

第3章 数据库的设计

数据库设计是建立本课程领域的核心技术,也是信息系统开发和建设的关键环节,具体说,数据库设计是指对于一个给定的应用环境,构造最优的数据库模式、建立数据库及其应用系统,使之能够有效地存储数据,满足各种用户的应用需求(信息要求和处理要求)。这个问题是数据库在应用领域的主要研究课题。在数据库领域内,常常把使用数据库的各类系统统称为数据库应用系统。本章主要以培训类学校学生信息管理系统为案例,详细介绍了数据库设计的步骤以及各步骤之间的关系,从而指导学生以实际应用程序的角度来理解数据库设计的全过程并能够更深入地掌握数据库设计技术。

【案例分析】

下面以某培训类学校学生信息管理系统为例,来分析一下数据库设计的过程,该培训类学校主要涉及学生信息管理、学生选课、收费以及一些维护处理等功能,详见图 3-1 所示。

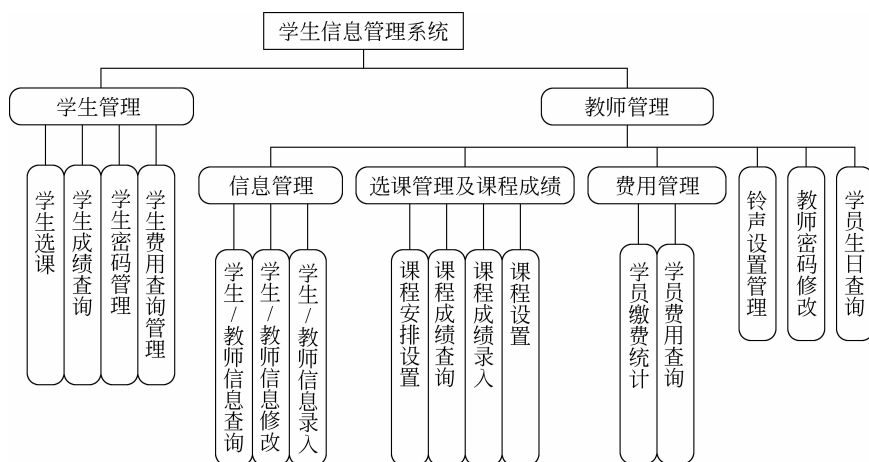


图 3-1 系统总功能

针对该软件各模块功能,应该如何来设计相应的数据库呢?下面将按照数据库的设计步骤来完成该软件系统的数据库设计。

3.1 数据库设计的基本步骤

数据库设计过程具有一定的规律和标准。在设计过程中,通常采用“分阶段法”,即“自顶向下,逐步求精”的设计原则。将数据库设计过程分解为若干相互依存的阶段,称之为步骤。每一阶段采用不同的技术、工具解决不同的问题,从而将一个大的问题局部化,减少局部问题对整体设计的影响及依赖,并利于多人合作。

