项 目 1

数据库设计

目前数据库系统的应用已与人们的生活密不可分，其中数据库设计是数据库系统开发中重要的环节，其设计的好坏直接影响后面系统的具体使用情况。数据库设计主要分为6个阶段，分别为需求分析阶段、概念结构设计阶段、逻辑结构设计阶段、物理结构设计阶段、数据库实施以及数据库运行与维护。

项目主要任务：

- 任务1　设计概念结构；
- 任务2　设计逻辑结构；
- 任务3　设计物理结构。

能力培养目标：

- 具备数据库分析的能力；
- 具备数据库概念结构设计的能力；
- 具备数据库逻辑结构设计的能力；
- 具备数据库E-R图设计的能力；
- 具备数据库物理结构设计的能力。

素质培养目标：

- 培养学生解决实际问题的独立思考能力；
- 培养学生沟通交流的能力；
- 培养学生的创新思维；
- 培养学生数据库设计人员的职业素养。

1.1 项目导入与需求分析

1. 项目导入

你正在大学就读，学校现在需要你设计一个包含学生所有信息的数据库系统：学生信息管理系统。学生信息管理系统是针对学校大量业务处理工作而开发的管理软件，主要用于学校学生信息管理，总体任务是实现学生信息管理的系统化、科学化、规范化和自动化，其主要任务是对学生各种信息进行日常管理，如查询、修改、增加、删除，另外还考虑到学生选课等功能。

2. 需求分析

现在，我们需要设计学生信息管理系统数据库，怎么样设计呢？自己坐在计算机前设想应该有哪些功能？当然不是。我们需要通过与用户协商等方式了解用户需求，确定学生信息管理系统中各个用户对数据的使用要求。

通过需求分析，我们得到了客户所需要的学生信息管理系统的主要功能：实现学生信息的查询，学籍信息的输入、修改和删除；学生课表的查询，学校基本课程信息的输入、修改和删除；学生成绩信息的统计、查询、修改、删除等功能。

1.2 相 关 知 识

通过上面的需求分析，我们已经掌握了客户的主要需求，那么，在正式开展数据库设计工作之前，大家需要了解设计数据库的一些相关知识。

1.2.1 数据库相关概念

1. 数据

现实世界中的数字数据、文本、图像、声音、视屏等信息都可以称为数据，所以概括起来数据(Data)就是描述现实世界中各种具体事物或者抽象概念的符号记录。

2. 数据库

数据库(DataBase，DB)指长期存储在计算机内有组织的、可共享的数据集合，即在计算机系统中按一定的数据模型组织、存储和使用的相关联的数据集合。它不仅包括描述事物的本身还包括相关数据之间的联系。

3. 数据库管理系统

数据库管理系统(DataBase Management System，DBMS)指支持用户对数据库进行定义、创建、维护及控制访问的软件系统，即一个与用户的应用程序和数据库相互作用的软件。常用的数据库管理系统有 MySQL、Oracle、SQL Server 等。

4. 数据库应用系统

数据库应用系统是为了满足用户对于数据库的管理而开发设计的应用软件。用户通过数据库应用系统向数据库管理系统提出请求，从而与数据库交互作用。

1.2.2　概念结构设计基本方法

系统的需求分析反映了用户的需求，我们要将其转化为信息世界的结构，这就是概念结构设计阶段主要要完成的内容。

1. 概念结构的相关概念

(1) 实体(Entity)。实体即客观现实存在并可相互区分的事物，如学生、教师、课程等。

(2) 属性(Attribute)。属性即实体所具有的某一特征，一个实体可由多个属性描述，如学生实体可以由姓名、学号、性别等属性描述。属性包含属性名和属性值，如“张力、20200001，男”则分别为对应姓名、学号、性别三个属性名的属性值。

(3) 码(Key)。码也称为键，即唯一标识实体的属性集。

(4) 域(Domain)。域即属性的取值范围，如学生实体的性别属性的取值范围为{“男”，“女”}。

(5) 实体型(Entity Type)。实体型即实体类型，指用实体名和该实体所具有的属性名的集合抽象化地标识同一类实体，如学生实体的实体名为“学生”，属性名的集合为{姓名、学号、性别}，那么就可以用实体名与属性名的集合来描述其实体型。

