

合理的数据库结构是数据库应用系统性能良好的基础和保证,但数据库的设计和开发却是一项庞大而复杂的工程。从事数据库设计的人员不仅要具备数据库知识和数据库设计技术,还要有程序开发的实际经验,掌握软件工程的原理和方法;数据库设计人员必须深入应用环境,了解用户具体的专业业务;在数据库设计的前期和后期与应用单位人员密切联系、共同开发,可大大提高数据库设计的成功率。

本章主要讲述数据库设计过程中的需求分析、概念结构设计、逻辑结构设计、物理结构设计、数据库实施、运行和维护等内容以及两个数据库设计实例。

3.1 数据库设计概述

数据库设计是根据用户需求设计数据库结构的过程。具体一点来讲,数据库设计是对于给定的应用环境在关系数据库理论的指导下构造最优的数据库模式,在数据库管理系统上建立数据库及其应用系统,使之能有效地存储数据,满足用户的各种需求的过程。

数据库设计方法有多种,概括起来分为 4 类,即直观设计法、规范设计法、计算机辅助设计法和自动化设计法。按照规范设计的方法,考虑数据库及其应用系统开发全过程,数据库设计可分为 6 个阶段,即需求分析阶段、概念结构设计阶段、逻辑结构设计阶段、物理结构设计阶段、数据库实施阶段以及数据库运行和维护阶段。

3.2 需求分析

需求分析是数据库设计的起点,需求分析就是数据库设计人员通过仔细地调查和向用户详细地咨询掌握用户的需求、理解用户的需求。需求分析的结果是否准确地反映了用户的实际要求将直接影响到后面各个阶段的设计,并影响到设计结果是否合理、实用。如果投入大量的人力、物力、财力和时间开发出的软件却没人要,那所有的投入都是徒劳。如果费了很大的精力开发一个软件,最后却不满足用户的要求,从而要重新开发,这种返工是让人痛心疾首的。

总之,需求分析为数据库的开发起到了决策的作用,提供了开发的方向,并指明了开发的策略,在数据库开发及维护中均起到了举足轻重的作用。可以说在一个大型数据库系统的开发中,它的作用要远远大于其他各个阶段。记住数据库设计的合理、可行,满足用户需求才是最重要的。

3.2.1 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)的工作概况,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,不能仅仅按当前的应用需求来设计数据库。

调查的重点是“数据”和“处理”,通过调查、收集与分析获得用户对数据库的以下要求。

(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数据要求,即在数据库中需要存储哪些数据。

(2) 处理要求:指用户要完成什么处理功能,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。

(3) 安全性与完整性要求:安全性要求是指对数据库的用户、角色、权限、加密方法等安全保密措施的要求。完整性要求是指对数据的取值范围、数据之间各种联系的要求等。

确定用户的最终需求往往是一件很困难的事,这是因为一方面用户缺少计算机知识,开始时无法确定计算机究竟能为自己做什么,不能做什么,往往不能准确地表达自己的需求,所提出的需求往往不断变化;另一方面,设计人员缺少用户的专业知识,不易理解用户的真正需求,甚至误解用户的需求。因此,设计人员必须不断深入地与用户交流,才能逐步确定用户的实际需求。

3.2.2 需求分析的方法

进行需求分析首先是调查清楚用户的实际要求,与用户达成共识,然后分析与表达这些需求。

1. 调查用户需求的具体步骤

(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) 在熟悉业务活动的基础上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、完全性与完整性要求,这是调查的又一个重点。

(4) 确定新系统的边界,对前面调查的结果进行初步分析,确定哪些功能由计算机完成或将来准备让计算机完成,哪些活动由人工完成。由计算机完成的功能就是新系统应该实现的功能。在调查过程中可以根据不同的问题和条件使用不同的调查方法。

2. 常用的调查方法

(1) 跟班作业:通过亲身参加业务工作来了解业务活动的情况。这种方法可以比较准确地理解用户的需求,但比较耗费时间。

(2) 开座谈会:通过与用户座谈来了解业务活动情况及用户需求。在座谈时,参加者之间可以相互启发,一般可按职能部门组织座谈会。

(3) 询问或请专人介绍:一般应包括领导、管理人员、操作员等。

(4) 设计调查表请用户填写需求:如果调查表设计得合理,这种方法很有效,也易于用户接受。

(5) 查阅记录:查阅与原系统有关的数据记录。

在做需求调查时往往需要同时使用上述多种方法。但无论使用何种调查方法,都必须有用户的积极参与和配合,最好能建立由双方人员参加的项目实施保障小组负责沟通联系。

3.2.3 数据流图和数据字典

数据流图(Data Flow Diagram,DFD)和数据字典(Data Dictionary,DD)是对需求分析结果进行描述的两个主要工具。

1. 数据流图

数据流图表达了数据和处理过程的关系,反映的是对事务处理所需的原始数据和经处理后的数据及其流向。在结构化分析方法中,任何一个系统都可以抽象成图 3-1 所示的数据流图。