目前数据库设计主要采用以逻辑数据库设计和物理数据库设计为核心的规范化设计方法。即将数据库设计分为:需求分析、概念结构设计、逻辑结构设计、数据库物理设计、数据

库实施、数据库运行和维护 6 个阶段,如图 3-2 所示。

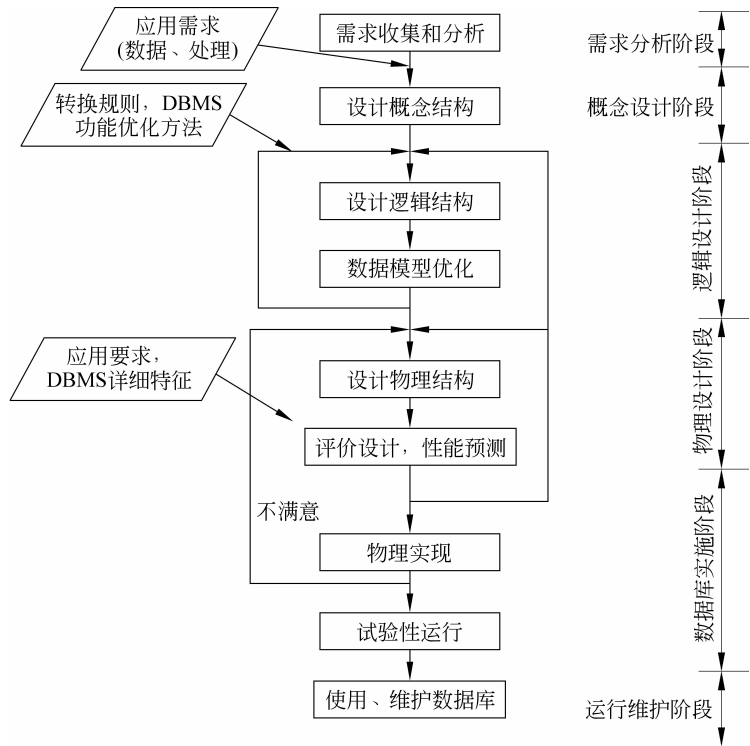


图 3-2 数据库设计步骤

1. 需求分析阶段

需求分析是对用户提出的各种要求加以分析,对各种原始数据加以综合、整理,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阶段,是以后各阶段任务的基础。因此,对用户的各种需求及数据,能否作出准确无误、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. 概念结构设计阶段

概念结构设计是对用户需求进行进一步抽象、归纳,并形成独立于 DBMS 和有关软、硬件的概念数据模型的设计过程,这是对现实世界中具体数据的首次抽象,完成从现实世界到信息世界的转化过程。数据库的逻辑结构设计和物理结构设计,都是以概念设计阶段所形成的抽象结构为基础进行的。因此,概念结构设计是数据库设计的一个重要环节。数据库的概念结构通常用 E-R 模型等来刻画。

3. 逻辑结构设计阶段

逻辑结构设计是将概念结构转化为某个 DBMS 所支持的数据模型,并进行优化的设计过程。由于逻辑结构设计是一个基于具体 DBMS 的实现过程,所以选择什么样的数据模型尤为重要,其次是数据模型的优化。数据模型有:层次模型、网状模型、关系模型、面向对象模型等,设计人员可选择其中之一,并结合具体的 DBMS 实现。在逻辑结构设计阶段后期的优化工作,已成为影响数据库设计质量的一项重要工作。

4. 数据库物理设计阶段

数据库物理设计阶段,是将逻辑结构设计阶段所产生的逻辑数据模型,转换为某种计算机系统所支持的数据库物理结构的实现过程。这里,数据库在相关存储设备上的存储结构和存取方法,称之为数据库的物理结构。完成物理结构设计后,对该物理结构作出相应的性能评价,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,对该物理结构作出相应的修改,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,直至评价结构最终满足原设计要求为止。

5. 数据库实施阶段

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 来定义、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. 数据库运行和维护阶段

数据库实施阶段结束,标志着数据库系统投入正常运行工作的开始。

严格地说,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,任何哪怕只有稍微的结构改变,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。

3.2 需求分析

需求分析是数据库设计的第一阶段,明确地把它作为数据库设计的第一步十分重要,这一阶段收集到的基础数据和一组数据流图(Data Flow Diagram,DFD)是下一步设计概念结构的基础。

3.2.1 需求描述与分析

原始数据库由于都比较简单,数据量小,所以设计人员主要将重点放到对数据库物理参数(如物理块大小、访问方法)的优化上。这种情况下,设计一个数据库所需要的信息常由一些简单的统计数据组成(如使用频率、数据量等)。目前,数据库应用越来越普及,而且结构也越来越复杂,整个企业可以在同一个数据库上运行。此时,为了支持所有用户的运行,数据库设计就变得异常复杂。要是没有对信息进行充分的事先分析,这种设计将很难取得成功。因此,需求分析工作就被置于数据库设计过程的前沿。

