

数据库设计和建模

数据库设计是进行数据库应用系统开发的重要环节,它是一种非常专业化的工作,如同服装设计、建筑设计、舞美设计等在各自领域有专门的意义一样。



3-1

3.1 数据库生命周期

数据库的应用有着严格的阶段划分,也称为生命周期。一个完整的数据库生命周期中涵盖了数据库设计、数据库实施(测试、生成、部署、维护、监视和备份)和数据库应用系统开发等一系列活动。本节介绍与数据库设计的生命周期相关的内容。

1. 什么是数据库设计

数据库设计是根据用户需求以及具体的数据对象的特征,选择相应的数据库管理系统,来设计某一具体的数据库应用系统的数据库模型。它是设计适宜用户使用的数据库组织结构和系统构造的过程。主要任务是通过现实系统中的数据进行抽象,得到符合现实系统需求,又能被数据库管理系统(DBMS)支持的数据模型。

数据库设计是“三分理论,七分设计”,设计者必须灵活地运用数据库理论,根据实际情况决定创建什么样的数据库,以及数据库中包含什么信息、数据表之间如何联系等。设计者必须在深刻地体会数据库原理本质的基础上,善于从管理对象抽象出有用的信息,并建立数学模型,这种能力不只是靠专门的知识 and 技艺本身,更依赖于对知识的综合利用。

数据库设计决定了数据库应用系统的底层设计的好坏,制约着整个数据库应用系统的成败,在实际工作中,人们常常因为数据库设计的不够完善导致系统需求的改变,严重的情况下甚至会导致应用系统开发无法进行,以至于要重新进行数据库设计。

2. 数据库设计方法与步骤

按照规范化设计方法,可将数据库设计归纳为如下 6 个阶段:

(1) 需求分析;

- (2) 概念结构设计；
- (3) 逻辑结构设计；
- (4) 物理结构设计；
- (5) 数据库实施；
- (6) 数据库运行和维护。

这6个阶段构成了一个完整的数据库设计生命周期。在数据库设计生命周期中,各阶段的任务目标和设计工作过程不尽相同。要设计出一个完善而高效的数据库模型,须认真做好每一个阶段的工作。以下分别介绍前4个阶段。

3.2 需求分析



3-2

需求分析阶段是数据库设计的基础,是数据库设计的最初阶段。这一阶段要收集大量的支持系统目标实现的各类基础数据、用户需求信息和信息处理需求,并加以分析归类和初步规划,确定设计思路。需求分析做得好与坏,决定了后续设计的质量和速度,制约着数据库应用系统设计的全过程。需求分析阶段是数据库设计的第一步,也是其他设计阶段的依据,是最困难、最耗费时间的阶段。

3.2.1 需求分析阶段的目标及任务

需求分析阶段要通过详细调查,深入了解需要解决的问题,了解用户对象所给数据的性质及其存在状态和使用情况,了解数据的流程、流向、流量等,并要仔细地分析用户在数据处理上的目标任务,以及在数据格式、数据处理、数据库安全性、可靠性以及数据的完整性方面的需求。

1. 需求分析阶段的目标

需求分析阶段的目标是对数据库应用系统所要处理的对象进行全面了解,大量收集支持系统目标实现的各类基础数据,调查用户对数据库信息的需求、对基础数据进行加工处理的需求、对数据库安全性和完整性的要求,按一定规范要求写出设计者和用户都能理解的需求分析说明书。

需求分析说明书通常包括:

- 分析用户活动过程与状态,产生业务流程图;
- 确定系统范围,产生系统范围图;
- 分析用户活动涉及的数据集。

2. 需求分析阶段的工作任务

需求分析阶段的工作任务是利用数据库设计理论和方法,对现实世界服务对象的现行系统进行详细调查,收集支持系统目标的基础数据及数据处理需求,撰写需求分析报告。

其具体工作任务如下:

(1) 调查数据库应用系统所涉及的用户各部门的组成情况、各部门职责、各部门业务及其流程,确定系统功能范围,明确哪些业务活动的工作可由计算机完成,哪些由人工来做。

(2) 了解用户对数据库应用系统的各种要求,包括信息要求、处理要求、安全性和完整性要求,如各个部门输入和使用什么数据,如何加工处理这些数据,处理后的数据的输出内容、格式及发布的对象等。

(3) 深入分析用户的各种需求,并用数据流图描述整个系统的数据流向以及对数据进行处理的过程,描述数据与处理之间的联系,也可用数据字典描述数据流图中涉及的各数据项、数据结构、数据流、数据存储和处理过程。

3.2.2 需求分析阶段的工作过程

需求分析阶段的工作过程中,数据库设计者要对用户进行需求调查。在进行调查时,最好深入用户的工作场所进行详细了解,与用户交流,明确用户需求并确定系统服务边界,最终形成需求分析报告。

需求分析阶段工作过程,如图 3-1 所示。

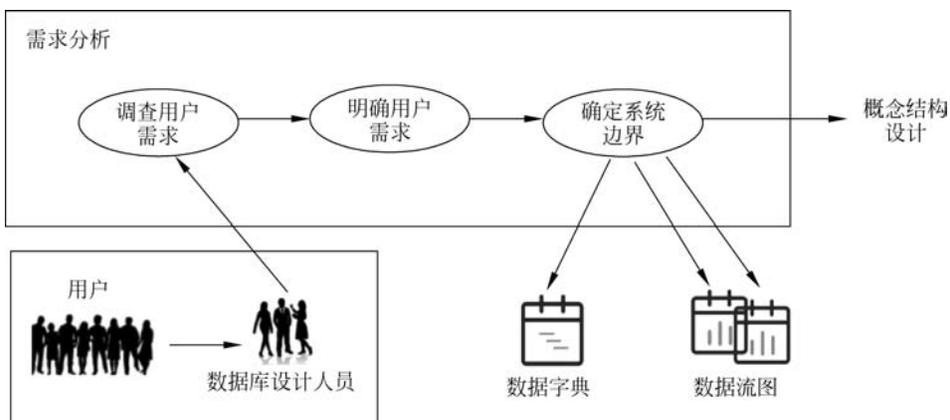


图 3-1 需求分析阶段工作过程

例 3-1 进行“新华大学学生信息管理系统”的需求分析。

“新华大学学生信息管理系统”主要用于教务人员对学校学生成绩信息的数字化

管理。以学生信息、教师信息、课程信息和学习行为的数据为例,简述系统的业务需求和系统功能如下:

(1) 系统业务需求如图 3-2 所示。

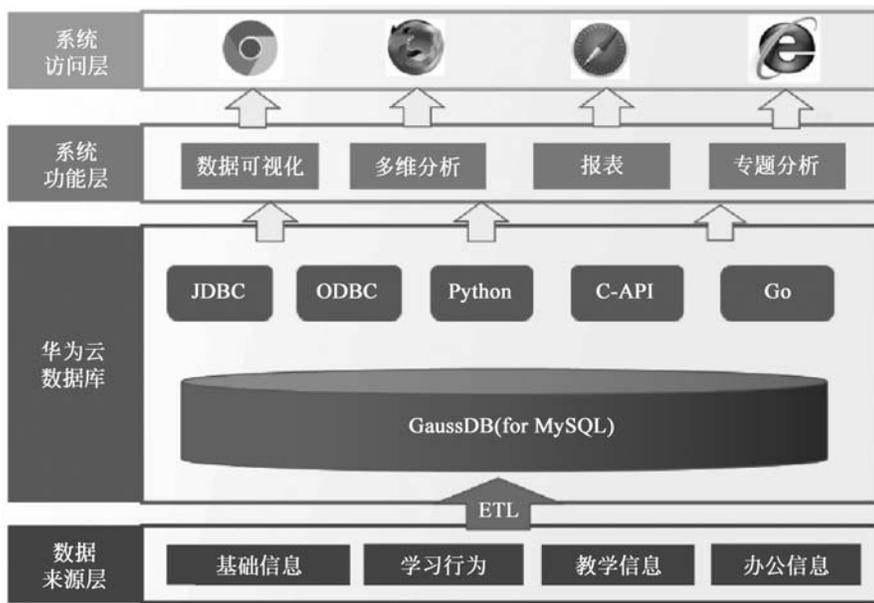


图 3-2 系统业务需求

(2) “新华大学学生信息管理系统”功能框图如图 3-3 所示。

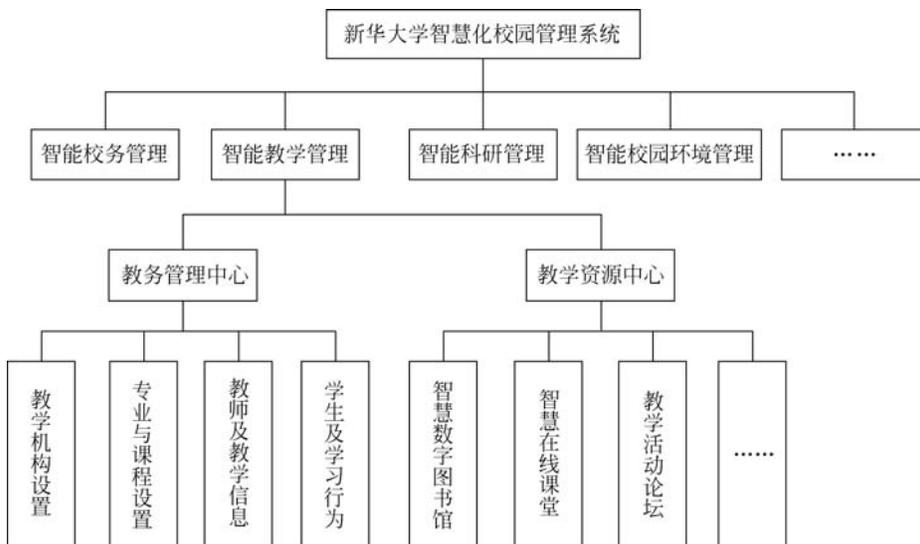


图 3-3 系统功能框图

(3) 支撑业务功能实现的数据集有如下内容:

① 基础信息管理: 包括学校、系、班级信息管理、学生信息、教师信息和课程数据管理等功能模块数据支撑。

② 教务信息管理: 包括课程教学管理、学生学习行为管理,以及数据分析等功能模块数据支撑。



3-3

3.3 概念结构设计

数据库概念结构设计阶段主要设计数据库的整体概念结构,也就是把需求分析结果抽象为反映用户需求信息和信息处理需求的概念模型。

概念模型独立于特定的数据库管理系统,也独立于数据库逻辑模型,还独立于计算机和存储介质上的数据库物理模型。

3.3.1 概念结构设计的目标及任务

1. 概念结构设计目标

概念结构设计目标是在需求分析的基础上,进行分析、归纳、抽象,形成一个符合数据对象实际、用户需求及工作要求的、独立于具体 DBMS 和计算机硬件结构的整体概念结构,即提出概念模型。

2. 概念结构设计任务

概念结构设计的具体工作任务流程如下:

- ① 进行数据抽象;
- ② 设计局部概念模式,得到局部 E-R 图;
- ③ 将局部概念模式综合成全局概念模式,得到全局 E-R 图;
- ④ 评价全局概念模式与优化,得到优化的全局 E-R 图。

3.3.2 概念结构设计的一般策略和方法

概念结构设计是有策略和方法可循的,以下介绍为一般性的总结。

1. 策略

(1) 自顶向下: 先定义全局 E-R 模式框架,然后逐步细化,即先从抽象级别高且普

遍性强的实体集开始设计,然后逐步进行细化、具体化与特殊化处理。

(2) 自底向上:首先定义各局部应用的概念结构,然后将它们集成起来,得到全局概念结构。先从具体的实体开始,然后逐步进行抽象化,经过普遍化与一般化,最后形成一个较高层次的抽象实体集。

