

技能目标及知识目标

- 了解数据库、数据库管理系统的概念；
- 了解关系数据库的相关概念；
- 了解数据完整性的概念；
- 掌握使用 E-R 模型的方法；
- 掌握数据库逻辑模型的设计方法；
- 掌握使用范式规范数据库设计的方法。

项目导引

数据库的发展经历了人工管理阶段（20 世纪 50 年代）、文件系统阶段（20 世纪 50 年代中期到 60 年代中期）、数据库系统阶段（20 世纪 60 年代后）。自数据库系统阶段至今，人们将软件工程的理论应用于数据库设计，形成了一个完整的数据库设计实施方法，整个过程包括需求分析、概念模型设计、逻辑模型设计、物理模型设计、数据库实现 5 个阶段。

本项目主要学习概念模型设计、逻辑模型设计，并通过两个典型工作任务讲解如何进行数据库概念设计、逻辑设计和数据库规范化。

任务 1 数据库设计

一、任务背景

小 Q 是计算机相关专业大二的学生，趁着假期到企业实习，企业工程师老李交给他一个简单的任务，要求他设计一个能保存员工基本资料和部门信息的数据库。小 Q 平时学习很认真，虽然没有系统学习过数据库，但了解过数据库相关知识，心想这任务也太简单了，马上根据具体要求做了一个数据表，并输入了员工信息，如表 1.1.1 所示。

表 1.1.1 员工信息

员工编号	姓名	性别	手机	部门	部门电话
11001	张三	男	13311112222	市场部	85556666
11002	李四	女	13311112245	开发部	85552222
11003	李明	男	13311112212	开发部	85552222
.....

当小 Q 得意地把数据库拿给工程师看时，工程师摇摇头说：“你这个数据库有很多问题，必须好好设计。”

小 Q 问：“数据库也要设计？我直接建库、建表就可以了。”

老李说：“正如我们建造房子一样，如果是建造一间简易平房，也许不需要设计直接开工建造就行，但如果是一幢大楼或者一个楼盘，那么不预先设计好肯定是不让开工的。我们在系统开发中，如果表比较多，表之间的关系比较复杂，那么不进行数据库设计，后面的程序开发工作就无法进行了。”

随后，工程师老李指出了小 Q 所设计的数据库存在的几个问题。

(1) 数据冗余。目前开发部有 100 多人，你的数据库中，开发部的电话重复保存了 100 多处，浪费存储空间。

(2) 修改异常。如果开发部改变了电话，你需要查找整个数据表，将所有开发部员工的电话都进行修改，否则会出现同一部门但电话不同的情况。

(3) 插入异常。如果要增加一个部门“测试部”，但部门暂时没有员工，你的数据库怎样记录这个部门呢？

(4) 删除异常。如果要删除张三员工的资料，会导致对应部门电话信息丢失。

面对这些问题，小 Q 才发现原来数据库设计还有这么多学问。

要解决小 Q 所遇到的问题，就要进行数据库设计。当数据库比较复杂（如数据量大、表较多、业务关系复杂）时，就更需要数据库设计了。

一个好的数据库设计应做到以下几点。

- (1) 节省数据的存储空间。
- (2) 能够保证数据的完整性，不产生数据的插入异常、修改异常、删除异常等问题。
- (3) 方便进行数据库应用系统的开发。

二、任务需求

某公司计划开发一个简单的 B2C 电子商务平台，销售手机配件，请为其设计数据库。数据库的内容包括以下几类。

- (1) 产品信息，包括产品的分类、编号、名称、价格、库存量等信息。
- (2) 客户信息，包括客户编号、名称、E-mail、生日、地址等信息。
- (3) 订单信息，包括订单时间、明细等信息。
- (4) 登录日志。

注意

作为教学案例，数据库可以尽量简单一些，产品分类只考虑一级分类；登录日志只需要记录用户登录时间和登录 IP 地址。

三、任务分析

针对任务需求，需要进行详细的数据库设计，才能保证数据库符合电子商务平台的需求。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。

数据库设计步骤如下。

(1) 需求分析。需求分析是数据库设计的第一个阶段，是概念设计的基础，该阶段主要任务是对现实世界要处理的对象（组织、部门、企业等）进行详细调查，明确新系统的功能，收集支持系统目标的基础数据及其处理。

(2) 概念结构设计。这是整个数据库设计的关键，概念结构独立于数据逻辑结构，独立于 DBMS，即与具体使用哪一种 DBMS（SQL Server、Oracle、Access 等）无关。概念结构设计是对现实世界的一种抽象，目前广泛使用的方法是 E-R 法。

(3) 逻辑结构设计。该阶段的主要任务就是把概念结构转换为与选用的 DBMS 支持的数据模型（网状、关系等模型）相符合的过程。如果概念设计采用 E-R 法，则向关系模型转换相当容易。

(4) 数据库物理设计。该阶段用于为逻辑数据模型选取一个合适的应用环境。

四、知识要点

1. 数据库和关系数据库

数据是描述事物的符号记录，数据的种类有数字、文字、图形、图像、声音等。

数据库（Database，DB）是一个长期存储在计算机内的、有组织的、可共享的、统一管理的数据集合。数据库中的数据是按照一定的数据模型组织、描述和存储的，有较少的冗余度、较高的数据独立性和易扩展性。