(6) 实体集(Entity Set)。实体集即同一类型实体的集合，如计算机学院的学生就是实体集。

(7) 联系(Relationship)。联系即实体与实体之间的联系，主要有三种情况，分别为一对一(1∶1)、一对多(1∶N)和多对多(M∶N)。例如每个教师都有一个工作证，教师和工作证之间是一对一的联系；一个班级有多个学生，班级与学生之间是一对多的联系；一个学生可以选修多门课程，一门课程可以被多个学生选修，学生与课程之间是多对多的联系。

(8) 实体联系模型(Entity-Relationship，E-R)。实体联系模型即 E-R 模型，指从现实世界中抽取出实体与实体之间的联系，主要由实体、联系和属性构成。

2. 概念结构设计的方法

(1) 自顶向下法。自顶向下法的主要思想是先建立一个全局 E-R 模型，即先认定用户最关注的实体与实体间的联系，然后再对该 E-R 模型逐步地进行细化，并且添加属性，得到局部 E-R 模型，最终设计出整体的 E-R 模型。

(2) 自底向上法。自底向上法的主要思想是先用需求分析阶段获得的数据元素作为基本输入，通过分析将它们形成局部 E-R 模型，然后再逐步汇总从而形成全局 E-R 模型。

(3) 逐步扩张法。逐步扩张法的主要思想为先将最核心的 E-R 模型进行定义，然后逐渐外扩，最终生成整体的 E-R 模型。

(4) 混合策略法。混合策略法的主要思想是先将整个系统划分为不同的相对独立的功能，然后针对每个功能设计局部 E-R 模型，最后通过归纳、合并消去冗余和不一致，从而形成全局 E-R 模型。

3. 概念结构设计的步骤

概念结构设计可以分为两个步骤：首先设计局部 E-R 模型；其次集成各局部 E-R 模型形成全局 E-R 模型。

(1) 设计局部 E-R 模型。

① 划分局部 E-R 图的功能。将整个数据库系统的功能进行划分，划分应当遵循独立性

和适度性原则，即划分的局部功能应独立完整，与其他功能应有最少的联系；局部 E-R 图的规模应当适当，一般以 6 个左右实体为宜。

② 确定局部 E-R 图实体。根据需求分析阶段获得的需求分析说明书，确定每个局部功能中的具体实体。

③ 确定局部 E-R 图实体的属性。根据局部 E-R 图中确定的实体确定描述该实体的所有属性。

④ 确定局部 E-R 图各实体间的联系。分析上面步骤确定的局部 E-R 图中确定的实体与属性，确定实体间的具体联系。

(2) 集成局部 E-R 模型。上面得到的局部 E-R 模型只能反映局部功能，还不可以作为逻辑结构设计的依据，那么，我们需要将各个局部 E-R 模型进行集成，得到全局 E-R 模型。

集成局部 E-R 模型有一次集成和逐步集成两种方式，其中一次集成复杂难度较大，逐步集成每次只集成两个局部 E-R 图，集成较为简单。

无论使用哪一种集成方式，都要通过合并和优化两个步骤。合并就是消除每个集成的局部 E-R 模型之间的不一致，形成初步的 E-R 图；优化就是消除不必要的冗余，形成最终的全局 E-R 图。

1.2.3 逻辑结构设计基本方法

1. 逻辑结构的相关概念

(1) 关系模型。关系模型是指用二维表结构表示实体与实体之间的联系的数据模型。简单地说，关系模型就是一张二维表。采用关系模型的数据库就是关系数据库。下面为关系模型中常见的一些术语。

- 关系：一个关系对应通常称为一张表。
- 元组：表中的一行即称为一个元组。
- 属性：表中的一列即称为一个属性，为每一个属性起一个名称就叫作属性名。
- 码：表中的某个属性组，该属性组可以唯一确定一个元组。
- 域：属性的取值范围。
- 分量：元组中的一个属性值。

(2) 关系模式。对关系模型的描述就是关系模式。因此，关系模式要指出这个关系由哪些属性组成，属性的域是什么，以及属性与域之间的映射关系。

2. 逻辑结构设计的步骤

逻辑结构设计主要分为 3 个步骤：首先将 E-R 图转化为关系模式；其次对关系模式进行优化；最后设计用户子模式。

(1) 将 E-R 图转化为关系模式。将 E-R 图转化为关系模式即将实体、属性、联系转化为关系模式。

要遵守以下原则将 E-R 图转化为关系模式。

① 实体的转化。

- 将每个实体类型转换成一个关系模式。
- 实体的属性即为关系模式的属性。
- 实体标识符即为关系模式的键。

② 联系的转化。根据需求分析阶段获得的需求分析说明书，确定每个局部功能中的具体实体。

- 若实体间联系是1∶1，可以在两个实体类型转换成的两个关系模式中任意一个关系模式的属性中加入另一个关系模式的键和联系类型的属性。
- 若实体间联系是1∶N，则在N端实体类型转换成的关系模式中加入1端实体类型的键和联系类型的属性。
- 若实体间联系是M∶N，则将联系类型也转换成关系模式，其属性为两端实体类型的键加上联系类型的属性，而键为两端实体键的组合。