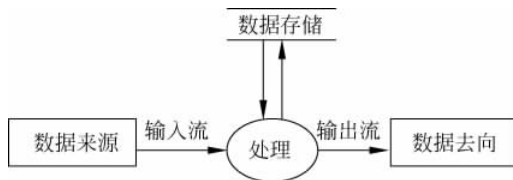


图 3-1 数据流图

1) 数据流

在数据流图中,用箭头表示数据流,数据流由一组确定的数据组成;名字表示流经的数据,箭头表示数据流动的方向。

2) 处理

在数据流图中,用圆圈表示处理,处理是对数据进行的操作。

3) 数据存储

在数据流图中,用双线段表示存储的信息,文件数据暂时存储或永久保存的地方可以是数据库存储文件,如学生表、开课计划表等。

4) 外部实体

外部实体指独立于系统存在的,但又和系统有联系的实体。它表示数据的外部来源和最后的去向,确定系统与外部环境之间的界限,从而可确定系统的范围。在数据流图中,用矩形表示外部实体。外部实体可以是某人员、组织、系统或某事物。

数据流图清楚地表达了数据与处理之间的关系。在结构化分析方法中,处理过程常常借助判定表或判定树来描述,而系统中的数据则用数据字典来描述。

2. 数据字典

数据字典是数据库系统中各类数据详细描述的组合。在数据库设计中,它提供了对各类数据描述的集中管理,是一种数据分析、系统设计和管理的有力工具。数据字典要有专人或专门小组进行管理,及时对数据字典进行更新,保证字典的安全、可靠。

数据字典通常包括数据项、数据存储、数据流和数据加工 4 个部分。这里以货物销售网站的数据库为例介绍数据字典的设计。

1) 数据项

数据项是最小的数据单位,通常包括属性名、含义、别名、类型、长度、取值范围、与其他数据项的逻辑联系等。数据项的定义格式如下。

数据项名称: 货物编号

别名: Goods-No

含义：公司所有货物的编号

类型：字符串

长度：10

其取值范围及含义如下。

第1位字符：进口/国产

第2~4位字符：类别

第5~7位字符：规格

第8~10位字符：品名编号

2) 数据存储

数据存储是数据停留并保存的地方,也是数据流的来源和去向之一。它可以是手工文档或凭单,也可以是计算机文档。它包括数据存储名、说明、输入/输出数据流、组成的成分(数据结构或数据项)、存取方式、操作方式等。数据存储的定义格式如下。

数据存储名称：库存

别名：Inventory

说明：货物存放于仓库的情况

组成：货物编号+仓库编号+存放位置+库存量

输入：进货单

输出：供货单

存取方式：索引文件,以货物编号+仓库编号为关键字

操作方式：立即查询

3) 数据流

数据流表示数据项或数据结构在某一加工过程中的输入或输出。数据流包括数据流名、说明、输入/输出的加工名、组成的成分。数据流的定义格式如下。

数据流名称：订单

说明：顾客订货时填写的项目

输入：顾客提交

输出：订单检验

组成：编号+订货日期+顾客编号+地址+电话+银行账号+货物名称+规格+数量

4) 数据加工

数据加工的具体处理逻辑一般用判定表或判定树来描述,包括加工名称、说明、输入/输出数据流、加工过程简介等。加工条目是用来说明 DFD 中基本加工的处理逻辑的,由于上层的加工是由下层的基本加工分解而来,只要有了基本加工的说明,就可以理解其他加工。基本加工的定义格式如下。

加工名：查阅库存

操作条件：接收到合格订单时

输入：合格订单

输出：可供货订单或缺货订单

加工过程：根据订单数量和总库存数量可以判断供货或缺货

数据字典是在需求分析阶段建立,在数据库设计过程中不断修改、充实、完善的。

3.3 概念结构设计

系统需求分析报告反映了用户的需求,但只是现实世界的具体要求,这是远远不够的。我们还要将其转换为信息(概念)世界的结构,这就是概念设计阶段所要完成的任务。

数据库概念结构设计是整个数据库设计的关键,此阶段要做的工作不是直接将需求分析得到的数据格式转换为 DBMS 能处理的数据模型,而是将需求分析得到的用户需求抽象为反映用户观点的概念模型。以此作为各种数据模型的共同基础,从而能更好、更准确地用某一 DBMS 实现这些需求。

描述概念结构的模型应具有以下几个特点。

(1) 有丰富的语义表达能力:能表达用户的各种需求,反映现实世界中的各种数据及其复杂的联系,以及用户对数据的处理要求等。

(2) 易于交流和理解:概念模型是系统分析师、数据库设计人员和用户之间主要的交流工具。

(3) 易于修改:概念模型能灵活地加以改变,以反映用户需求和环境的变化。

(4) 易于向各种数据模型转换:设计概念模型的最终目的是向某种 DBMS 支持的数据模型转换,建立数据库应用系统。

人们提出了多种概念设计的表达工具,其中最常用、最有名的是 E-R 模型。

3.3.1 概念结构设计的方法

概括起来,设计概念模型的总体策略和方法可以归纳为下面 4 种。

(1) 自顶向下法:首先认定用户关心的实体及实体间的联系,建立一个初步的概念模型框架,即全局 E-R 模型,然后再逐步细化,加上必要的描述属性,得到局部 E-R 模型。

(2) 自底向上法:有时又称属性综合法,先将需求分析说明书中的数据元素作为基本输入,通过对这些数据元素的分析把它们综合成相应的实体和联系,得到局部 E-R 模型,然后在此基础上进一步综合成全局 E-R 模型。