从数据库设计的角度考虑,要求分析阶段的目标是:对现实世界要处理的对象(组织,部门,企业等)进行全面的、详细的调查,确定企业组织的目标,收集支持系统总的设计目标的基础数据和对这些数据的要求,确定用户的需求,确定新系统功能,并把这些要求写成用户和数据库设计者都能够接受的文档。

需求分析阶段应该对系统的整个应用情况作详细的调查,需求分析中调查分析的方法

很多,通常的办法是对不同层次的企业管理人员进行个人访问,内容包括业务处理和企业组织中的各个数据。访问的结果应该包括数据的流程、过程之间的接口以及访问者和职员两方面对流程和接口语义上的核对说明和结论。某些特殊的目标和数据库的要求,应该从企业组织中的最高层机构得到。

需求分析阶段必须强调用户的参与,本阶段一个重要而困难的任务是设计人员还应该了解系统将来要发生的变化,收集未来应用所涉及的数据,充分考虑到系统可能的扩充和变动,使系统设计更符合未来发展的趋向,并且易于改动,以减少系统维护的代价。

3.2.2 需求分析分类

在需求分析总体包括两类,即信息需求和处理需求,如图 3-3 所示。

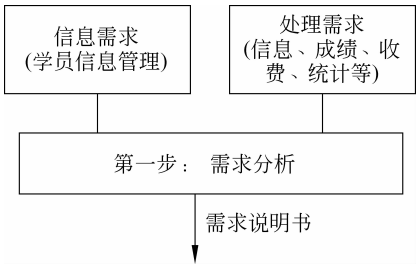


图 3-3 需求分析阶段的输入输出

信息需求定义了未来系统用到的所有信息,描述了数据之间本质上和概念上的联系,描述了实体、属性、组合及联系的性质。

在“学生信息管理系统”开发过程中,首先要做的是要了解该系统所涉及的各类体,属性以及它们之间的关系,诸如学生、教师、课程等关系。

处理需求中定义了未来系统的数据处理的操作,描述了操作的优先次序、操作执行的频率和场合、操作与数据之间的联系。

在信息需求和处理需求的定义说明的同时还应定义安全性和完整性约束。

这一阶段的输出是“需求说明书”,其主要内容是系统的数据流图和数据字典。需求说明书应是一份既切合实际,又具有远见的文档,是一个描述新系统的轮廓图。

3.2.3 需求分析的内容与方法

1. 需求分析的内容

(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。包括了解各个部门输入和使用什么数据、如何加工处理这些数据、输出什么数据、输出到什么部门、输出结果的格式等,这是调查的重点。

(3) 熟悉业务活动,并协助用户明确对新系统的各种要求,包括信息要求、处理要求、安全性与完整性要求,这是调查的另一重点。

(4) 确定新系统的边界。对前面的调查结果进行初步分析,确定哪些功能由计算机完成,哪些功能由人工完成。由计算机完成的功能就是新系统应完成的功能。

2. 需求分析的方法

需求分析的方法有多种,下面归纳了 6 种主要方法。

- (1) 跟班作业。
- (2) 开调查会。
- (3) 请专人介绍。

- (4) 找人询问。
- (5) 设计调查表请用户填写。
- (6) 查阅记录。

在本案例中,首先做的是去一所相关的学校做实地调查,咨询了校长、教师、学生的一些实际情况,并对学生的场地、上课时间、课程设置、费用等做了详尽的调查,并以此做了记录,供下一阶段分析使用。

3.2.4 需求分析的步骤

1. 分析用户活动,产生用户活动图

这一步主要了解用户当前的业务活动和职能,分析其处理流程(即业务流程)。如果一个业务流程比较复杂,就要把处理分解成若干个子处理,使每个处理功能明确、界面清楚,分析之后画出用户活动图(即用户的业务流程图)。

