审计署计算机审计中级培训系列教材

计算机审计

——数据采集与分析技术

董化礼　刘汝焯等　编著
王智玉　　审定

清 华 大 学 出 版 社

（京）新登字 158 号

内 容 简 介

面向数据的计算机审计，是计算机审计的一个重要领域。本书紧密结合我国国民经济信息化发展的实际，尤其是财政财务收支和相关业务活动信息化建设的实际，紧紧围绕着数据采集—数据转换—数据分析这一主线，系统论述了面向数据的计算机审计，既注意阐述必要的基础理论知识，又注意侧重于实际操作和开发技术，同时附有《审计数据采集分析 2.0》光盘和审计案例。

本书内容共分为 10 章，分别介绍了电子数据的组织、处理和存储；审计接口和数据库访问技术；审计数据的采集、转换和分析技术；常用的 SQL 语言以及《审计数据采集分析 2.0》软件。

本书可操作性、指导性强，非常适合包括政府审计、社会审计、内部审计在内的各类审计人员学习，同时也适合所有关心财政财务收支和相关经济活动真实、合法、效益状况的读者，包括各级领导、纪检监察干部、公检法干部、投资者学习。既能引导读者掌握计算机审计的操作知识，又能为他们的审计软件开发和理论研究提供帮助。

书　　名：计算机审计　数据采集与分析技术
作　　者：董化礼　刘汝焯　等编著
出 版 者：清华大学出版社（北京清华大学学研大厦，邮编 100084）
http://www.tup.tsinghua.edu.cn
责任编辑：王　青
印 刷 者：北京市清华园胶印厂
发 行 者：新华书店总店北京发行所
开　　本：787×1092 1／16　**印张：**17　**字数：**387 千字
版　　次：2002 年 6 月第 1 版　2002 年 6 月第 1 次印刷
书　　号：ISBN 7-900641-90-4
印　　数：0001~5000
定　　价：36.00 元(含光盘)

前　言

面向数据的计算机审计，也就是对计算机信息系统中输入、处理和输出的电子数据进行审计，是计算机审计的一个重要领域。本书紧密结合我国国民经济信息化发展的实际，尤其是财政财务收支和相关业务活动信息化建设的实际，紧紧围绕着数据采集—数据转换—数据分析这一主线，系统论述了电子数据的特点、计算机信息系统中数据文件的设置、审计接口开发、数据库访问技术、数据采集、数据转换、数据分析和审计软件等知识。本书既注意介绍必要的理论知识，如数据库原理、ODBC、数据库访问接口、SQL语言，又注意侧重于实际操作和开发技术，同时附有《审计数据采集分析 2.0》光盘和审计案例，非常适合广大审计工作者学习，也可为计算机审计软件开发人员提供参考。

努力提高可操作性和指导性是本书编写过程中贯彻始终的指导思想。第 1、2 两章深入浅出地介绍了电子数据的特点，力求使读者一开始就清晰地了解什么是计算机中的凭证、账簿、报表，它和纸质的数据有什么不同；第 3、4、5 章论述审计接口和数据库访问技术，告诉读者通过哪些渠道才能把计算机中的电子数据采集过来，以及怎样采集；第 6、7、8 章讲解如何将采集到的电子数据转换成审计人员需要的格式，如何进行查询、挖掘、验证分析，收集审计线索；第 9、10 两章介绍了审计软件的概念、特点和操作。本书按照计算机审计操作实务的逻辑来安排章节结构，力求使本书介绍的知识与计算机审计的实际相结合，做到易学、易懂、易用。

服务主体明确，适应范围广，是本书追求的又一个目标。本书是写给审计人员看的，包括政府审计、社会审计、内部审计的管理者和审计师，但又不仅限于此；本书同时是写给所有关心财政财务收支和相关经济活动真实、合法、效益状况的读者看的，包括各级领导、纪检监察干部、公检法干部、投资者，因为国民经济信息化已经进入了我们的生活，无纸化的趋势越来越紧迫，看不懂电子账就不能了解情况，就成了高科技环境下的“文盲”，就不能公正、准确地判断是非——这已经是不争的事实。就“审计人员”这一群体来讲，作者也力求适应两个层次，一是在审计第一线进行计算机审计实务的审计部门、社会审计组织以及内部审计人员，让他们掌握计算机审计的操作知识，用本书介绍的知识指导审计实践；二是计算机审计软件的开发人员，为他们的软件开发和理论研究提供帮助。