(3) 逐步扩张法:先定义最重要的核心概念 E-R 模型,然后向外扩充,以滚雪球的方式逐步生成其他概念 E-R 模型。

(4) 混合策略:将单位的应用划分为不同的功能,每一种功能相对独立,针对各个功能设计相应的局部 E-R 模型,最后通过归纳合并消除冗余与不一致,形成全局 E-R 模型。

其中最常用的策略是自底向上法,即先进行自顶向下的需求分析,再进行自底向上的概念设计。

3.3.2 概念结构设计的步骤

概念结构设计可以分为两步:进行数据抽象,设计局部 E-R 模型,即设计局部视图;集成各局部 E-R 模型,形成全局 E-R 模型,即视图的集成。

1. 设计局部 E-R 模型

局部 E-R 模型的设计步骤包括以下 4 步。

(1) 确定局部 E-R 图描述的范围:根据需求分析所产生的文档可以确定每个局部 E-R 图描述的范围。通常采用的方法是将单位的功能划分为几个系统,将每个系统又分为几个子系

统。设计局部 E-R 模型的第一步就是划分适当的系统或子系统,划分的过细或过粗都不太合适,划分的过细将造成大量的数据冗余和不一致,过粗有可能漏掉某些实体。

用户一般可以遵循以下两条原则进行功能的划分。

- 独立性原则:划分在一个范围内的应用功能具有独立性与完整性,与其他范围内的应用有最少的联系。
- 规模适度原则:局部 E-R 图的规模应适度,一般以 6 个左右实体为宜。

(2) 确定局部 E-R 图的实体:根据需求分析说明书将用户的数据需求和处理需求中涉及的数据对象进行归类,指明对象的身份是实体、联系还是属性。

(3) 定义实体的属性:根据上述实体的描述信息来确定其属性。

(4) 定义实体间的联系:在确定了实体及其属性后就可以定义实体间的联系了。实体间的联系按其特点可分为 3 种,即存在性联系(如学生有所属的班级)、功能性联系(如教师要教学生)、事件性联系(如学生借书)。实体间的联系方式分为一对一、一对多、多对多 3 种。

在设计完成某一局部结构的 E-R 模型后,看还有没有其他的局部 E-R 模型,有则转到第(2)步继续,直到所有的局部 E-R 模型都设计完为止。

2. 局部 E-R 模型的集成

由于局部 E-R 模型反映的只是单位局部子功能对应的数据视图,可能存在不一致的地方,还不能作为逻辑设计的依据,这时可以去掉不一致和重复的地方,将各个局部视图合并为全局视图,即局部 E-R 模型的集成。

一般来说,视图集成有两种方式:第一种是多个分 E-R 图一次集成;第二种是逐步集成,用累加的方式一次集成两个分 E-R 图。第一种方式比较复杂,做起来难度较大;第二种方式每次只集成两个分 E-R 图,可以降低复杂度。

无论采用哪种集成方法,每一次集成都分为两步:第一步是合并,以消除各局部 E-R 图之间不一致的情况,生成初步的 E-R 图;第二步是优化,消除不必要的冗余数据,包括冗余的数据和实体间冗余的联系,生成全局 E-R 图。

3.4 逻辑结构设计

数据库概念设计阶段得到的数据模式是用户需求的形式化,它独立于具体的计算机系统和 DBMS。为了建立用户所要求的数据库,必须把上述数据模式转换成某个具体的 DBMS 所支持的数据模式,并以此为基础建立相应的外模式,这是数据库逻辑设计的任务,是数据库结构设计的重要阶段。

逻辑设计的主要目标是产生一个 DBMS 可处理的数据模型和数据库模式。该模型必须满足数据库的存取、一致性及运行等各方面的用户需求。

其逻辑结构设计阶段一般要分 3 步进行:将 E-R 图转化为关系数据模型、关系模式的优化、设计用户外模式。

3.4.1 将 E-R 图转化为关系数据模型

关系数据模型是一组关系模式的集合,而 E-R 图是由实体、属性和实体之间的联系三要素组成的,所以将 E-R 图转化为关系数据模型实际上是将实体、属性和实体之间的联系转化为关系模式。

在转化过程中要遵循以下原则：

1. 实体的转换

一个实体转化为一个关系模式，实体的属性就是该关系模式的属性，实体的主码就是该关系模式的主码。

2. 联系的转换

在 E-R 图中，用菱形表示实体间的联系，用无向边分别把菱形与有关实体连接起来，并在无向边旁注明了联系的类型。而在关系数据模型中，实体和联系都用关系来表示，所以需要把联系转换到关系模式中。

(1) 两实体集间的 1:1 联系可以转换为一个独立的关系模式，也可以与任意一端对应的关系模式合并。

① 转换为一个独立的关系模式：转换后的关系模式中关系的属性包括与该联系相连的各实体的主码以及联系本身的属性，关系的主码为两个实体的主码的组合。

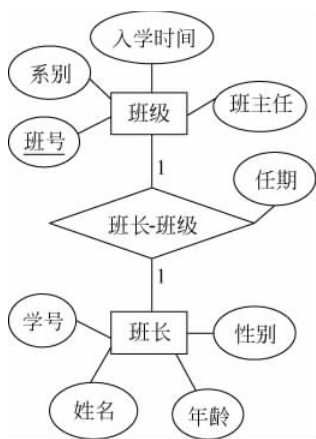


图 3-2 1:1 联系的 E-R 图

【例 3-1】 将图 3-2 所示的 E-R 图按方法①转换为关系模式。

转换成以下关系模式：

班级(班号, 系别, 班主任, 入学时间)