2. 确定系统范围,产生系统范围图

这一步是确定系统的边界。在和用户经过充分讨论的基础上,确定计算机所能进行数据处理的范围,确定哪些工作由人工完成,哪些工作由计算机系统完成,即确定人机界面。

3. 分析用户活动所涉及的数据,产生数据流图

在这一过程中,要深入分析用户的业务处理过程,以数据流图形式表示出数据的流向和对数据所进行的加工。

数据流图(Data Flow Diagram,DFD)是从“数据”和“对数据的加工”两方面表达数据处理系统工作过程的一种图形表达法,具有直观、易于被用户和软件人员双方都能理解的一种表达系统功能的描述方式。

DFD有4个基本成分:数据流(用箭头表示)、加工或处理(用圆圈表示)、文件(用双线线段表示)和外部实体(数据流的源点或终点,用方框表示)。图3-4是一个简单的DFD。

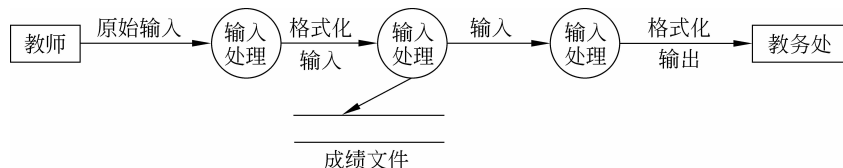


图 3-4 一个简单的 DFD

在众多分析和表达用户需求的方法中,自顶向下逐步细化是一种简单实用的方法。为了将系统的复杂度降低到人们可以掌握的程度,通常把大问题分割成若干个小问题,然后分别解决,这就是“分解”。分解也可以分层进行,既先考虑问题最本质的属性,暂把细节略去,以后再逐层添加细节,直到涉及最详细的内容,这称之为“抽象”。

DFD可作为自顶向下逐步细化时描述对象的工具。顶层的每一个圆圈(加工处理)都可以进一步细化为第二层;第二层的每一个圆圈都可以进一步细化为第三层;……直到最底层的每一个圆圈已表示一个最基本的处理动作为止。DFD可以形象地表示数据流与各业务活动的关系,它是需求分析的工具和分析结果的描述手段。

图3-5给出了某学校学生课程管理子系统的数据流图。该子系统要处理的工作是学生

根据开设课程提出选课送教务部门审批,对已批准的选课单进行上课安排,教师对学生上课情况进行考核,给予平时成绩和允许参加考试资格,对允许参加考试的学生根据考试情况给予考试成绩和总评成绩。