(3) 由内向外:首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,直至生成总体概念结构。即先从最基本与最明显的实体集着手,逐步扩展至非基本、不明显的其他实体集。

(4) 混合策略:将“自顶向下”和“自底向上”相结合,用“自顶向下”策略设计一个全局概念结构的框架,同时以它为骨架集成用“自底向上”策略设计各局部概念结构。

2. 方法

(1) 集中式设计法:根据用户需求由一个统一的机构或人员一次性设计出数据库的全局 E-R 模式。其特点是容易保证 E-R 模式的统一性与一致性,但它仅适用于小型或并不复杂的数据库设计问题,而在设计大型的或语义关联复杂的数据库时并不适用。

(2) 分散-集成设计法:设计过程分解成两步,首先,根据某种原则将一个企业或部门的用户需求分解成若干部分,并对每个部分设计局部 E-R 模式;然后将各个局部 E-R 模式进行集成,并消除集成过程中可能会出现冲突,最终形成一个全局 E-R 模式。其特点是设计过程比较复杂,但能较好地反映用户需求,对于解决大型和复杂的数据库设计问题比较有效。

3.3.3 概念结构设计阶段工作过程

概念结构设计阶段工作过程,是先设计局部概念结构,再整合全局概念结构。

1. 局部概念结构设计

(1) 确定概念结构的范围:将用户需求划分成若干个部分,其划分方法有两种:
①根据企业的组织机构对其进行自然划分,并逐一设计其概念结构。
②根据数据库提供的服务种类进行划分,使得每一种服务所使用的数据明显地不同于其他种类,并将每一类服务设计成局部概念结构。

(2) 定义实体型:每一局部的概念结构包括哪些实体型,要从选定的局部范围内的用户需求出发进行选定,即逐一确定每一个实体型的属性及其属性名和主码。

设计的内容包括:

- ① 区分实体与属性。
- ② 给实体集与属性命名。其原则是清晰明了、便于记忆,并尽可能采用用户熟悉

的名字,减少冲突,方便使用。

③ 确定实体标识,即确定实体集的主码。在列出实体集的所有候选码的基础上,选择一个作为主码。

④ 非空值原则:保证主码中的属性不出现空值。

(3) 定义联系:判断实体集之间是否存在联系,并定义实体集之间联系的类型。

① 确定实体集之间是否存在联系,而且同时确定联系类型。

② 定义联系的方法。

③ 为实体集之间的联系命名:联系的命名应反映联系的语义性质,通常采用动词命名。

④ 确定每个联系的存在属性,并为其命名。

2. 合并局部概念结构设计

合并局部 E-R 模式为全局 E-R 模式的过程包括区分公共实体型、合并局部概念结构设计和消除冲突 3 步。

① 区分公共实体型:一般根据实体型名称和主码来认定公共实体型。

② 合并局部概念结构设计:首先,将具有公共实体型的局部概念结构设计进行合并,然后加入独立的局部概念结构设计,这样即可获得全局概念结构设计。

③ 消除冲突:消除合并过程中局部概念结构设计之间出现的不一致描述。

两个局部 E-R 模式之间可能出现的冲突类型如下:

① 命名冲突:主要指同名异义和异名同义两种冲突,包括属性名、实体型名、联系名之间的冲突。同名异义,即不同意义的对象具有相同的名字(编号);异名同义,即同一意义的对象具有不同的名字。

② 结构冲突:同一对象在不同的局部概念结构设计中的抽象不一致,同一实体在不同的局部 E-R 模式中的属性组成不同。

3. 优化全局概念结构

全局 E-R 模式的优化标准:能全面、准确地反映用户需求,且具有实体型的个数尽可能少;实体型所含属性个数尽可能少;实体型之间联系无冗余等。

(1) 全局概念结构的优化方法:首先将实体型进行合并,将两个有联系的实体型合并为一个实体型;然后消除属性的冗余,即消除合并为全局 E-R 模式后产生的冗余属性;最后消除联系的冗余,也就是消除全局模式中存在的冗余联系。