班长(学号, 姓名, 性别, 年龄)

班级-班长(班号, 学号, 任期)

② 与某一端对应的关系模式合并：合并后关系模式的属性包括自身关系模式的属性和另一关系模式的主码及联系本身的属性，合并后关系的主码不变。

【例 3-2】 将图 3-2 所示的 E-R 图按方法②转换为关系模式。

转换成以下关系模式：

班级(班号, 系别, 班主任, 入学时间)

班长(学号, 姓名, 性别, 年龄, 班号, 任期)

或：

班级(班号, 系别, 班主任, 入学时间, 班长学号, 班长任期)

班长(学号, 姓名, 性别, 年龄)

(2) 两实体集间的 1:N 联系可以转换为一个独立的关系模式，也可以与 N 端对应的关系模式合并。

① 转换为一个独立的关系模式：关系的属性包括与该联系相连的各实体的主码以及联系本身的属性，关系的主码为 N 端实体的主码。

【例 3-3】 将图 3-3 所示的 E-R 图按方法①转换为关系模式。

对应的关系模式如下：

系(系号, 系名, 系主任)

教师(教师号, 教师名, 年龄, 职称)

工作(教师号, 系号, 入系日期)

② 与 N 端对应的关系模式合并：合并后关系的属性包括在 N 端关系中加入 1 端关系的

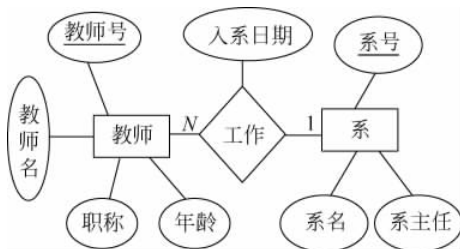


图 3-3 1:N 联系的 E-R 图

主码和联系本身的属性,合并后关系的主码不变。

【例 3-4】 将图 3-3 所示的 E-R 图按方法②转换为关系模式。

对应的关系模式如下:

系(系号,系名,系主任)

教师(教师号,教师名,年龄,职称,系号,入系日期)

注意: 在实际使用中,两实体集间的 1:1 和 1:N 联系通常都采用方法②进行转换,以减少关系模式,因为多一个关系模式就意味着在查询过程中要进行连接运算,从而降低查询的效率。

(3) 同一实体集内实体间的 1:N 联系:可以在这个实体集所对应的关系模式中多设一个属性,用来作为与该实体相联系的另一个实体的主码。

比如学生实体集中有部分学生是班长,就可以在学生关系模式中加入一个“班长学号”属性,表示这个学生所在班的班长。

(4) 两实体集间的 M:N 联系:必须为联系产生一个新的关系模式,在该关系模式中至少包含被它所联系的双方实体的主码,若联系中有属性,也要并入该关系模式中。如果没有指定另外的属性(例如此联系的 ID)作为该关系的主码,则该关系的主码为双方实体的主码的组合。

【例 3-5】 将图 3-4 所示的 E-R 图转换成对应的关系模式。

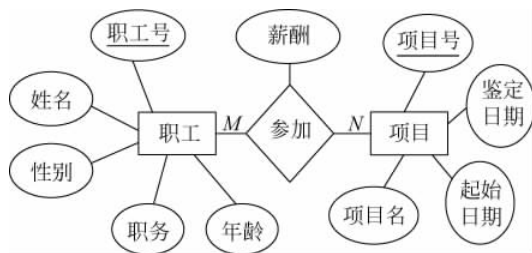


图 3-4 两实体集间 M:N 联系的 E-R 图

对应的关系模式如下:

职工(职工号,姓名,性别,年龄,职务)

项目(项目号,项目名,起始日期,鉴定日期)

参加(职工号,项目号,薪酬)

(5) 同一实体集内实体间的 M:N 联系:必须为联系产生一个新的关系模式,在该关系模式中至少包含被它所联系的双方实体的主码,若联系有属性,也要并入该关系模式中。

例如学生实体集中有部分学生是社团的负责人,而一个学生可以参加多个社团,一个社团

也可以有多个学生,所以学生实体集中的学生和社团负责人就存在 $M:N$ 联系。学生实体集转换为一个学生关系,学生实体集内实体间的 $M:N$ 联系就生成一个新的关系模式,在该关系中至少包含“学号”和“社团负责人编号”两个属性。

(6) 两个以上实体集之间的 $M:N$ 联系:必须为联系产生一个新的关系模式,在该关系模式中至少包含被它所联系的所有实体的主码,若联系有属性,也要并入该关系模式中,关系的主码可以指定一个单独的属性(例如此联系的 ID),否则为它所联系的所有实体的主码的组合。

3.4.2 关系模式的优化

在通常情况下,数据库逻辑设计的结果不是唯一的。为了进一步提高数据库应用系统的性能,还应努力减少关系模式中存在的各种异常,改善完整性、一致性和存储效率。规范化理论是数据库逻辑设计的重要理论基础和进行关系模式优化的有力工具,规范化的具体过程详见第2章。