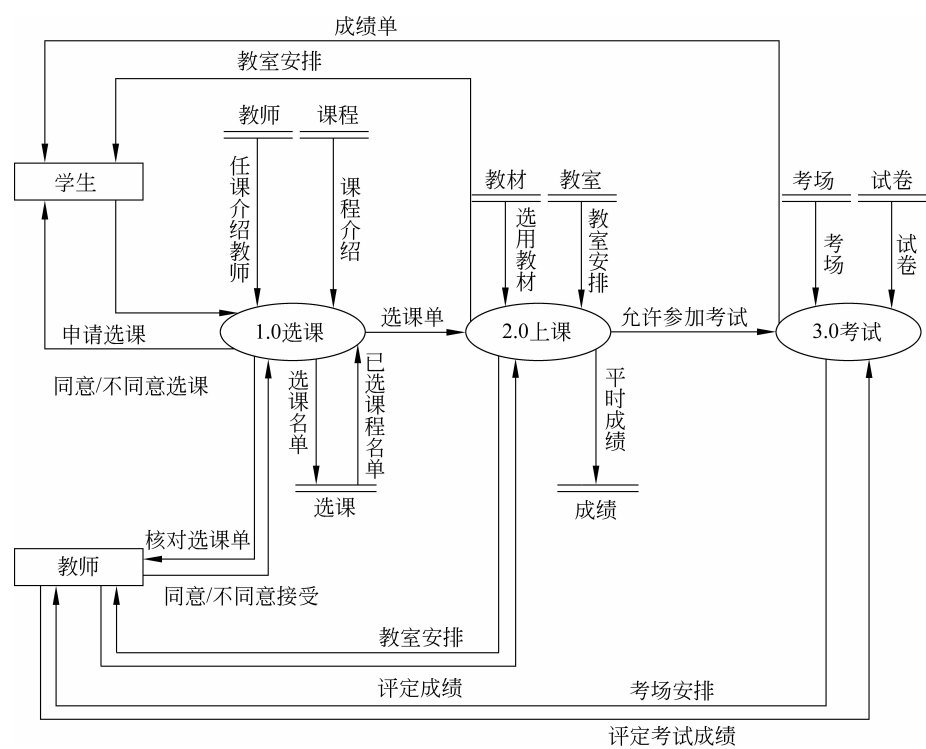


图 3-5 某学校学生课程管理子系统的数据流程图

4. 分析系统数据,产生数据字典

仅仅有 DFD 并不能构成需求说明书,因为 DFD 只表示出系统有哪几部分组成和各部分之间的关系,并没有说明各个成分的含义。只有对每个成分都给出确切定义后,才能较完整地描述系统。

3.2.5 数据字典

数据字典提供对数据库描述的集中管理,它的功能是存储和检索各种数据描述(称为元数据 Meta Data),如叙述性的数据定义等,并且为 DBA 提供有关的报告。对数据库设计来说,数据字典是进行详细的数据收集和数据分析所获得的主要成果。因此在数据库设计中占有很重要的地位。

数据字典中通常包括数据项、数据结构、数据流、数据存储和处理过程 5 个部分。其中数据项是数据的最小组成单位,若干个数据项可以组成一个数据结构,数据字典通过对数据项和数据结构的定义来描述数据流以及数据存储的逻辑内容。

1. 数据项

数据项是数据的最小单位,对数据项的描述,通常包括数据项名、含义、别名、类型、长度、取值范围以及其他数据项的逻辑关系等。

【例 3-1】 在图 3-6 中有一个数据流选课单,每张选课单有一个数据项为选课单号。在数据字典中可对此数据项作如下描述。

2. 数据结构

数据结构反映了数据之间的组合关系。一个数据结构可以由若干个数据项组成,也可以由若干个数据结构组成,或有若干个数据项和数据结构混合而成。它包括数据结构名、含义及组成该数据结构的数据项名或数据结构名。

数据项名:	选课单号
说 明:	标识每张选课单
类 型:	CHAR(8)
长 度:	8
别 名:	选课单号
取值范围:	00 000 001~99 999 999

图 3-6 一个数据流选课单

3. 数据流

数据流可以是数据项,也可以是数据结构,表示某一加工处理过程的输入或输出数据。对数据流的描述应包括数据流名、说明、流出的加工名、流入的加工名以及组成该数据流的数据结构或数据项。

【例 3-2】 在图 3-7 中考场安排是一个数据流,在数据字典中可对考场安排描述如下。

数据流名:	考场安排
说 明:	由各课程所选学生数,选定教室、时间,确定考试安排。
来 源:	考试
去 向:	教师
数据结构:	考场安排
	—— 考试课程
	—— 考试时间
	—— 教学楼
	—— 教室编号

图 3-7 考场安排数据流

【例 3-3】 在例 3-2 中描述了数据流“考场安排”的细节,在数据字典中,对于数据结构“考试课程”还有如下详细的说明,如图 3-8 所示。

数据结构名:	考试课程
说 明:	作为考试安排的组成部分,说明某门课程哪位老师代,以及所选学生人数。
组 成:	课程号、教师号
选课人数	

图 3-8 考试课程的详细说明

4. 数据存储

数据存储是处理过程中要存储的数据,可以是手工凭证、手工文档或计算机文档。对数据存储的描述应包括:数据存储名、说明、输入数据流、输出数据流、数据量(每次存取取多少数据)、存取频率(单位时间内存取次数)和存取方法(是批处理,还是联机处理;是检索,还是更新;是顺序存取,还是随机存取)等。