(2) 关系模式优化。数据库逻辑设计的结果不是唯一的。关系数据模型的优化通常以规范化理论为指导，方法有以下几种。

- 确定数据依赖。
- 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。
- 按照数据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、多值依赖等，确定各关系模式分别属于第几范式。
- 根据需求分析阶段得到的处理要求分析对于这样的应用环境这些模式是否合适，确定是否对某些模式进行合并或分解。
- 对关系模式进行必要分解，提高数据操作效率和存储空间利用率。

(3) 设计用户子模式。子模式也叫外模式，也就是用户可以直接访问的数据模式。通过子模式对逻辑模式进行屏蔽，可以适应不同用户对于数据的需求。在同一个系统中，不同的用户可以有不同的子模式。

在设计用户子模式时需要考虑到用户的方便与习惯。例如使用符合用户习惯的别名；对不同的用户设计不同的子模式，以保证系统的安全性；简化用户的使用流程。

1.2.4　物理结构设计基本方法

1. 物理结构的相关概念

数据库的物理结构设计是指数据库在物理设备上的存储结构与存取方法。

2. 物理结构设计的步骤

数据库的物理结构设计可以分为两个步骤：首先确定数据库的物理结构，在关系数据库中主要指存储结构和存取方法；其次是对物理结构进行评价，评价的重点是时间和空间效率。

(1) 确定数据库的物理结构。

① 确定数据库的存储结构。确定数据库的存储结构主要包括确定关系、索引、聚簇、日志、备份等的存储安排和存储结构。

② 确定数据库的存取方法。常用的存取方法有3类：索引方法、聚簇(Cluster)方法和Hash方法。

③ 确定数据库的存放位置。为了提高系统性能，应该根据实际应用将数据库中数据的易变部分与稳定部分、常存取部分、存取频率较低部分分开存放。有多个磁盘的计算机可以采用下面几种存取位置的分配方案。

将表和该表的索引放在不同的磁盘中。在查询时，两个磁盘驱动器并行操作，提高了物理I/O的读、写效率；将比较大的表分别放在两个磁盘上，以加快存取速度，这在多用户环

境下特别有效；将日志文件与数据库的对象放在不同的磁盘上，以改进系统的性能。对于经常存取或存取时间要求高的对象应放在高速存储器上；对于存取频率小或存取时间要求低的对象（如数据库的数据备份和日志文件备份等，只在故障恢复时才使用），如果数据量很大，则可以存放在低速存储设备上。

④ 确定系统配置。DBMS产品一般都提供了一些系统配置变量、存储分配参数，以供设计人员和DBS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。这些初始值并不一定适合每种应用环境，在进行物理设计时，需要重新对这些变量赋值，以改善系统的性能。系统配置变量很多，如同时使用数据库的用户数、同时打开数据库的对象数、内存分配参数、缓冲区分配参数（使用的缓冲区长度、个数）、存储分配参数、物理块的大小、物理块装填因子、时间片大小、数据库的大小、锁的数目等。这些参数值会影响存取时间和存储空间的分配，因此在进行物理设计时，要根据应用环境确定这些参数值，以使系统性能最佳。

（2）评价数据库的物理结构。

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，其结果可以产生多种方案。评价物理数据库的方法完全依赖于所选用的关系数据库管理系统，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。

1.2.5 拓展知识 非关系型数据库

NoSQL泛指非关系型的数据库。随着互联网Web 2.0网站的兴起，传统的关系数据库在处理Web 2.0网站，特别是超大规模和高并发的SNS类型的Web 2.0纯动态网站已经显得力不从心，出现了很多难以克服的问题，而非关系型的数据库则由于其本身的特点得到了非常迅速的发展。NoSQL数据库的产生就是为了解决大规模数据集合多重数据种类带来的挑战，尤其是大数据应用难题。

NoSQL最常见的解释是non-relational，Not Only SQL也被很多人接受。NoSQL仅仅是一个概念，泛指非关系型的数据库，区别于关系数据库，它们不保证关系数据的ACID特性。NoSQL是一项全新的数据库革命性运动，其拥护者们提倡运用非关系型的数据存储，相对于铺天盖地的关系型数据库运用，这一概念无疑是一种全新的思维的注入。