(2) 全局概念结构的优化原则:在存储空间、访问效率和维护代价之间进行权衡,对实体型进行恰当的合并,适当消去部分冗余属性和冗余联系。

概念结构设计阶段工作过程,如图 3-4 所示。

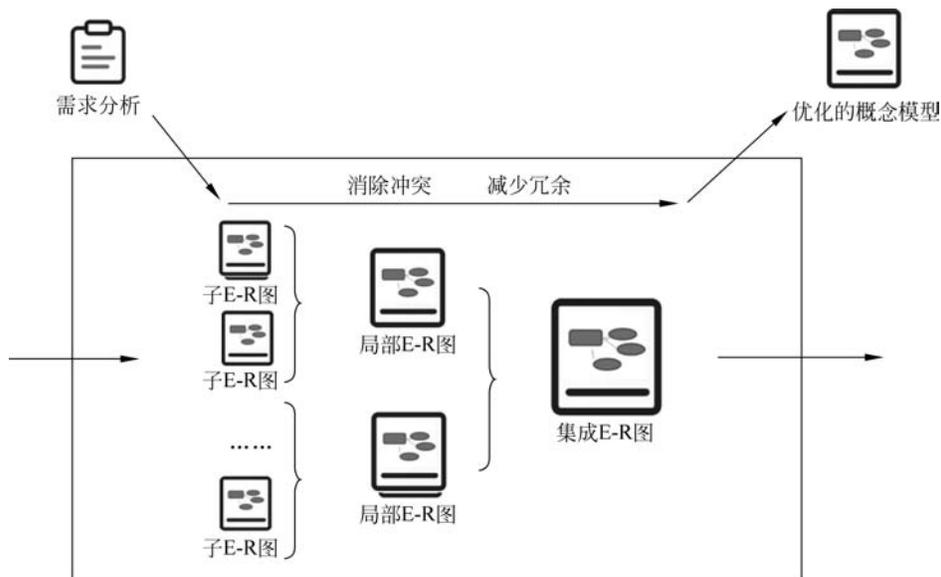


图 3-4 概念结构设计阶段工作过程

例 3-2 “新华大学学生信息管理系统”的全局概念结构设计。

根据需求设计的“新华大学学生信息管理系统”全局概念结构,如图 3-5 所示。

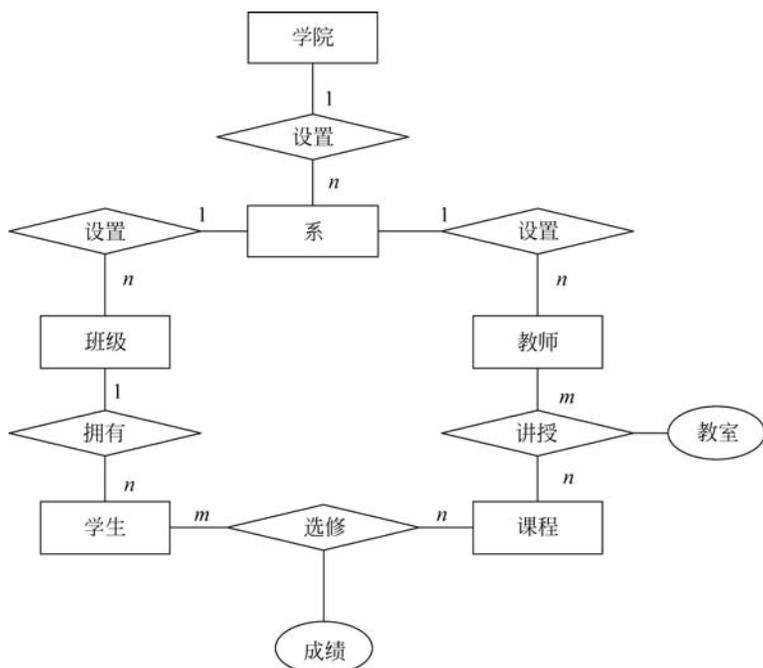


图 3-5 全局概念结构



3-4

3.4 逻辑结构设计

数据库逻辑结构设计是在概念模型的基础上进行的,是把概念模型转换成某个数据库管理系统支持的数据模型。设计者需要详细了解前一个阶段数据库设计的全过程,重点是概念设计中的 E-R 模型的设计方法,研究确定逻辑结构设计中,E-R 模型向关系模型转换的方法,还要考虑物理设计中索引的建立。

3.4.1 逻辑结构设计的目标及任务

1. 逻辑结构设计目标

逻辑结构设计目标是在概念结构设计的基础上,在一定原则的指导下,将概念结构转换为与某具体 DBMS 支持的数据模型相符合的、经过优化的逻辑结构。

2. 逻辑结构设计工作任务

逻辑结构设计的具体工作任务包括:

- ① 选定 DBMS;
- ② 将概念模型转换 DBMS 支持的数据模型(全局关系模式);
- ③ 利用规范化原则优化(良好全局关系模式);
- ④ 实现数据模型完整性(关系的完整性相关约束)。

3.4.2 概念结构转换成逻辑结构的方法

将概念模型转换成逻辑结构时通常采用“二步式”,一是按转换规则直接转换,二是进行关系模式的优化。