数据库管理系统（Database Management System，DBMS）是使用和管理数据库的系统软件，负责对数据进行统一的管理和控制。

数据库系统（Database Systems，DBS）是由数据库、数据库管理系统及其相关应用软件、支撑环境所组成的系统。DBMS 是数据库系统的基础和核心。

最常用的数据模型有层次模型、网状模型和关系模型，目前主流是关系模型。

关系模型就是用二维表格结构来表示实体及实体之间联系的模型。

1) 关系模型的基本概念

关系（Relation）：一个关系对应一张二维表，每个关系有一个关系名。在 SQL Server 中，一个关系就是一个表对象。

元组（Tuple）：二维表中水平方向的一行，有时也称一条记录。

属性（Attribute）：表格中的一列，相当于记录中的一个字段。

关键字（Key）：可唯一标识元组的属性或属性集，也称为关系键或主键。

域（Domain）：属性的取值范围，如性别的域是（男，女）。

分量：每一行对应的列的属性值，即元组中的一个属性值。

关系模式：对关系的描述。一个关系模式对应一个关系结构，一般表示为关系名（属性 1，属性 2，……，属性 n）。

2) 关系模型的性质

① 关系中不允许出现相同的元组。因为数学的集合中没有相同的元素，而关系是元组的集合，所以作为集合元素的元组应该是唯一的。

② 关系中元组的顺序（即行序）是无关紧要的，在一个关系中 can 任意交换两行的

次序。因为集合中的元素是无序的，所以作为集合元素的元组也是无序的。根据关系的这个性质，可以改变元组的顺序使其具有某种排序，然后按照顺序查询数据，以提高查询速度。

③ 关系中属性的顺序是无关紧要的，即列的顺序可以任意交换。交换时，应连同属性名一起交换，否则将得到不同的关系。

④ 同一属性名下的各个属性值必须来自同一个域，是同一类型的数据。

⑤ 关系中各个属性必须有不同的名称，不同的属性可来自同一个域，即它们的分量可以取自同一个域。

⑥ 关系中每一分量必须是不可分的数据项，或者说所有属性值都是原子的，是一个确定的值，而不是值的集合。

3) 关系数据库完整性

① 实体完整性 (Entity Integrity)。实体完整性是指主关系键的值不能为空或部分为空。

② 参照完整性 (Referential Integrity)。如果关系 R2 的外部关系键 X 与关系 R1 的主关系键相符，则 X 的每个值或者等于 R1 中主关系键的某一个值，或者取空值。

③ 域完整性。域完整性是针对某一具体关系数据库的约束条件。它反映了某一具体应用所涉及的数据必须满足的语义要求。

2. 概念设计

数据库概念设计主要应用实体-联系图 (Entity-Relation Diagram, E-R 图) 来完成。

实体-联系图用来建立数据模型，在数据库系统概论中属于概念设计阶段，形成一个独立于机器，独立于 DBMS 的 E-R 图模型。通常将它简称为 E-R 图，相应的，可把用 E-R 图描绘的数据模型称为 E-R 模型。E-R 图提供了表示实体 (即数据对象)、属性和联系的方法，用来描述现实世界的概念模型。

E-R 图是由美籍华人陈平山于 1976 年提出来的。E 表示实体，A 表示属性，R 表示实体和实体之间的关系。涉及的主要概念如下。

(1) 实体：客观存在并可互相区分的事物。实体可以是人，可以是物，也可以指某些概念。例如，一个职工，一个部门，一门课等。

(2) 属性：实体所具有的某一特性。一个实体可以由若干个属性来刻画。例如，学生可以由学号、姓名、年龄、性别、系、联系电话等属性组成。

(3) 关键字：唯一标识实体的最小属性集。

(4) 联系：现实世界的事物之间是有联系的。一般存在两类联系：一是实体内部组成实体的属性之间的联系，二是实体之间的联系。我们讨论的是实体之间的联系。

两个实体之间的联系可以分为以下 3 类。

① 一对一联系 (1:1)，如一个部门有一个经理，而每个经理只在一个部门任职，则部门与经理之间具有一对一的联系。

② 一对多联系 (1:n)，如一个部门有若干职工，而每个职工只在一个部门工作，则部门与职工之间是一对多的联系。

③ 多对多联系 (m:n)，如一个项目有多个职工参加，而一个职工可以参加多个项目的工作，则项目与职工之间是多对多联系。

使用的基本符号如下。

- ：矩形，表示实体，框内注明实体名。
- ◇：菱形，表示实体间的联系，框内注明联系名。
- ：椭圆，表示实体的属性，框内注明属性名。
- ：无向边，连接实体与属性，或者实体与联系。