董化礼设计了全书的总体框架，制定了编写要求，组织了全书的编写。

刘汝焯、万建国、杨小虎、潘连安、朱文明、孙兴国、马社亮、杨晓宇、程建勤参加了 1~10 章的具体编写。各章成稿后由万建国、杨小虎对有关章节做了修改和充实。陈虹、黄维江、罗锋、熊立新、刘素合参加了审计案例的撰写。全书由刘汝焯统稿。

王智玉审定了全书。

审计署驻上海特派员办事处特派员刘海彬对本书的撰写给予了热情鼓励和大力支持，在此谨向他，向关心和支持本书出版的所有同志表示衷心的感谢。

我国的计算机审计目前正处在全面起步的阶段，无论是理论研究还是审计实务，都在探索之中，计算机审计的技术、方法、手段、评价指标体系都依赖于审计工作者深入的研究和开发。由于作者水平有限，加之时间仓促，书中的错误之处在所难免，敬请读者不吝赐教。

刘汝焯

2002 年 3 月 9 日于扬州

目　录

第 1 章　电子数据的组织、处理和存储

内容提要：本章介绍电子数据处理的特点，包括电子数据的组织结构与存储管理，重点介绍关系数据模型和关系数据库。通过一组财务数据文件的实例引出并阐述了会计信息系统的文件设计和代码设计。这些知识对于读者认识电子数据、了解电子数据、掌握电子数据的特点是非常宝贵的。

1.1　电子数据处理的特点

在计算机信息系统中，输入的原始数据、处理的中间结果和最后结果一般都是以数据文件的形式存储的。要对被审计信息系统输出的信息的真实性、正确性、合法性等进行评价，必须对数据文件进行审计。面向数据的审计就是对数据文件的审计。数据文件审计可以分为对打印输出的数据文件的审计或对存储在磁性介质（本书所指的磁性介质不仅包括磁带、磁盘，也包括光盘等其他存储介质，下同）上的数据文件的审计。本书所指的数据文件的审计，是指对存储在磁性介质上的数据文件的审计。

磁性介质上的数据文件不同于手工环境下的纸质凭证、账簿和报表，它有着明显不同的特点。

- 数据记录的载体不同，数据记录在磁、光、电等介质上，是无纸化的。由于记录载体的不同，带来了许多信息输入、处理、输出的深刻变化。
- 肉眼不可见，是以计算机可读的形式存在的，一旦机器出现故障，信息就无法取出和使用。要审查这些数据文件，必须利用计算机技术。
- 修改或删除不留痕迹，就像在日常生活中轻易地把录音带或录像带上原来录有的内容洗掉或重录上新的内容而不会留下痕迹一样。在已发现的计算机犯罪案件中，有不少就是通过直接篡改数据文件来实现的。
- 存储电子数据的磁盘、磁带、光盘易于损坏，它们不能折叠，害怕潮湿、高温、灰尘、电磁场。如果保管不善，存储在磁性介质上的电子数据很容易因介质的损坏而丢失。
- 由于存储介质和数据组织的变化，数据的操作变得十分便利，如检索、查询、排序、统计、计算，效率大大提高，数据的保存和转移也变得十分容易，为审计的计算机化提供了良好的前提条件。
- 计算机信息系统的数据库平台不同，数据文件的格式也就不同。对这些数据文件的访问要求高，对数据文件的采集必须运用相应的计算机技术。
- 计算机信息系统中的凭证、账簿和报表的设置和手工不同，有明显差异。