1. 概念模型转换成逻辑结构的原则

(1) 实体型的转换:对于概念结构中的每个实体型,设计一个关系模式与之对应,使该关系模式包含实体型的所有属性。通常用下划线来表示关系模式的主码所包含的属性。

(2) 联系的转换:联系的转换方法是由联系的类型决定的,具体方法如下:

① 1:1 联系的转换:先将两个实体型分别转换为两个对应的关系模式,再将联系的属性和其中一个实体型对应关系模式的主码属性加入到另一个关系模式中。

② $1:n$ 联系的转换：先将两个实体型分别转换为两个对应的关系模式，再将联系的属性和 1 端对应关系模式的主码属性加入到 n 端对应的关系模式中。

③ $m:n$ 联系的转换：先将两个实体型分别转换为两个对应的关系模式，再将联系转换为一个对应的关系模式，其属性由联系的属性和前面两个关系模式的主码属性构成。

2. 关系模式的优化

优化关系模式的方法如下：

(1) 确定数据依赖：按需求分析阶段所得到的语义，分别写出每个关系模式内部各属性之间的数据依赖以及不同关系模式属性之间的数据依赖。

(2) 消除冗余的联系：对于各个关系模式之间的数据依赖进行极小化处理，消除冗余的联系。

(3) 确定所属范式：根据数据依赖的理论对关系模式逐一进行分析，确定各关系模式分别属于第几范式。注意，并不是规范化程度越高的关系就越好，一般说来，第三范式就足够了。

(4) 确定数据处理是否合适：根据需求分析阶段得到数据处理的要求，分析关系模式是否合适。若不合适，对其进行合并或分解。

3.4.3 逻辑结构设计阶段工作过程

逻辑结构设计阶段工作过程较为简单，它的设计结果完全依赖于“概念结构”。首先要选定 DBMS，然后将概念结构转换为 DBMS 支持的数据模型，最后利用规范化原则优化数据模型。