如图 1.1.1 所示，表示供应商这一实体；如图 1.1.2 所示，表示学生这一实体。

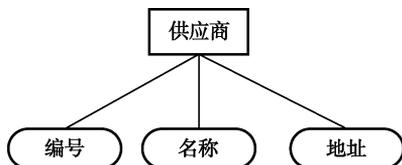


图 1.1.1 供应商实体图

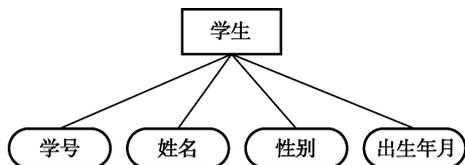


图 1.1.2 学生实体图

图 1.1.3 分别表示前面所说的三种联系。

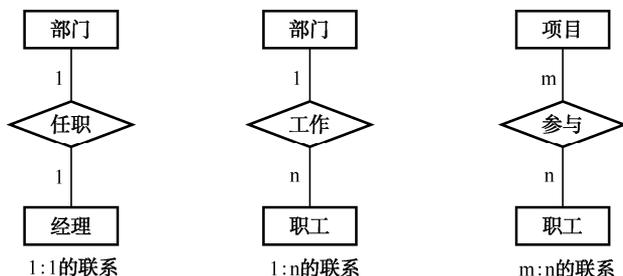


图 1.1.3 三种实体间的联系

联系也可以带有属性，如学生与课程存在学习的关系，学习有“成绩”这一属性；仓库存储零件有“库存量”的属性，如图 1.1.4 所示。

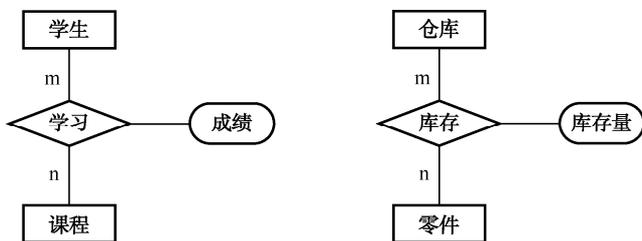


图 1.1.4 带有属性的联系

1) 三元关系

E-R 还可以表达更加复杂的关系。图 1.1.5 (a) 表示了课程、老师、参考书之间的关系：一门课可以由若干老师教授，用若干本参考书，而某一老师或某一本参考书只对应一门课。图 1.1.5 (b) 表示了供应商、项目、零件之间的关系：一个供应商可以供给若干项目多种零件，而每个项目可以使用不同供应商供应的零件，每种零件可由不同供应商供给。因此，供应商、项目和零件之间是多对多的联系。

2) 同一实体集实体之间的联系

同一实体集内的各实体之间可以存在某种联系，如职工实体集内具有领导和被领导的关系，学生实体集内具有管理和被管理的关系(如班长管理其他学生)。可以用图 1.1.6 表示这里所提的关系。

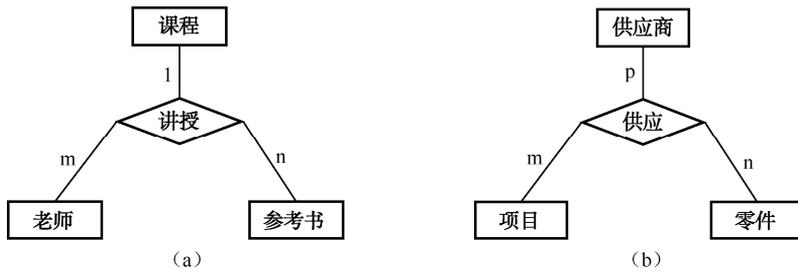


图 1.1.5 三元关系

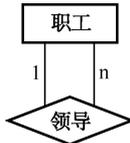


图 1.1.6 同一实体集内的联系

实例 1：物资管理。

物资管理涉及的实体有以下几点。

仓库：属性有仓库号、电话号码。

零件：属性有零件号、名称、单价。

供应商：属性有供应商编号、名称、电话。

项目：属性有项目编号、预算。

职工：属性有职工号、姓名、年龄、职称。

这些实体之间的联系如下。

(1) 一个仓库可以存放多种零件，一种零件可以存放在多个仓库中，因此仓库和零件具有多对多的联系。

(2) 一个仓库有多个职工当仓库保管员，一个职工只能在一个仓库工作，因此仓库和职工之间是一对多的联系。

(3) 职工之间具有领导与被领导的关系，即仓库主任领导若干保管员。

(4) 供应商、项目和零件三者之间是多对多的关系，参见上文所述。

省略各实体的属性，上述的“物资管理”可以用以下 E-R 图表示，如图 1.1.7 所示。

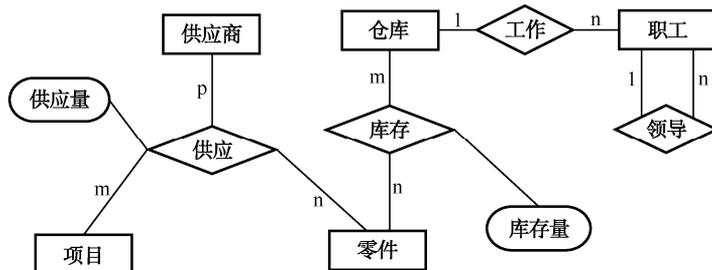


图 1.1.7 物资管理的 E-R 图

3. E-R 图向关系数据库转换

1) 关系数据库基本知识

我们把一个二维表称为一个关系。二维表由行和列组成，一列对应于一个字段，称为属

性；一行对应于一条记录，称为一个元组。

关系具有以下性质。

- ① 不允许有两行完全相同的记录。
- ② 行序不重要。
- ③ 每一个属性（列）是基本的、不可分裂的。
- ④ 每一列都有不同的名称，即在一个关系中属性的名称唯一。
- ⑤ 列序不重要。