- 数据表示代码化，这是计算机信息系统数据处理的一个显著特点。在手工作业系统中，纸性介质上的信息是用人们熟悉的文字和数字表示的。在计算机信息系统中，为了使信息更便于计算机处理，为了提高计算机处理的速度和节省存储空间，也为了方便操作人员的操作，尽可能减少汉字的输入，大量的数据要用代码来表示。例如会计科目、部门、职工、产成品、原材料、固定资产、主要的顾客或供应商等等，都常用适当设计的代码来表示，甚至记账方向和部分较规范、常用的摘要也用代码表示。用各种代码表示有关的业务信息有利于计算机的处理，但不便于审计人员的使用。

1.2 电子数据的存储与管理

电子数据处理的中心问题是数据存储和管理，它指的是对数据的分类、组织、编码、存储、检索和维护。随着计算机硬件和软件的发展，对电子数据的存储和管理经历了三个阶段：

- 人工管理阶段
- 文件系统阶段
- 数据库系统阶段

1.2.1 人工管理阶段

20 世纪 50 年代中期以前，计算机主要应用于科学计算。数据一般不需要长期保存，同时数据量较小，所以数据的组织方式没有数据文件的概念，只是用磁带、卡片、纸带来存储，也没有软件系统来对数据进行管理。程序员不仅要规定数据的逻辑结构，还要在程序中设计物理结构，包括存储结构、存取方法、输入输出方式等。因此程序中用于存取数据的子程序随着数据的逻辑结构或物理结构的改变而改变，即数据和程序之间不具有独立性，一组数据只对应一个应用程序。

人工管理阶段应用程序与数据之间的关系如图 1-1 所示。

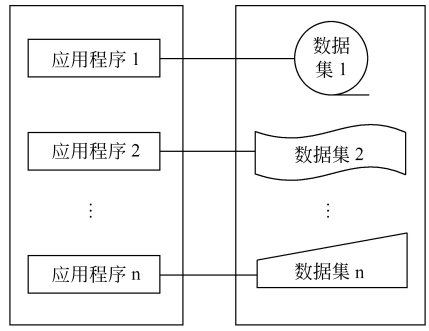


图 1-1　人工管理阶段应用程序与数据之间的对应关系

1.2.2 文件系统阶段

从 20 世纪 50 年代后期到 60 年代中期，计算机不仅用于科学计算，还大量应用于管理领域。这时候有了磁盘、磁鼓等存储设备。在软件方面，计算机操作系统中已经有了专门的管理数据的软件，通常称为文件系统。从处理方式来看，不仅有了文件批处理，而且能够联机实时处理。电子数据在这一阶段是以数据文件的形式存储在存储设备上的。

数据文件是以某种数据结构将电子数据组织、保存起来，以方便数据存取的文件。在文件系统阶段，数据文件已经多样化，有了索引文件、链接文件、直接存取文件等。在数据文件中，数据的存取基本上是以记录为单位的。

文件系统是计算机操作系统的重要组成部分，对于由电子数据组织成的相互独立的数据文件，利用“按文件名访问，按记录进行存取”的管理技术，进行数据修改、插入和删除等操作。

文件系统与人工管理阶段相比具有很多优点。由软件进行数据管理，程序和数据之间由软件提供的存取方法进行转换，有共同的数据查询修改的管理模块。文件的逻辑结构与存储结构由系统进行转换，程序与数据之间有了一定的独立性。

但是，文件系统仍有很多缺点，主要有：

- 数据冗余度（redundancy）大

文件系统中的文件基本上对应于某个应用程序。也就是说，数据还是面向应用的。当不同的应用程序所需要的数据有部分相同时，仍必须建立各自的文件，而不能共享相同的数据，因此数据冗余度大，浪费存储空间。相同数据的重复存储与各自管理给数据的修改和维护带来了困难。极易造成数据的不一致性。