为了提高数据库应用系统的性能,对规范化后的关系模式还需要进行修改、调整结构,这就是关系模式的进一步优化,通常采用合并或分解的方法。比如两个或多个关系经常进行连接查询,在确定不会造成数据操作异常的前提下可以对它们的关系模式进行合并;又如一个数据记录非常多的关系,可以将其关系模式进行横向分解,将旧数据和新数据分别存储;再如一个属性非常多的关系,可以将其关系模式进行纵向分解,将常用属性和不常用属性分别存储。

关系模式的优化方法如下:

- (1) 确定函数依赖。
- (2) 对于各个关系模式之间的函数依赖进行极小化处理,消除冗余的联系。
- (3) 按照函数依赖的理论对关系模式逐一进行分析,考查是否存在部分函数依赖、传递函数依赖等,确定各关系模式分别属于第几范式,对不符合规范的关系模式进行规范化。
- (4) 按照需求分析阶段得到的各种应用对数据处理的要求,分析对于这样的应用环境这些模式是否合适,确定是否要对它们进行合并或分解。
- (5) 对关系模式进行必要的合并或分解。

规范化理论为数据库设计人员判断关系模式的优劣提供了理论标准,可用来预测关系模式可能出现的问题,使数据库设计工作有了严格的理论基础。

3.4.3 设计用户外模式

外模式也叫子模式,是用户可直接访问的数据模式。在同一系统中,不同用户可以有不同的外模式。外模式来自逻辑模式,但在结构和形式上可以不同于逻辑模式,所以它不是逻辑模式简单的子集。

外模式的作用为:通过外模式对逻辑模式屏蔽,为应用程序提供了一定的逻辑独立性;可以更好地适应不同用户对数据的需求;为用户划定了访问数据的范围,有利于数据的保密等。

定义数据库全局模式主要是从系统的时间效率、空间效率、易维护等角度出发。由于用户外模式与模式是相对独立的,因此在定义用户外模式时要注意考虑用户的习惯与方便。这些习惯与方便如下:

- (1) 使用符合用户习惯的别名。
- (2) 可以对不同级别的用户定义不同的外模式,以保证系统的安全性。
- (3) 简化用户对系统数据的使用。

如果某些局部应用中经常要使用一些很复杂的查询,为了方便用户,可以将这些复杂的查询定义为外模式,这样用户就可以每次只对定义好的外模式进行查询,从而大大方便了用户的使用。

3.5 物理结构设计

数据库最终要存储在物理设备上,将逻辑设计中产生的数据库逻辑模型结合指定的 DBMS 设计出最适合应用环境的物理结构的过程称为数据库的物理结构设计。

为了设计数据库的物理结构,设计人员必须充分了解所用 DBMS 的内部特征;充分了解数据系统的实际应用环境,特别是数据应用处理的频率和响应时间的要求;充分了解外存储设备的特性。

数据库的物理结构设计分为下面两个步骤:

- (1) 确定数据库的物理结构。
- (2) 对所设计的物理结构进行评价。

如果所设计的物理结构的评价结果满足原设计要求则可进入到物理实施阶段,否则需要重新设计或修改物理结构,有时甚至要返回逻辑设计阶段修改数据模型。

3.5.1 确定数据库的物理结构

数据库物理设计内容包括确定数据的存储结构、设计数据的存取路径、确定数据的存放位置和确定系统配置。

1. 确定数据的存储结构

在确定数据库的存储结构时要综合考虑存取时间、存储空间利用率和维护代价 3 个方面的因素。这 3 个方面常常是相互矛盾的,例如消除一切冗余数据虽然能够节约存储空间,但往往会导致检索代价的增加,因此必须进行权衡,选择一个折中方案。通常,确定数据的存储结构包括为各行记录分配连续或不连续的物理块等。

2. 设计数据的存取路径

DBMS 常用的存取方法有索引方法(目前主要是 B+ 树索引方法)、聚簇(CLUSTER)方法和 HASH(哈希)方法。

1) 索引方法

在关系数据库中,选择存取路径主要是指确定如何建立索引。例如,应把哪些域作为次码建立次索引,是建立单码索引还是组合索引,建立多少个合适,是否建立聚集索引等。

2) 聚簇方法

为了提高某个属性(或属性组)的查询速度,把这个或这些属性(称为聚簇码)上具有相同值的元组集中存放在连续的物理块上称为聚簇。

聚簇的用途:大大提高按聚簇属性进行查询的效率;聚簇功能不仅适用于单个关系,也适用于多个关系,假设用户经常要按系别查询学生成绩单,这一查询涉及学生关系和选修关系的连接操作,即需要按学号连接这两个关系,为提高连接操作的效率,可以把具有相同学号值

的学生元组和选修元组在物理上聚簇在一起,这就相当于把多个关系按“预连接”的形式存放,从而大大提高连接操作的效率。

3) HASH 方法

有些数据库管理系统提供了 HASH 存取方法。如果一个关系的属性主要出现在等值连接条件中或主要出现在相等比较选择条件中,而且满足下列两个条件之一,则此关系可以选择 HASH 存取方法。

(1) 该关系的属性主要出现在等值连接条件中或主要出现在相等比较选择条件中。

(2) 该关系的大小可预知且关系的大小不变,或该关系的大小动态改变但所选用的 DBMS 提供了动态 HASH 存取方法。

3. 确定数据的存放位置

为了提高系统性能,用户应该把数据根据应用情况将易变部分与稳定部分分磁盘存放、把经常存取部分和存取频率较低部分分磁盘存放,以及把数据表和索引分磁盘存放、把数据和日志分磁盘存放,等等。