逻辑结构设计阶段工作过程，如图 3-6 所示。

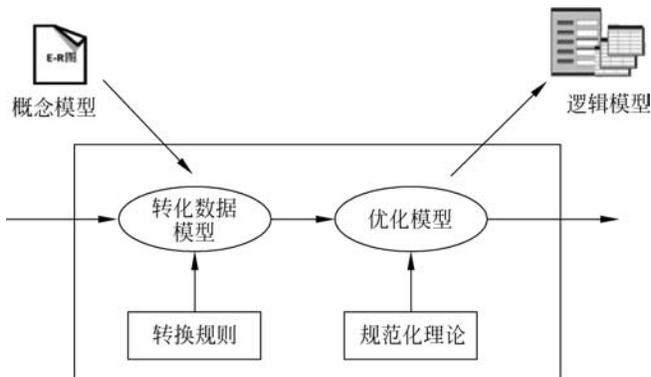


图 3-6 逻辑结构设计阶段工作过程

例 3-3 “新华大学学生信息管理系统”应用系统的逻辑结构设计。

根据“新华大学学生信息管理系统”全局概念结构,我们将“新华大学学生信息管理系统”逻辑结构设计如下:

学院(学院编号,学院名称,院长,电话,地址)

系(系编号,系名称,系主任,电话,班级个数,学院编号)

班级(班级编号,班级名称,班级人数,班长姓名,专业名称,系编号)

学生(学号,姓名,性别,出生年月,籍贯,班级编号)

教师(教师编号,姓名,性别,职称,系编号)

课程(课程编号,课程名称,学时,学分,学期)

学生成绩(学号,课程编号,成绩)

教师授课(教师编号,课程编号,教室编号)



3.5 物理结构设计

3-5

数据库物理结构设计阶段针对一个给定的数据库逻辑模型,选择最适合的应用环境。换句话说,就是能够在应用环境中的物理设备上,由全局逻辑模型产生一个能在特定 DBMS 上实现的关系数据库模式。

3.5.1 物理结构设计的目标及任务

1. 物理结构设计阶段目标

物理结构设计阶段目标是为逻辑数据结构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等。

2. 物理结构设计阶段工作任务

物理结构设计阶段的具体工作任务包括以下 6 部分。

- (1) 存储记录结构设计(表的结构);
- (2) 确定数据存放位置;
- (3) 存取方法的设计(触发器与存储过程);
- (4) 完整性和安全性考虑;
- (5) 对物理结构进行评价;
- (6) 程序设计(前台代码的设计)。

3.5.2 物理结构设计时的注意事项

- (1) 确定数据的存储结构：设计关系、索引等数据库文件的物理存储结构，需注意存取时间、空间效率和维护代价间的平衡；
- (2) 选择合适的存取路径：确定哪些关系模式建立索引，索引关键字是什么等；
- (3) 确定数据的存放位置：确定数据存放在一个磁盘上还是多个磁盘上；
- (4) 确定存取分布：许多 DBMS 都提供了一些存储分配参数供设计者使用(如缓冲区的大小和个数、块的长度、块因子的大小等)。

3.5.3 物理结构设计阶段工作过程

物理结构设计阶段首先要设计存储记录的表结构，然后确定数据存放位置和存取方法，同时也要设计数据的完整性和安全性。

物理结构设计阶段工作过程，如图 3-7 所示。

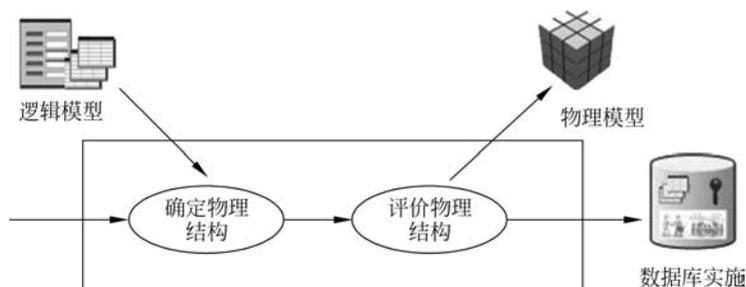


图 3-7 物理结构设计阶段工作过程

例 3-4 “新华大学学生信息管理系统”应用系统的物理结构设计。

“新华大学学生信息管理系统”数据中心的表结构设计如表 3-1~表 3-8 所示。

表 3-1 School 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
School_id	学院编号	char	1	有(无重复)	主键
School_name	学院名称	char	10	—	—
School_dean	院长姓名	char	6	—	—
School_tel	电话	char	13	—	—
School_addr	地址	char	10	—	—