2) 转换方法

(1) 实体：每一个实体转换为一个关系模式，即一个二维表，其属性为二维表中的列。

(2) 关系。

① 对于 1:n 的联系，可以将该关系对应于 1 的实体的关键字作为一个属性插入到 n 的实体关系中。

如在学生管理系统中：为了反映学生与班级之间的联系，可以把该联系中，对应于 1 的班级的关键字（班级编号）作为实体学生的一个属性，即作为学生资料表的一个列。

② 对于 m:n 的联系，应该将联系转换为一个新的关系模式，并且将关联实体的关键字作为这个关系模式的属性。

如在学生管理系统中，为了反映学生和课程的联系（即学生学习课程的成绩），应建立一个新的关系模式：成绩表（*学号、*课程编号、成绩）。

③ 对于 1:1 关系，则可以根据实际情况，看作 1:n 的特例，任选一方的关键字作为属性，插入到另一个关系中。

④ 对于三元关系，或其他多于 2 个实体之间的关系，一般应转换为一个新的关系模式，并且将关联实体的关键字作为这个关系模式的属性。

实例 2：库存管理。

某工厂中生产若干产品，每种产品由不同的零件组成，有的零件可用在不同的产品上。这些零件由不同的原材料制成，不同零件所用的材料可以相同，这些零件分类放在仓库中，原材料按照类别放在若干仓库中，用 E-R 图画出此工厂产品、零件、材料、仓库的概念模型，如图 1.1.8 所示。

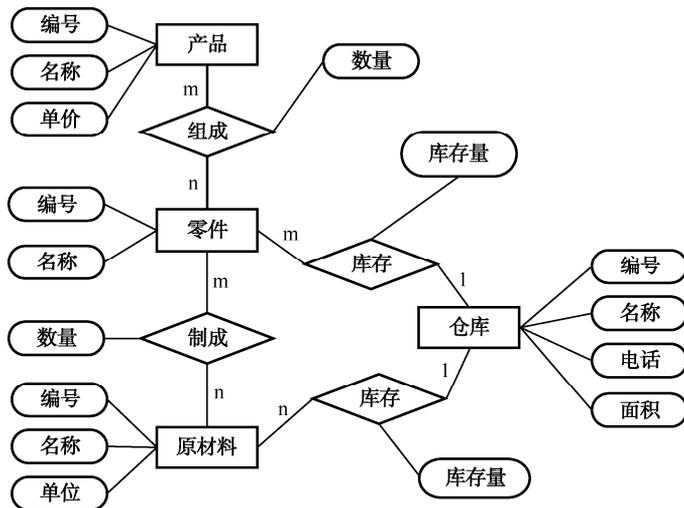


图 1.1.8 库存管理 E-R 图

注意

零件、原材料、仓库三个实体中，两两之间存在联系，但并非三元关系，所以不能作为三元关系处理。

根据转换规则，每个实体转换为一个关系模型；关系转换中，注意“零件-组成-产品”、“原材料-制成-零件”是多对多的联系，需要将该联系转换为一个新的关系模型。转换结果如下。

- ◆ 产品资料表 (*编号、名称、单价)。
- ◆ 零件资料表 (*编号、名称、仓库编号)。
- ◆ 原材料资料表 (*编号、名称、单位、仓库编号)。
- ◆ 仓库资料表 (*编号、名称、电话、面积)。
- ◆ 产品_零件表 (*产品编号、*零件编号、数量)。
- ◆ 零件_原材料表 (*零件编号、*原材料编号、数量)。

注意

前加*为关键字，下画线表示该实体属性是由联系转化而来的。

实例 3: 学生管理系统。

某学校有多个系，每个系包含一定数量的老师和班级，每个班级包含一定数量的学生，一个老师可担任一个或多个本系班级的管理。学生学习多门课程，一位老师可以教授一门或多门课程，某一门课程也可由多位老师任教，但某一班级某一门课程只能由一位老师任教。E-R 图如图 1.1.9 所示。