NoSQL有如下优点：易扩展，NoSQL数据库种类繁多，但是一个共同的特点都是去掉关系数据库的关系型特性。数据之间无关系，这样就非常容易扩展。无形之间也在架构的层面上带来了可扩展的能力。大数据量，高性能，NoSQL数据库都具有非常高的读写性能，尤其在大数据量下，同样表现优秀。这得益于它的无关系性，数据库的结构简单。

1.3 任 务 分 解

数据库设计是系统开发中重要的环节，其设计的好坏直接影响后面系统的使用情况。其设计主要分为6个阶段，分别为需求分析阶段、概念结构设计阶段、逻辑结构设计阶段、物理结构设计阶段、数据库实施以及数据库运行与维护。1.2节已对项目进行了需求分析，所

以这里我们主要针对概念结构设计、逻辑结构设计、物理结构设计三个部分进行任务分解，数据库实施以及数据库运行与维护将在后续章节进行介绍。

1.3.1　任务 1　设计概念结构

1. 任务目标

■ 学会设计数据库概念结构设计的方法；

■ 学会设计 E-R 图。

2. 任务描述

每个学生可以选修多门课程，每门课程可以被多名学生选修，学生选修课程会有对应的成绩，每个学生都有学号、姓名、性别、出生年月、班级、所属系、政治面貌、民族、生源地、QQ号码、个人电话属性，每门课程具有课程号、课程名、学分、授课教师、授课学期属性。

3. 课堂任务

设计完成 E-R 图。

1.3.2　任务 2　设计逻辑结构

1. 任务目标

■ 学会设计数据库逻辑结构设计的方法；

■ 学会将 E-R 图转化为关系模式。

2. 任务描述

将任务 1 中设计的 E-R 图转化为关系模式。

3. 课堂任务

设计完成学生信息管理系统关系模式。

1.3.3　任务 3　设计物理结构

1. 任务目标

■ 学会设计数据库物理结构设计的方法；

■ 学会设计物理结构。

2. 任务描述

设计学生信息管理系统数据库的存储结构与存储方法。

3. 课堂任务

掌握物理结构设计的步骤。

1.4　任 务 实 施

1.4.1　实施任务 1　概念结构设计

1. 确定实体、属性以及实体间的联系

通过对任务 1 的分析，从“每个学生可以选修多门课程，每门课程可以被多名学生选

修”这句话中，可以找到两个实体——“学生”和“课程”，两个实体之间的联系就是“选课”，并且是多对多的关系；而“学生选修课程会有对应的成绩”，则可以知道联系“选课”具有属性“成绩”；通过“每个学生都有学号、姓名、性别、出生年月、班级、所属系、政治面貌、民族、生源地、QQ 号码、个人电话属性”这句可以知道学生实体具有学号、姓名、性别、出生年月、班级、所属系、政治面貌、民族、生源地、QQ 号码、个人电话属性；通过“每门课程具有课程号、课程名、学分、授课教师、授课学期属性”这句可知课程实体所具有的属性。

2. 设计实体及其属性图

通过上面的分析，可知系统中所涉及的实体和联系较少，为了让大家更好地理解 E-R 图的设计方法，现将局部 E-R 图的设计分为两个实体的 E-R 图设计。

（1）确定学生实体及属性图。由于学生实体涉及学号、姓名、性别、出生年月、班级、所属系、政治面貌、民族、生源地、QQ 号码、个人电话属性，所以学生实体及属性图如图 1-1 所示。