- 数据和程序缺乏独立性

文件系统中的数据文件是为某一特定应用服务的。文件的逻辑结构对该应用程序来说是优化的。因此，要想对现有的数据再增加一些新的应用是很困难的，系统不容易扩充。一旦数据的逻辑结构发生改变，必须修改应用程序，修改文件结构的定义。而应用程序的改变，如应用程序所使用的高级语言的变化等，也将影响文件的数据结构的改变。数据和程序之间缺乏独立性。因此文件系统仍然是一个不具有弹性的数据集合，文件与文件之间是孤立的，不能反映现实世界事物之间的内在联系。

文件系统阶段应用程序与数据之间的关系如图 1-2 所示。

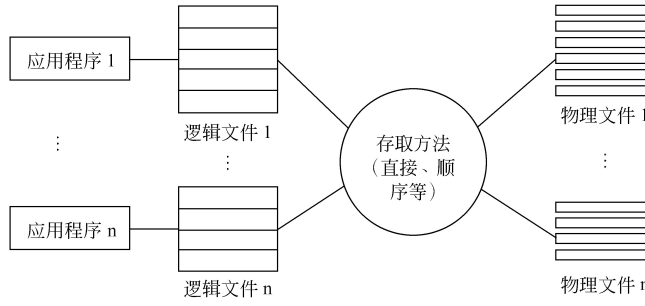


图 1-2　文件系统阶段应用程序与数据之间的对应关系

1.2.3 数据库系统阶段

这一时期（从 20 世纪 60 年代后期开始至今）的背景是：

- 计算机用于管理的规模更为庞大，应用越来越广泛，数据量急剧增长，而且数据的共享要求越来越强。
- 有了大容量的磁盘。
- 联机实时处理要求更多了，并开始提出和考虑分布处理。
- 软件价格上升，硬件价格下降。为编制和维护系统软件及应用程序所需的成本相对增加。

为了解决多用户、多应用共享数据的需求，使数据为尽可能多的应用服务，就出现了数据库、数据库管理系统、数据库系统。

数据库（database）是长期存储在计算机内的、有组织的、可共享的数据集合。

数据库管理系统（database management system，DBMS）是建立在操作系统的基础上，对数据库的建立、使用和维护进行管理的软件。

数据库管理系统作为界于用户和操作系统之间的一层数据管理软件，它的主要功能包括以下几个方面：

- 数据定义功能。用户使用 DBMS 提供的数据定义语言（DDL），可以方便地定义数据。
- 数据操纵功能。用户通过 DBMS 提供的数据操纵语言（DML）可以实现查询、插入、删除、修改等数据操作。
- 数据库运行管理。这是 DBMS 的核心部分，包括并发控制、安全性检查、完整性约束条件的检查和执行、数据库内部维护等（如索引、数据字典的自动维护）。
- 数据库的建立和维护功能。包括数据库初始数据的输入、转换功能，数据库的转存、恢复功能，数据库的重组织功能和性能监视、分析功能等等。

数据库管理系统在计算机系统中的地位如图 1-3 所示。

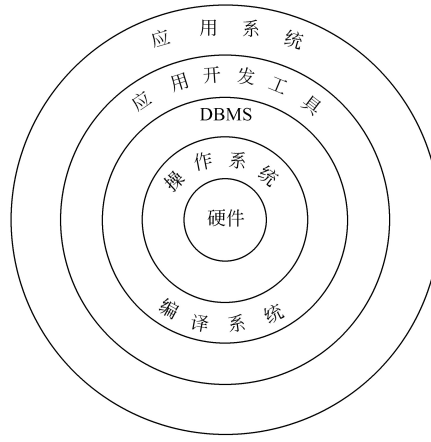


图 1-3　数据库管理系统在计算机系统中的地位

数据库系统（database system，DBS）通常是指带有数据库的计算机系统。因此，广义地讲，数据库系统不仅包括数据库本身，还包括相应的硬件、软件和各类相关的人员。其中，硬件是指计算机（有足够大的内存、外存和系统处理能力）；软件是指 DBMS、支持 DBMS 的操作系统以及以 DBMS 为核心的应用开发工具；人员是指数据库的开发、管理、使用过程中的数据库管理人员（DBA）、系统分析员、应用程序员和最终用户。