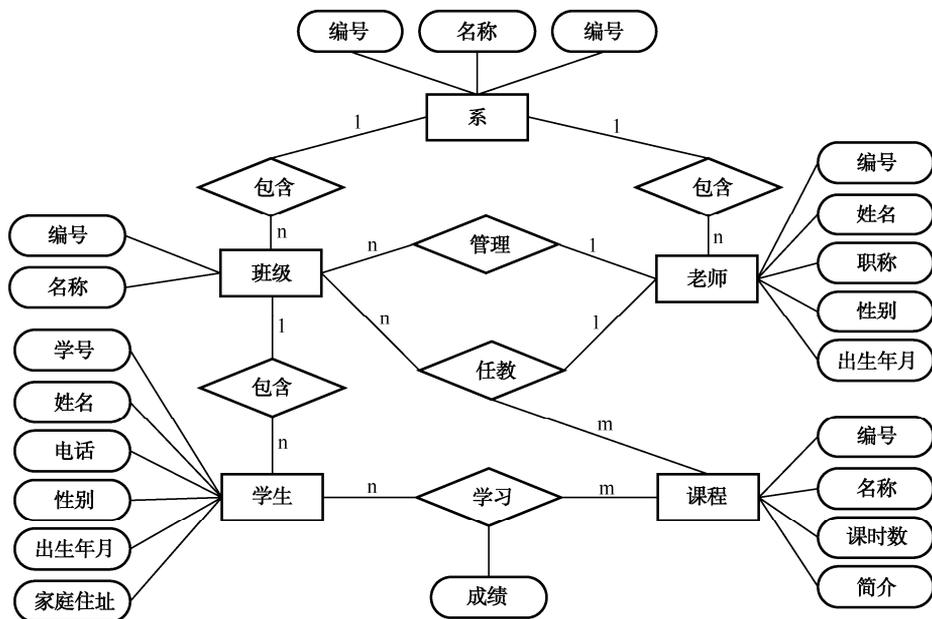


图 1.1.9 学校概念模型

根据转换规则，每个实体转换为一个关系模型；关系转换中，注意“学生-学习-课程”是

多对多的联系，需要将该联系转换为一个新的关系模型。转换结果如下。

- ◆ 系资料表 (*系编号、名称、电话)。
- ◆ 班级资料表 (*班级编号、班级名称、系编号、班主任)。
- ◆ 学生资料表 (*学号、姓名、电话、性别、出生年月、家庭住址、班级编号)。
- ◆ 老师资料表 (*编号、姓名、性别、出生年月、职称、系编号)。
- ◆ 课程资料表 (*课程编号、名称、课时数、简介)。
- ◆ 班级任课表 (*班级编号、*课程编号、教师编号)。
- ◆ 成绩表: *学号、*课程编号、成绩。

注意

前加*为关键字，下画线表示该实体属性是由联系转化而来的。

五、任务实施

1. 概念模型设计

针对任务需求中的 B2C 电子商务平台主要涉及如下实体：分类、产品、客户、订单、登录日志。E-R 模型如图 1.1.10 所示。

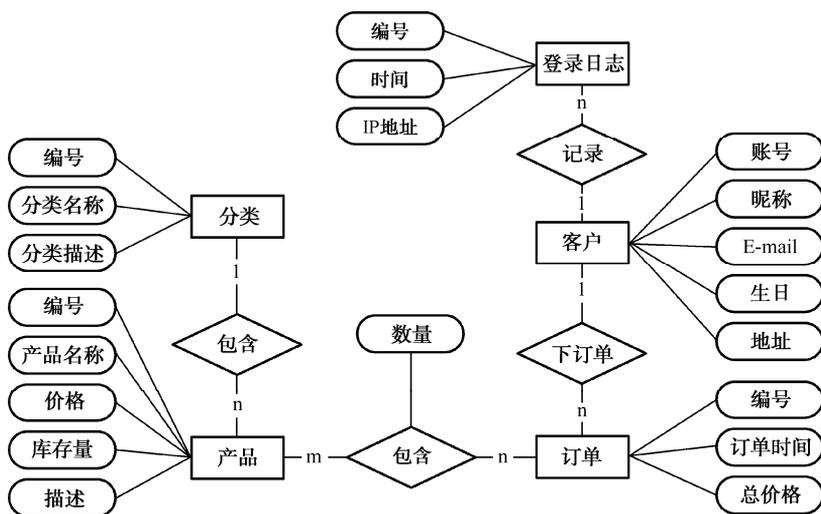


图 1.1.10 电子商务平台 E-R 图

2. 逻辑设计

逻辑设计的过程其实就是 E-R 图向关系模型转换的过程。根据转换规则，转换过程如下。

(1) 每一个实体转换为一个关系模型，转换结果如下。

- ◆ 分类信息表 (*编号、分类名称、分类描述)。
- ◆ 产品信息表 (*编号、产品名称、价格、库存量、描述)。
- ◆ 客户信息表 (*账号、昵称、E-mail、生日、地址)。
- ◆ 订单信息表 (*编号、订单时间、总价格)。
- ◆ 登录日志表 (*编号、时间、IP 地址)。

(2) 对于 1:n 的联系, 将该关系对应于 1 的实体的关键字作为一个属性插入到 n 的实体关系中。所以分类产品信息表中需要增加字段“编号”, 以表示产品和分类之间的关系; 同理, 登录日志表中增加字段“客户账号”, 订单表中增加字段“客户账号”; 修改后, 上述三个表变更为

- ◆ 产品信息表 (*编号、产品名称、价格、库存量、描述、分类编号)。
- ◆ 订单信息表 (*编号、订单时间、总价格、客户账号)。
- ◆ 登录日志表 (*编号、时间、IP 地址、客户账号)。

(3) 对于多对多的联系“订单-包含-产品”转换为一个新的关系模式, 并且将关联实体的关键字作为这个关系模式的属性, 形成新的“订单明细表”。

订单明细表 (*编号、订单编号、产品编号、数量)。

完整的转换结果如下。

- ◆ 分类信息表 (*编号、分类名称、分类描述)。
- ◆ 客户信息表 (*账号、昵称、E-mail、生日、地址)。
- ◆ 产品信息表 (*编号、产品名称、价格、库存量、描述、分类账号)。
- ◆ 订单信息表 (*编号、订单时间、总价格、客户编号)。
- ◆ 登录日志表 (*编号、时间、IP 地址、客户账号)。
- ◆ 订单明细表 (*编号、订单编号、产品编号、数量)。

六、课堂互动

(1) 请用 E-R 图表示出“出版社”、“图书”、“作者”三者之间的关系; 再将其转换为关系模型。注意, 出版社属性包括编号、名称、电话; 图书属性包括编号、名称、ISBN、价格; 作者属性包括姓名、年龄、电话。

(2) 在一个简单的 BBS 系统中, 请用 E-R 图表示出帖子 (主帖和跟帖)、用户、版块之间的关系。

注意

分两种情况讨论: 用户最多可以担任一个版块的版主、一个版块可以有多个版主; 用户可以担任多个版块的版主, 一个版块也可以有多个版主。

任务 2 数据库规范化

一、任务背景

通过前面任务 1 的学习, 小 Q 大概掌握了如何设计数据库, 但心中的疑问还有很多, 于是找到老李问: “不同的人针对同一个项目需求所设计出来的 E-R 图是相同的吗?”