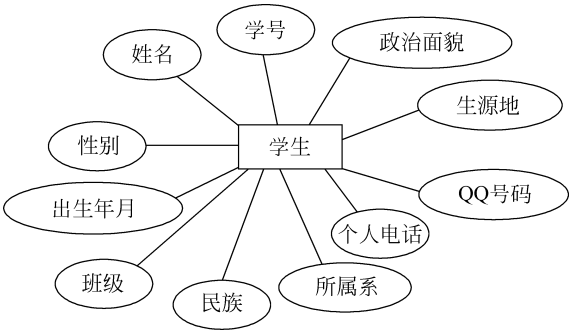

图 1-1　学生实体及属性图

（2）确定课程实体及属性图。由于课程实体具有课程号、课程名、学分、授课教师、授课学期属性，所以课程实体及属性图如图 1-2 所示。

（3）确定选课联系及属性图。由于选课联系具有成绩属性，所以选课联系及属性图如图 1-3 所示。

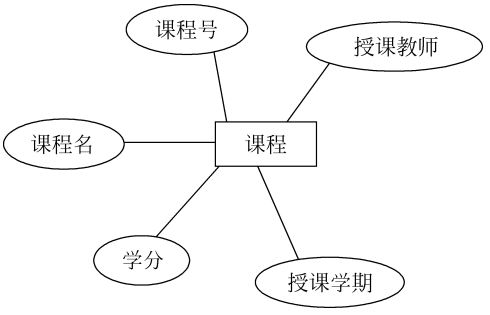

图 1-2　课程实体及属性图

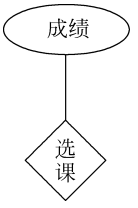

图 1-3　选课联系及属性图

（4）确定选课关系 E-R 图。由于学生和课程之间具有多对多的选课联系，所以 E-R 图可以设计如图 1-4 所示。

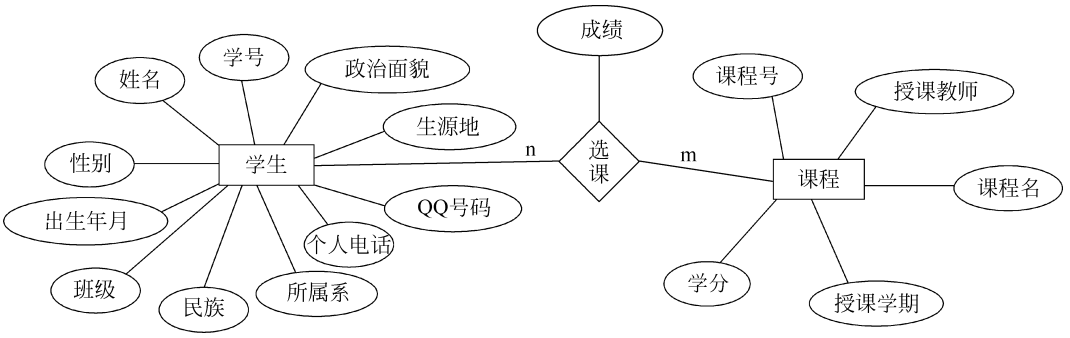

图 1-4　选择关系 E-R 图

1.4.2　实施任务 2　逻辑结构设计

根据实施任务 1 的 E-R 图，可以写出如下关系模式。

- 学生(学号、姓名、性别、出生年月、班级、所属系、政治面貌、民族、生源地、QQ 号码、个人电话)；
- 课程(课程号、课程名、学分、授课教师，授课学期)；
- 选课(成绩)。

1.4.3　实施任务 3　物理结构设计

1. 确定数据库的存储结构

确定数据库的存储结构主要包括确定关系、索引、聚簇、日志、备份等的存储安排和存储结构。

2. 确定数据库的存取方法

将存取方法设为索引方法。

3. 确定数据库的存放位置

将表和该表的索引放在不同的磁盘中。

4. 确定系统配置

配置同时使用数据库的用户数、同时打开数据库的对象数、内存分配参数、缓冲区分配参数(使用的缓冲区长度、个数)、存储分配参数、物理块的大小、物理块装填因子、时间片大小、数据库的大小、锁的数目等。

1.5　项 目 实 训

1.5.1　设计图书馆信息管理系统数据库

按照需求分析、概念结构设计、逻辑结构设计、物理结构设计的步骤设计一个图书馆信息管理系统数据库，要求有详细的设计报告。

1.5.2 创建、管理网上书店信息管理系统数据库

按照需求分析、概念结构设计、逻辑结构设计、物理结构设计的步骤设计一个网上书店信息管理系统，要求有详细的设计报告。

1.6 项 目 小 结

本项目主要是通过对学生信息管理系统进行设计，在对项目进行初步的需求分析之后，将项目分了 3 个任务进行处理。使学生通过学习具备数据库分析的能力、具备数据库概念结构设计的能力、具备数据库逻辑结构设计的能力、具备数据库 E-R 图设计的能力、具备数据库物理结构设计的能力。

习 题

(1) 试述概念结构设计、逻辑结构设计的作用。

(2) 请给出三个实际情况的 E-R 图，要求实体型之间分别具有一对一、一对多、多对多各种不同的联系。

(3) 请写出习题(2)中的三个实际情况的 E-R 图的关系模式。