【例 3-4】 图 3-7 中课程是个数据存储,在数据字典中可对其作如下描述,如图 3-9 所示。

数据存储名:	课程
说 明:	对每门课程的名称、学分、先行课程号和摘要的描述。
输出数据流:	课程介绍
数 据 描 述:	课程号
	课程名
	学分数
	先行课程号
	摘要
数 量:	每年 500 种
存 取 方 式:	随机存取

图 3-9 课程数据字典的描述信息

5. 加工过程

对加工处理的描述包括加工过程名、说明、输入数据流、并简要说明处理工作、频度要求、数据量及响应时间等。

【例 3-5】 对图 3-7 中的“选课”，在数据字典中可对其作如下描述，如图 3-10 所示。

处理过程:	确定选课名单
说 明:	对要选某门课程的每一个学生,根据已选修课程确定其是否可选该课程。 在根据学生选课的人数选择适当的教室,制定选课单。
输 入:	学生选课 可选课程 已选课程
输 出:	选课单
程序提要:	a. 对所选课程在选课表中查找其是否已选此课程。 b. 若未选过此课程,则在选课表中查找是否已选此课程的先行课程。 c. 若 a,b 都满足,则在选课表中增加一条选课记录。 d. 处理完全部学生的选课处理后,形成选课单。

图 3-10 选课数据字典的描述信息

数据字典是在需求分析阶段建立,并在数据库设计过程中不断改进、充实和完善。

3.3 概念设计

概念设计的目标是产生反映企业组织信息需求的数据库概念结构,即概念模式。概念模式是独立于计算机硬件结构,独立于支持数据库的 DBMS。

3.3.1 概念设计的必要性及要求

在进行数据库设计时,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,就会感到非常不方便,注意力往往被转移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具体机器的信息结构,这种信息结构不是 DBMS 支持的数据模型,而是概念模型。然后再把概念模型转换成具体机器上 DBMS 支持的数据模型,设计概念模式的过程称为概念

设计。

1. 将概念设计从数据库设计过程中独立开来的优点

(1) 各阶段的任务相对单一化,设计复杂程度大大降低,便于组织管理。

(2) 不受特定的 DBMS 的限制,也独立于存储安排和效率方面的考虑,因而比逻辑模式更为稳定。

(3) 概念模式不含具体的 DBMS 所附加的技术细节,更容易为用户所理解,因而才有可能准确地反映用户的信息需求。

2. 概念模型的要求

(1) 概念模型是对现实世界的抽象和概括,应真实、充分地反映现实世界中事物和事物之间的联系,有丰富的语义表达能力,能表达用户的各种需求,包括描述现实世界中各种对象及其复杂的联系、用户对数据对象的处理要求的手段。

(2) 概念模型应简洁、清晰、独立于机器、容易理解,方便数据库设计人员与应用人员交换意见,使用户能积极参与数据库的设计工作。

(3) 概念模型应易于变动。当应用环境和应用要求改变时,容易对概念模型进行适当的修改和补充。

(4) 概念模型应很容易向关系、层次或网状等各种数据模型转换。易于从概念模式导出与 DBMS 有关的逻辑模式。

选用何种模型完成概念结构设计任务,是进行概念数据库设计前应考虑的首要问题。用于概念设计的模型既要有足够的表达能力,使之可以表示各种类型的数据及其相互间的联系和语义,又要简明易懂,能够为非专业数据库设计人员所接受。这种模型有很多种,如 20 世纪 70 年代提出的 E-R 模型,以及后来提出的语义数据模型、函数数据模型等。其中,E-R 模型提供了人们对数据模型描述既标准、规范,又具体、直观的构造手法,从而使得 E-R 模型成为应用最广泛的数据库概念结构设计工具。