4. 确定系统配置

DBMS 产品一般都提供了一些存储分配参数,供设计人员和 DBA 对数据库进行物理优化。在初始情况下,系统都为这些变量赋予了合理的默认值,但是这些值不一定适合每一种应用环境。

对于系统配置的变量,例如同时使用数据库的用户数、同时打开的数据库对象数、缓冲区分配参数、物理块装填因子、数据库的大小、锁的数目等,在进行物理设计时应根据应用环境确定这些参数值,以使系统的性能最佳。

3.5.2 评价物理结构

在数据库物理设计过程中需要对时间效率、空间效率、维护代价和各种用户要求进行权衡,可能产生多种方案,数据库设计人员必须对这些方案进行细致的评价,从中选择一个较优的方案作为数据库的物理结构。

评价物理数据库的方法完全依赖于所选用的 DBMS,主要是从定量估算各种方案的存储空间、存取时间和维护代价入手对估算结果进行权衡、比较,选择出一个较优的、合理的物理结构。如果该结构不符合用户需求,则需要修改设计。

3.6 数据库的实施、运行和维护

在数据库正式投入运行之前还需要完成很多工作,例如在模式和子模式中加入对数据库安全性、完整性的描述,完成应用程序和加载程序的设计,数据库系统的试运行,并在试运行中对系统进行评价。如果评价结果不能满足用户要求,还需要对数据库进行修正设计,直到用户满意为止。数据库正式投入使用并不意味着数据库设计生命周期的结束,而是数据库维护阶段的开始。

3.6.1 数据库的实施

根据逻辑和物理设计的结果在计算机上建立起实际的数据库结构并装入数据、进行试运行和评价的过程叫数据库的实施(或实现)。

1. 建立实际的数据库结构

用 DBMS 提供的数据库定义语言 (DDL) 编写描述逻辑设计和物理设计结果的程序 (一般称为数据库脚本程序), 经计算机编译处理和执行后就生成了实际的数据库结构。

2. 数据的加载

数据库应用程序的设计应该与数据库设计同时进行。通常, 应用程序的设计应该包括数据库加载程序的设计。在数据加载前必须对数据进行整理。由于用户缺乏计算机应用背景的知识, 常常不了解数据的准确性对数据库系统正常运行的重要性, 因而未对提供的数据做严格的检查。所以在数据加载前要建立严格的数据登录、录入和校验规范, 设计完善的数据校验与校正程序, 排除不合格数据。

3. 数据库的试运行和评价

在加载了部分必需的数据和应用程序之后, 就可以开始对数据库系统进行联合调试了, 称为数据库的试运行。一般将数据库的试运行和评价结合起来, 目的如下:

- (1) 测试应用程序的功能;
- (2) 测试数据库的运行效率是否达到设计目标, 是否为用户所容忍。

测试的目的是为了发现问题, 而不是为了说明能达到哪些功能, 所以在测试中一定要有非设计人员的参与。

对于数据库系统的评价比较困难, 需要估算不同存取方法的 CPU 服务时间及 I/O 服务时间。为此, 一般从实际试运行中进行估价, 确认其功能和性能是否满足设计要求, 对空间占用率和时间响应是否满意等。

3.6.2 数据库的运行与维护

数据库试运行结果符合设计目标之后, 数据库就可以真正投入运行了。数据库投入运行标志着开发任务的基本完成和维护工作的开始。

对数据库设计进行评价、调整、修改等维护工作是一个长期的任务, 也是设计工作的继续和提高。

概括起来, 维护工作包括数据库的转储和恢复; 数据库的安全性和完整性控制; 数据库性能的监督、分析和改造; 数据库的重组织和重构造。

3.7 数据库设计实例

3.7.1 图书借阅管理系统设计

1. 需求分析

与用户协商, 了解用户的需求, 了解需要哪些数据和操作 (主要是查询), 确定系统中应包含书籍、员工、部门和出版社实体。

书籍的属性确定为图书号、分类、书名、作者、单价、数量; 员工的属性确定为工号、姓名、性别、出生年月; 部门的属性确定为部门号、部门名称、电话; 出版社的属性确定为出版社名、地址、电话、联系人。

其中, 每个员工可以借阅多本书, 每本书也可以由多个员工借阅, 每个员工每借一本书都有一个借阅日期和应还日期; 每个员工只属于一个部门; 每本图书只能由一个出版社出版。

2. 概念结构设计

画出图书借阅管理系统的 E-R 图,如图 3-5 所示。

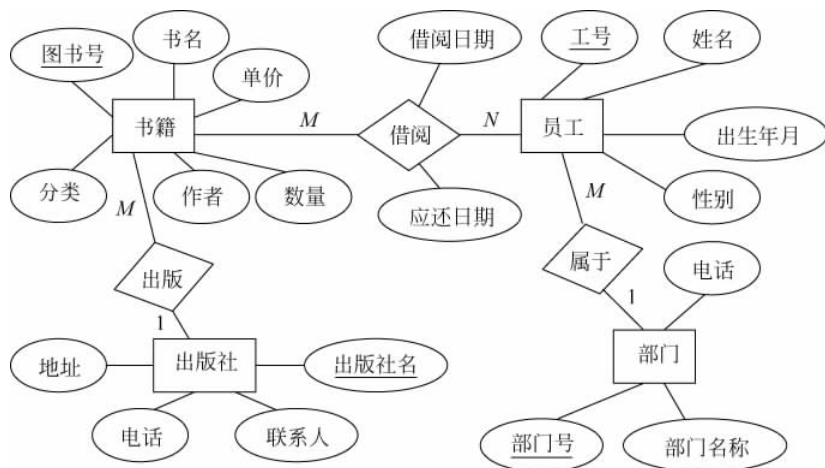


图 3-5 图书借阅管理系统的 E-R 图

3. 逻辑结构设计

根据前面的转换原则,图书借阅管理系统的关系模式设计如下:

书籍(图书号,分类,书名,作者,出版社名,单价,数量)

员工(工号,部门号,姓名,性别,出生年月)

部门(部门号,部门名称,电话)

出版社(出版社名,地址,电话,联系人)

借阅(工号,图书号,借阅日期,应还日期)

表结构设计如表 3-1~表 3-5 所示。

表 3-1 书籍表的表结构

属性名	类型	宽度	主键	取值范围
图书号	字符型	15	是	
分类	字符型	10		
书名	字符型	30		
作者	字符型	10		
出版社名	字符型	20		参考出版社表主键
单价	实型			0.00~999.99
数量	整型			

表 3-2 员工表的表结构

属性名	类型	宽度	主键	取值范围
工号	字符型	5	是	
部门号	字符型	4		参考部门表主键
姓名	字符型	10		
性别	字符型	2		(男,女)
出生年月	日期型			1956-1-1~2050-1-1

表 3-3 部门表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
部门号	字符型	4	是	
部门名	字符型	20		
电话	字符型	11		数字字符

表 3-4 出版社表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
出版社名	字符型	20	是	
地址	字符型	40		
电话	字符型	11		数字字符
联系人	字符型	10		

表 3-5 借阅表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
工号	字符型	5	是	参考员工表主键
图书号	字符型	15		参考书籍表主键
借阅日期	日期型			2012-1-1~2030-1-1
应还日期	日期型			2012-1-1~2030-2-1

4. 物理结构设计

根据查询需求设计每一个关系的索引文件。

3.7.2 钢材仓库管理系统设计

1. 需求分析

与用户协商,了解用户的需求,了解需要哪些数据和操作(主要是查询),确定系统中应包含职工、仓库、钢材和供应商实体。

职工的属性确定为工号、姓名、性别、出生年月、工种(销售员、采购员、仓库管理员);仓库的属性确定为仓库编号、仓库名称、地址、联系电话、容量;钢材的属性确定为钢材号、钢材名、品种、规格;供应商的属性确定为供应商编号、供应商名称、地址、电话、联系人。

其中,一种钢材可以存放于多个仓库内,一个仓库也可以存放多种钢材;一个供应商可以供应多种钢材,一种钢材也可以由多个供应商供应,每个供应商供应一种钢材有一个报价;钢材、仓库与销售员之间有销售关系,它们是多对多的,每个销售员销售每个仓库的每种钢材都有一个销售单号、出库数量和出库日期;采购员、钢材与仓库之间有采购关系,它们是多对多的,每个采购员采购每种钢材都有一个入库单号、入库数量和入库日期;每个仓库有多名管理员,每个管理员只能管理一个仓库。

2. 概念结构设计

画出钢材仓库管理系统的 E-R 图,如图 3-6 所示。

3. 逻辑结构设计

根据前面的转换原则,钢材仓库管理系统的关系模式设计如下:

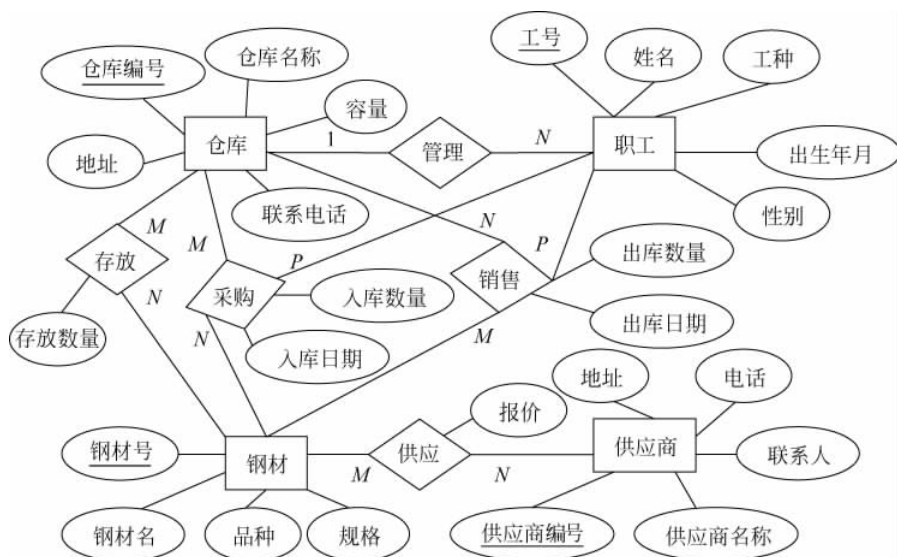


图 3-6 钢材仓库管理系统的 E-R 图

职工(工号,姓名,性别,出生年月,工种,仓库编号)