数据库系统可以用图 1-4 表示。

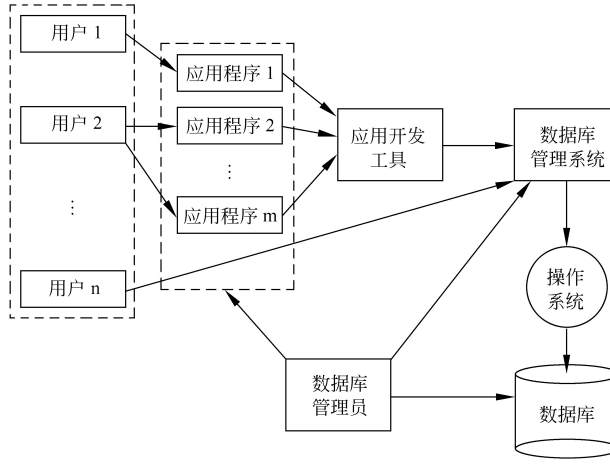


图 1-4　数据库系统

数据库系统与文件系统相比有以下特点：

- 面向全组织的复杂的数据结构

这就要求在描述数据时不仅描述数据本身，还要描述数据之间的联系。文件系统中尽管记录内部已有了某些结构，但记录之间是没有联系的、孤立的。因此，数据的结构化是数据库的主要特征之一，也是数据库系统与文件系统的根本区别。

- 数据冗余度小、易扩充

由于数据库从整体观点来看待和描述数据，数据不再是面向某一应用，而是面向整个系统，同样的数据不再需要重复存储，这就大大降低了数据的冗余度，既节约存储空间，缩短存取时间，又可避免数据之间的不相容性和不一致性。对数据库数据的应用可以有很灵活的方式，可以取整体数据的各种合理子集用于不同的应用系统。而且当应用需求改变或增加时，只要重新选取不同子集或者加上一小部分数据，便可以有更多的用途，满足新的要求。这就是弹性大，易扩充的特点。

- 数据和程序的独立性较高

数据库系统提供了两方面的映像功能：数据的存储结构与逻辑结构之间的映像或转换功能；数据的总体逻辑结构与某类应用所涉及的局部逻辑结构之间的映像或转换功能。

- 具有统一的数据控制功能

这些统一的数据控制功能包括对数据安全性、数据完整性、多用户并发性的控制。

数据库管理阶段应用程序与数据之间的关系如图 1-5 所示。

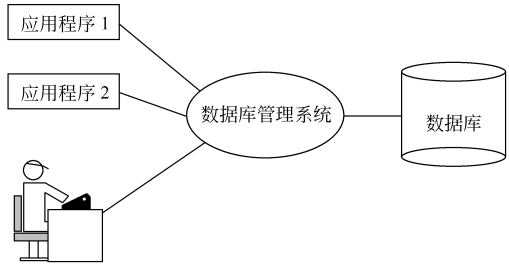


图 1-5　数据库管理阶段应用程序与数据之间的对应关系

1.3　电子数据的组织与结构

1.3.1　数据模型

数据库是某个企业、组织或部门所涉及的数据的综合。它不仅反映数据本身的内容，而且反映数据之间的联系。如何抽象、表示和处理现实世界中的数据和信息呢？在数据库中是用数据模型这个工具来对现实世界进行抽象的。数据模型是数据库系统中用于提供信息表示和操作手段的形式构架。

数据模型是提供模型化数据和信息的工具。根据应用的不同目的，可以将模型分为两类或者说两个层次。一是概念模型（也称信息模型），一是数据模型（如网状、层次、关系模型）。前者是按用户的观点来对数据和信息建模，后者是按计算机系统的观点对数据建模。

概念模型用于信息世界的建模。这类模型强调其语义表达能力，要能够较方便、直接地表达应用中各种语义知识。这类模型应当概念简单、清晰、易于用户理解，因为它是现实世界到信息世界的第一层抽象，是用户和数据库设计人员之间进行交流的语言。第二层次的模型用于机器世界。这类模型通常需要有严格的形式化定义，而且常常会加上一些限制或规定，以便于在机器上实现。它通常有一组严格定义了语法和语义的语言，人们可以使用它来定义、操纵数据库中的数据。