表 3-2 Department 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
Department_id	系编号	char	4	有(无重复)	主键
Department_name	系名称	char	14	—	—
Department_dean	系主任	char	6	—	—
Teacher_num	教师人数	smallint	默认值	—	—
Class_num	班级个数	smallint	默认值	—	—
School_id	学院编号	char	1	—	外键

表 3-3 Class 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
Class_id	班级编号	char	8	有(无重复)	主键
Class_name	班级名称	char	4	—	—
Student_num	班级人数	smallint	默认值	—	—
Monitor	班长姓名	char	6	—	—
Major	专业	char	10	—	—
Department_id	系编号	char	4	—	外键

表 3-4 Student 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
Student_id	学号	char	6	有(无重复)	主键
Student_name	姓名	char	6	—	—
Gender	性别	char	2	—	—
Birth	出生年月	datetime	默认值	—	—
Birthplace	籍贯	char	50	—	—
Class_id	班级编号	char	8	—	外键

表 3-5 Teacher 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
Teacher_id	教师编号	char	7	有(无重复)	主键
Teacher_name	姓名	char	6	—	—
Gender	性别	char	2	—	—
Title	职称	char	8	—	—
Department_id	系编号	char	4	—	外键

表 3-6 Course 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
Course_id	课程编号	char	5	有(无重复)	主键
Course_name	课程名称	char	12	—	外键
Period	学时	smallint	默认值	—	—
Credit	学分	smallint	默认值	—	—
Term	学期	smallint	1	—	—

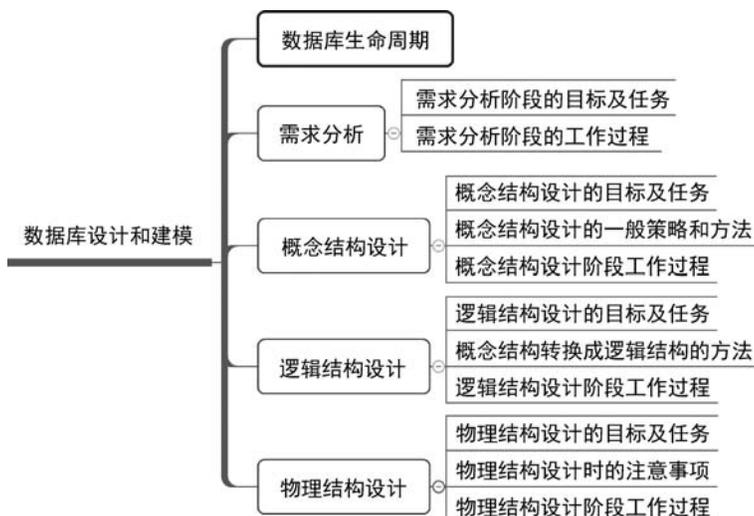
表 3-7 Score 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
Student_id	学号	char	6	有(无重复)	联合主键
Course_id	课程编号	char	5	有(无重复)	联合主键
Score	成绩	smallint	默认值	—	—

表 3-8 Assignment 表结构

字段名	字段别名	字段类型	字段长度	索引	备注
Teacher_id	教师编号	char	7	有(无重复)	联合主键
Course_id	课程编号	char	5	有(无重复)	联合主键
Classroom_id	教室编号	char	5	—	—

知识点树



思考题

- (1) 简述数据库设计的步骤。
- (2) 需求分析阶段主要工作是什么？
- (3) 简述数据库概念结构设计方法。
- (4) 简述数据库逻辑结构设计方法。
- (5) 简述数据库物理结构设计方法。
- (6) 数据库实施阶段的主要工作是什么？
- (7) 数据库对象有哪些？
- (8) 解释概念结构、逻辑结构和物理结构三者之间的关系。