3.3.2 概念设计的方法与步骤

1. 概念设计的方法

(1) 自顶向下方法。根据用户要求,先定义全局概念结构的框架,然后分层展开,逐步细化。

(2) 自底向上方法。根据用户的每一具体需求,先定义各局部应用的概念结构,然后将它们集成,逐步抽象化,最终产生全局概念结构。

(3) 逐步扩张方法。先定义最重要的核心概念结构,然后向外扩充以滚雪球的方式逐步生成其他概念结构,直至全局概念结构。

(4) 混合方式方法。将自顶向下和自底向上相结合,先用自顶向下方式设计一个全局概念结构框架,再以它为基础,采用自底向上法集成各局部概念结构。

在上一节介绍需求分析的实现方法中,较为普遍的是采用自顶向下法描述数据的层次结构化联系。但在概念结构的设计过程中却截然相反,自底向上法是普遍采用的一种设计策略。因此,在对数据库的具体设计过程中,通常先采用自顶向下法进行需求分析,得到每一集体的应用需求,然后反过来根据每一子需求,采用自底向上法分步设计产生每一局部的 E-R 模型,综合各局部 E-R 模型,逐层向上回到顶端,最终产生全

局 E-R 模型。

2. 概念设计的步骤

1) 进行数据抽象,设计局部概念模式

局部用户的信息需求,是构造全局概念模式的基础。因此,需要先从个别用户的需求出发,为每个用户建立一个相应的局部概念结构。在建立局部概念结构时,常常要对需求分析的结果进行细化、补充和修改,如有的数据项要分为若干子项,有的数据的定义要重新核实等。

2) 将局部概念模式综合成全局概念模式

综合各局部概念结构就可得到反映所有用户需求的全局概念结构。在综合过程中,主要处理各局部模式对各种对象定义的不一致问题,包括同名异义、异名同义和同一事物在不同模式中被抽象为不同类型的对象(例如,有的作为实体,有的又作为属性)等问题。把各个局部结构连接、合并,还会产生冗余问题,有可能导致对信息需求的再调整与分析,以确定准确的含义。

3) 评审

消除了所有冲突后,就可把全局结构提交评审。评审分为用户评审与 DBA 及应用开发人员评审两部分。用户评审的重点放在确认全局概念模式是否准确完整地反映了用户的信息需求和现实世界事物的属性间的固有联系;DBA 和应用开发人员评审则侧重于确认全局结构是否完整,各种成分划分是否合理,是否存在不一致性,以及各种文档是否齐全等。文档应包括局部概念结构描述、全局概念结构描述、修改后的数据清单和业务活动清单等。

3.3.3 E-R 模型的操作

在利用 E-R 模型进行数据库概念设计的过程中,常常需要对 E-R 图进行种种变换。这些变换又称为对 E-R 模型的操作,包括实体类型、联系类型和实体属性的分裂、合并、增加和删除等。

1. 实体类型的分裂

一个实体类型可以根据需要分裂成若干个实体类型。分裂方式有垂直分割和水平分割两种形式。

1) 垂直分割

垂直分割是指把一个实体类型的属性分成若干组,然后按组形成若干实体类型。例如图 3-11 中,可以把教师实体类型中经常变动的一些属性组成一个新的实体类型,而把固定不变的属性组成另一个实体类型。但应注意,在垂直分割中,码必须在分割后的各实体类型中都出现。

2) 水平分割

水平分割是指把一个实体类型分裂为互不相交的子类(即得到原实体类型的一个分割)。如对于有些数据库,不同的应用关心不同的内容,则可以将记录型水平分割成两个记录型。

这样可减少应用存取的逻辑记录数。例如可把教师实体类型水平分割为男教师与女教师两个实体类型,如图 3-12 所示。