老李: “不一定相同, 不同的人从不同的角度, 标识出不同的实体, 实体的属性划分可能又有不同, 最终得到的 E-R 图当然就不一定相同了。”

小 Q: “那如何判断哪个 E-R 图是正确的呢?”

老李: “其实, 针对同一项目, E-R 图不是唯一的, 不同的 E-R 图也可能都是正确的, 有些 E-R 图划分可能比较细, 分出来的表比较多, 便于数据的维护, 但查询的效率可能就比较低; 不同的 E-R 图则重点不同, 但不能说是不正确的。当然, E-R 图设计出来后必须符合一定的范式。”

小 Q: “范式?”

老李: “构造数据库必须遵循一定的规则, 在关系数据库中, 这种规则就是范式。要知道一个数据库的概念设计是否合理, 可以通过范式的理论去检验。”

如何应用范式理论去规范数据库设计呢?

二、任务需求

一个简单的 BBS 系统有若干版块, 注册用户可以在某一版块发表主帖, 也可以就某一个主帖发表跟帖; 用户可以担任多个版块的版主, 一个版块可以由多个用户担任。该系统数据库设计如下。

- ◆ 用户信息表 (用户编号*、用户名、密码、电子邮件、注册日期、联系电话)。
- ◆ 版块信息表 (版块编号*、版主用户编号*、版块名称)。
- ◆ 主帖表 (帖子编号*、标题、正文、发帖人编号、版块编号、版块名称、发帖时间)。
- ◆ 跟帖表 (帖子编号*、标题、正文、主帖编号、发帖人编号、发帖人姓名、发帖时间)。

请用范式理论检查该数据库设计是否符合第三范式, 如果不满足, 请优化数据库并使之满足第三范式。

三、任务分析

设计关系数据库时, 遵从不同的规范要求, 设计出合理的关系型数据库, 这些规范要求被称为范式, 各种范式呈递次规范, 越高的范式数据库冗余越小。而且, 高范式数据库不会发生插入 (Insert)、删除 (Delete) 和更新 (Update) 操作异常。反之, 如果数据库设计不遵循范式, 则会产生数据操作异常, 给编程人员制造了麻烦, 可能存储了大量不需要的冗余信息。

满足高等级的范式的先决条件是满足低等级范式, 如满足 2NF 一定要满足 1NF。针对该任务, 可以依次从低到高使用各种范式对数据库设计进行规范化。

四、知识要点

1. 范式理论基础

范式的目标主要有两个: 一是减少数据冗余; 二是消除异常, 包括插入异常、更新异常、删除异常。

目前, 关系数据库有 6 种范式: 第一范式 (1NF)、第二范式 (2NF)、第三范式 (3NF)、第四范式 (4NF)、第五范式 (5NF) 和第六范式 (6NF)。满足最低要求的范式是第一范式 (1NF)。在第一范式的基础上进一步满足更多要求的范式称为第二范式, 其余范式以此类推。一般来说, 数据库只需满足第三范式就可以了。

1) 第一范式

所谓第一范式是指在关系模型中，所有的域是原子性的、不可分割的，即数据库表的每一列都是不可分割的原子数据项。

简而言之，第一范式就是要求各个字段是不可分割的。

在任何一个关系数据库中，第一范式是对关系模式设计的基本要求，一般设计中都必须满足第一范式。

例如，某数据表中有“地址”字段，而系统中需要区分地址中的国家、城市等信息，则必须将“地址”字段划分成“国家”、“城市”、“街道”等字段。

2) 第二范式

第二范式是指关系模式在满足 1NF 的基础上，非主属性必须完全依赖于主键。2NF 在 1NF 基础上消除非主属性对主键部分函数的依赖。

第二范式是在第一范式的基础上建立起来的，即满足第二范式必须先满足第一范式。第二范式要求数据库表中的每个实例或记录必须可以被唯一地区分。选取一个能区分每个实体的属性或属性组，作为实体的唯一标识。

第二范式要求实体的属性完全依赖于主关键字。所谓完全依赖是指不能存在仅依赖主关键字一部分的属性，如果存在，那么这个属性和主关键字的这一部分应该分离出来形成一个新的实体，新实体与原实体之间是一对多的关系。为实现区分通常需要为表加上一个列，以存储各个实例的唯一标识。

简而言之，第二范式就是在第一范式的基础上属性完全依赖于主键。

例如，成绩表（学号*、课程编号*、学生姓名、课程名称、成绩），学号和课程编号两个字段组合成主键，成绩完全依赖于该主键（即成绩由学号、课程编号决定）；但是学生姓名和课程名称只是部分依赖于主键，学生姓名由学号决定，并不依赖于课程编号，同样，课程名称由课程编号决定，不依赖于学号。所以该关系表不符合 2NF。