仓库(仓库编号,仓库名称,地址,联系电话,容量)

钢材(钢材号,钢材名,品种,规格)

供应商(供应商编号,供应商名称,地址,电话,联系人)

存放(仓库编号,钢材号,存放数量)

供应(供应商编号,钢材号,报价)

销售(出库单号,钢材号,仓库编号,销售员工号,出库数量,出库日期)

采购(入库单号,钢材号,仓库编号,采购员工号,入库数量,入库日期)

其中,职工关系的工种属性如果取值为仓库管理员,则仓库编号属性的值为某个仓库的编号,否则为空值。

表结构设计如表 3-6~表 3-13 所示。

表 3-6 职工表的表结构

属性名	类型	宽度	主键	取值范围
工号	字符型	5	是	
姓名	字符型	10		
性别	字符型	2		(男,女)
出生年月	日期型	20		1956-1-1~2050-1-1
工种	字符型	10		(销售员,采购员,仓库管理员)
仓库编号	字符型	4		参考仓库表主键或空值

表 3-7 仓库表的表结构

属性名	类型	宽度	主键	取值范围
仓库编号	字符型	4	是	
仓库名称	字符型	20		
地址	字符型	20		
联系电话	字符型	10		数字字符
容量	整型			$\geq$ 总存放数量

表 3-8 钢材表的表结构

属 性 名	类 型	宽 度	主 键
钢材号	字符型	4	是
钢材名	字符型	20	
品种	字符型	10	
规格	字符型	10	

表 3-9 供应商表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
供应商编号	字符型	5	是	
供应商名称	字符型	30		
地址	字符型	40		
电话	字符型	10		数字字符
联系人	字符型	10		

表 3-10 存放表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
仓库编号	字符型	4	是	参考仓库表主键
钢材号	字符型	4		参考钢材表主键
存放数量	整型			

表 3-11 供应表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
供应商编号	字符型	5	是	参考供应商表主键
钢材号	字符型	4		参考钢材表主键
报价	实型			0.00~9999.99

表 3-12 销售表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
出库单号	字符型	10	是	
仓库编号	字符型	4		参考仓库表主键
销售员工号	字符型	5		参考员工表主键
钢材号	字符型	4		参考钢材表主键
出库数量	整型			≤ 存放数量
出库日期	日期型			2000-1-1~2030-1-1

表 3-13 采购表的表结构

属 性 名	类 型	宽 度	主 键	取 值 范 围
入库单号	字符型	10	是	
仓库编号	字符型	4		参考仓库表主键
采购员工号	字符型	5		参考员工表主键
钢材号	字符型	4		参考钢材表主键
入库数量	整型	2		
入库日期	日期型	8		2000-1-1~2030-1-1

4. 物理结构设计

根据查询需求设计每一个关系的索引文件。

习 题 3

1. 简述数据库设计过程。
2. 简述数据库设计过程的各个阶段上的设计任务。
3. 简述数据库设计的特点。
4. 简述数据库概念结构设计的重要性和设计步骤。
5. 什么是数据库的逻辑结构设计？试述其设计步骤。
6. 简述把 E-R 图转换为关系模型的转换规则。
7. 将图 3-7 所示的学生信息数据库系统的 E-R 图转换为关系模型。

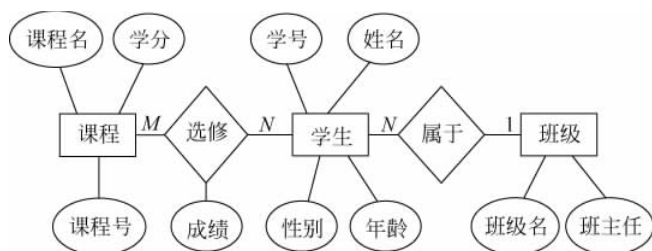


图 3-7 习题 7 图

8. 在一个设备销售管理系统中,设备实体包含设备编号、设备名称、型号规格、数量属性;部门实体包含部门编号、部门名称、部门经理、电话属性;客户实体包含客户编号、客户名称、地址、电话属性。其中,任何设备都可以销售给多个客户,每个客户购买任何一种设备都要登记购买数量;一个部门可以管理多种设备,一种设备仅由一个部门来管理。

根据以上情况完成以下设计:

- (1) 设计系统的 E-R 图;
- (2) 将 E-R 图转换为关系模式,标出每个关系模式的主码。

9. 某电子商务网站要求可随时查询库存中现有物品的名称、数量和单价,所有物品均由物品编号唯一标识;可随时查询顾客订货情况,包括顾客号、顾客名、所订物品编号、订购数量、联系方式、交货地点,所有顾客编号不重复;当需要时,可通过数据库中保存的供应商名称、电话、邮编与地址信息向相应供应商订货,一个编号的货物只由一个供应商供货。

根据以上要求完成以下任务:

- (1) 根据语义设计出 E-R 模型;
- (2) 将该 E-R 模型转换为一组等价的关系模式,并标出各关系模式的主码。

10. 根据转换规则,将第 1 章习题的第 13 题中的 E-R 模型转换成关系模型,要求标明每个关系模式的主键和外键(如果存在)。

11. 根据转换规则,将第 1 章习题的第 14 题中的 E-R 模型转换成关系模型,要求标明每个关系模式的主键和外键。