对象,就会感到比较复杂,注意力往往被牵扯到更多的细节限制方面,而不能集中在最重要的信息的组织结构和处理模式上,因此,通常是将现实世界中的客观对象首先抽象为不依赖任何 DBMS 支持的数据模型。故概念模型可以看成现实世界到机器世界的一个过渡的中间层次。概念模型是各种数据模型的共同基础,比数据模型更独立于机器、更抽象。将概念结构设计从设计过程中独立出来,可以带来以下好处。

- (1) 任务相对单一化,设计复杂程度大大降低,便于管理。
- (2) 概念模式不受具体的 DBMS 的限制,也独立于存储安排和效率方面的考虑,因此,更稳定。
- (3) 概念模型不含具体 DBMS 所附加的技术细节,更容易被用户理解,因而更能准确反映用户的信息需求。

设计概念模型的过程称为概念模型设计。

5.3.2 概念模型设计的特点

在需求分析阶段所得到的应用要求应该首先抽象为信息世界的结构,才能更好、更准确地用某一 DBMS 实现这些需求。

概念结构设计的特点有以下几点。

- (1) 易于理解,从而可以用它和不熟悉计算机的用户交换意见,用户的积极参与是数据库设计成功的关键。
- (2) 能真实、充分地反映现实世界,包括事物和事物之间的联系,能满足用户对数据的处理要求,是对现实世界的一个真实模型。