正确的做法是对“成绩表”进行划分，可划分为：学生资料表（学号*、姓名）；课程资料表（课程编号*、课程名称）；成绩表（学号*、课程编号*、成绩）。

3) 第三范式

第三范式是指在满足 2NF 的基础上，任何非主属性不依赖于其他非主属性。3NF 在 2NF 基础上消除了传递依赖。

简而言之，第三范式就是属性不依赖于其他非主属性。

例如，员工资料表（员工编号*、员工姓名、部门编号、部门电话），其中，部门电话依赖于非关键字属性部门编号，所以不符合 3NF。也可以这样理解：部门编号依赖于员工编号，而部门电话依赖于部门编号，这就构成了传递依赖，不符合 3NF。

正确的做法是增加一个部门资料表：员工资料表（员工编号*、员工姓名，部门编号）；部门资料表（部门编号*、部门电话）。

2. 范式总结

上面介绍的 3 种范式存在的关系以及范式规范化的过程如图 1.2.1 所示。

范式规范化的过程很多时候就是不断拆分表的过程，应用的范式等级越高，则表拆分得越多。表拆分过多带来的主要问题如下。

(1) 查询时要连接多个表，增加了查询的复杂度。

(2) 查询时要连接多个表，降低了数据库查询性能。

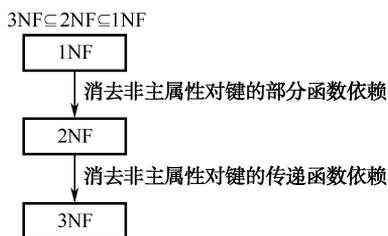


图 1.2.1 范式之间的关系

而以现在的情况，磁盘空间成本基本可以忽略不计，所以数据冗余所造成的问题也并不是应用数据库范式的理由。因此，并不是应用的范式等级越高越好，要根据实际情况而定。第三范式已经很大程度上减少了数据冗余，并且减少了造成的插入异常、更新异常和删除异常，多数情况下应用到第三范式已经足够。

五、任务实施

1. 第一范式

1NF 要求每个字段都不可再分割。针对任务需求中的数据表，观察论坛用户表。

论坛用户表（用户编号*、登录账号、密码、电子邮件、注册日期、联系电话）。

其中，联系电话不太明确，可以是手机号码，也可以是家庭电话，或者同时有手机号码和家庭电话。所以，根据具体需求将该字段划分成手机号码、家庭电话。

新的论坛用户表：

论坛用户表（用户编号*、登录账号、密码、电子邮件、注册日期、手机号码、家庭电话）。

2. 第二范式

2NF 要求属性完全依赖于主键。在该案例中用户可以担任多个板块的版主、一个板块可以由多个用户担任版主，设计数据库时使用了版块信息表来反映这种关系，仔细观察该表：

版块信息表（版块编号*、版主用户编号*、版块名称）。

该表使用了组合关键字版块编号、版主用户编号，而版块名称由版块编号决定，版块名称依赖于版块编号，即版块编号只是部分依赖于关键字，不符合第二范式。可以将该表进一步划分成以下两个表。

版块信息表（版块编号*、版块名称）。

版块版主信息表（版块编号*、版主用户编号*）。

3. 第三范式

3NF 要求不存在传递依赖，或者说非主属性不依赖于其他非主属性。仔细观察主帖表：

主帖表（帖子编号*、标题、正文、发帖人编号、版块编号、版块名称、发帖时间）。

版块名称和版块编号两个非主属性存在依赖关系，版块名称依赖于版块编号，版块信息表已经存在，所以在该表中只需记录版块的编号即可：

主帖表（帖子编号*、标题、正文、发帖人编号、版块编号、发帖时间）。

同样，跟帖表中发帖人编号、发帖人姓名两个非主属性也存在依赖关系，已经存在用户信息表，这里只需要发帖人编号即可：

跟帖表（帖子编号*、标题、正文、主帖编号、发帖人编号、发帖时间）。

至此，经过范式规范化后，得到的数据设计如下。

(1) 论坛用户表（用户编号*、登录账号、密码、电子邮件、注册日期、手机号码、家庭电话）。

(2) 版块信息表（版块编号*、版块名称）。

(3) 版块版主信息表（版块编号*、版主用户编号*）。

(4) 主帖表（帖子编号*、标题、正文、发帖人编号、版块编号、发帖时间）。

(5) 跟帖表（帖子编号*、标题、正文、主帖编号、发帖人编号、发帖时间）。

六、课堂互动

(1) 观察某商品库存信息表，看看其符合第几范式，如何优化？

商品库存信息表（商品编号*、商品名称、单价、数量、总价）。

(2) 观察图书借阅系统的数据库设计，看看是否符合 3NF，如果不符合，请将其优化使之符合 3NF。

图书资料表（图书编号*、图书名称、读者编号*、读者名称、借出日期、归还日期）。

拓展实训 1-1 图书管理系统数据库设计

【实训目的】

掌握完整的数据库设计方法，包括数据库需求分析、概念设计、逻辑设计，并应用范式理论优化数据库设计。

【实训内容】

某图书馆计划开发一个简单的图书管理系统，要求该系统能实现图书的检索、借阅、归还功能，图书馆的读者主要有学生、老师；系统暂不考虑超时罚款、图书征订、库存管理等功能。请根据该基本需求，进一步细化用户需求，并进行概念设计（画出 E-R 图）、逻辑设计（关系模型），并应用范式理论优化数据库设计。

【训练要点】

因为系统只考虑基本的借阅、归还功能，所以系统涉及最基本的实体有“读者”、“图书”，读者有类型的属性，不同类型读者的借书数量和借书期限可能不同，读者和图书之间存在“借阅”、“归还”关系。根据这些分析，可以基本上确定 E-R 图了。



本项目介绍了数据库、数据库系统、数据库管理系统等相关概念，并通过两个任务介绍了数据库设计的方法、步骤，从中可以了解到数据库设计的步骤，掌握使用 E-R 模型设计概念模型的方法和应用范式规范数据库的设计。



习题 1

一、选择题

- 在数据库设计中使用 E-R 图工具的阶段是 ()。
 - 需求分析阶段
 - 数据库物理设计阶段
 - 数据库实施阶段
 - 概念结构设计阶段
- 数据库设计中的逻辑结构设计的任务是把 () 阶段产生的概念数据库模式变换为逻辑结构的数据库模式。
 - 需求分析
 - 物理设计
 - 逻辑结构设计
 - 概念结构设计
- 公司中有多个部门和多名职员, 每个职员只能属于一个部门, 一个部门可以有多名职员, 从职员到部门的联系类型是 ()。
 - 多对多
 - 一对一
 - 多对一
 - 一对多
- 关系数据库中, 一个关系代表一个 ()。
 - 表
 - 查询
 - 行
 - 列
- 关系数据库中, 一个元组是一个 ()。
 - 表
 - 查询
 - 行
 - 列
- 数据库系统的核心是 ()。
 - 用户
 - 数据
 - 数据库管理系统
 - 硬件
- DBMS 代表 ()。
 - 用户
 - 数据
 - 数据库管理系统
 - 硬件
- 建立在操作系统之上, 对数据库进行管理和控制的一层数据管理软件是 ()。
 - 数据库
 - 数据库系统
 - 数据库管理系统
 - 数据库应用系统
- 学生社团可以接纳多名学生参加, 但每个学生也可参加多个社团, 从社团到学生之间的联系类型是 ()。
 - 多对多
 - 一对一
 - 多对一
 - 一对多
- 下列四项说法中不正确的是 ()。
 - 数据库减少了数据冗余
 - 数据库中的数据可以共享
 - 数据库避免了一切数据的重复
 - 数据库具有较高的数据独立性
- 下列四项中, 不属于数据库特点的是 ()。
 - 数据共享
 - 数据完整性
 - 数据冗余很高
 - 数据独立性高
- 在数据库系统中数据模型有三类, 它们是 ()。
 - 实体模型、实体联系模型、关系模型
 - 层次模型、网状模型、关系模型

C. 一对一模型、一对多模型、多对多模型

D. 实体模型、概念模型、存储模型

13. 一个学生只能就读于一个班级，而一个班级可以同时容纳多个学生，班级与学生之间是（ ）的关系。

A. 一对一

B. 一对多

C. 一对零

D. 多对多

14. E-R 图中，关系用下面的（ ）来表示。

A. 矩形

B. 椭圆形

C. 菱形

D. 圆形

15. 有如下表结构，带有*的字段代表主键或组合主键，一份订单可以订购多种产品。

产品：*产品编号，产品名称，产品价格。

订单：*订单编号，*产品编号，订购日期，订购数量（订单编号与产品编号是组合主键）。

该表最高符合第（ ）范式。

A. 一

B. 二

C. 三

D. 不符合任何范式

16. 有如下表结构，带有*的字段代表主键或组合主键，一份订单可以订购多种产品。

产品：*产品编号，产品名称，产品价格。

订单：*订单编号，总价，订购日期。

订单明细：*明细编号，订单编号，产品编号，订购数量。

该表最高符合第（ ）范式。

A. 一

B. 二

C. 三

D. 不符合任何范式

二、填空题

1. 已知有课程信息表（课程号、课程名称、课时数）和学生信息表（学号、姓名、性别）两个表，课程信息表的主键是_____，学生信息表的主键是_____；学生与课程之间是多对多关系，可以用“选课成绩表”这种关系，则“选课成绩表”包含的字段有_____，主键为_____。

2. 实体之间的联系可以分为3类：_____、_____、_____。

三、问答题

1. 举例说明什么是一对多关系。

2. 举例说明什么是多对多关系。

3. 数据库设计一般包含哪几个阶段？

4. 某医院的病房管理系统涉及的实体如下。

科室：科室名、科地址、科电话。

病房：病房号、床位号。

医生：姓名、职称、年龄、工作证号。

病人：病历号、姓名、性别。

相关业务规定如下。

① 一个科室有多个病房、多个医生；

② 一个病房只能属于一个科室；

③ 一个医生只属于一个科室，但可负责多个病人的诊治；

④ 一个病人的主管医生只有一个。

根据这些业务规定，制作 E-R 图，并将 E-R 图转换为关系模型。

5. 某销售部门子系统涉及的实体如下。

职工：职工号、姓名、地址和所在部门。

部门：部门所有职工、部门名、经理和销售的产品。

产品：产品名、制造商、价格、型号和产品内部编号。

制造商：制造商名称、地址、生产的产品名和价格。

相关业务规定如下。

① 部门有很多职工，职工仅在一个部门工作；

② 部门销售多种产品，这些产品也在其他部门销售；

③ 制造商生产多种产品，其他制造商也制造这些产品。

根据这些业务规定，制作该系统的 E-R 模型。