1.3.2　数据模型的三要素

一般地讲，数据模型是严格定义的概念的集合。这些概念精确地描述系统的静态特性、动态特性和完整性约束条件。因此，数据模型通常由数据结构、数据操作和完整性约束条件三部分组成。

1. 数据结构

数据结构是所研究的对象类型的集合。这些对象是数据库的组成成分。一般可分为两类：一类是与数据类型、内容、性质有关的对象；一类是与数据之间联系有关的对象。在数据库系统中通常按照数据结构的类型来命名数据模型，如层次结构、网状结构和关系结构的模型分别命名为层次模型、网状模型和关系模型。

2. 数据操作

数据操作是指对数据库中各种数据允许执行的操作的集合，包括操作及有关的操作规则。数据库主要有检索和更新（包括插入、删除、修改）两大类操作。数据模型要定义这些操作的确切含义、操作符号、操作规则（如优先级别）以及实现操作的语言。数据结构是对模型静态特性的描述；数据操作是对模型动态特性的描述。

3. 数据的约束条件

数据的约束条件是完整性规则的集合。完整性规则给定数据间的制约和依存规则，用以限定符合数据模型的数据库状态以及状态的变化，以保证数据正确、有效、相容。数据模型应该反映和规定符合这种数据模型所必须遵守的基本的通用的完整性约束条件。

此外，数据模型还应该提供定义完整性约束条件的机制，以反映某一部门的应用所涉及的数据必须遵守的特定的语义约束条件。

数据模型这三个方面的内容完整地描述了一个数据模型，其中数据结构是刻画模型性质最重要的方面。

1.3.3 概念模型

数据模型是数据库系统的核心和基础。各种机器上实现的 DBMS 软件都是基于某种数据模型的。为了把现实世界中的具体事物抽象、组织为 DBMS 支持的数据模型，人们常常首先将现实世界抽象为信息世界，然后将信息世界转换为机器世界。也就是说，首先把现实世界中的客观对象抽象为某一种信息结构，这种信息结构并不依赖于具体的计算机系统，不是某一个 DBMS 支持的数据模型，而是概念级的模型。然后再把概念模型转换为计算机上某一 DBMS 支持的数据模型。因此概念模型是现实世界到机器世界的一个中间层次。

实体—联系（entity-relationship，E-R）模型是应用最广泛、最方便的概念模型。在实体—联系模型中，最主要的两个概念是实体和联系。实体是指现实世界中客观存在并可相互区别的事物，例如学生、教师、课程、职工、部门、工资表等；而事物与事物间是有联系的，在 E-R 模型中实体与实体间的联系分为三类：

- 一对一的联系（1：1）

例如在学校里一个班级对应一个班主任，在单位里一个职工对应一张工资条，两者间是一一对应的。

- 一对多的联系（1：n）

例如在学校里一个班级中可以有多名学生，一个学生只能属于一个班级，或者在单位里一个部门可以有多名职工，一个职工只能属于一个部门。

- 多对多的联系（m：n）

例如在学校里一个学生可以学多门课程，一门课程可以有多名学生，或者在工厂里一个产品由多种原材料生产而成，一种原材料可以用来生产多种产品，两者间的对应关系是多对多。

为表述方便，E-R 模型一般用图表示，上面这三种联系分别可以用图 1-6、图 1-7 和图 1-8 来表示。

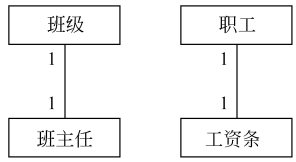


图 1-6　实体间一对一的联系

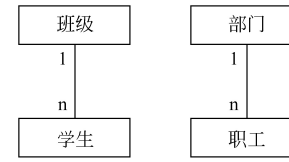


图 1-7　实体间一对多的联系

在建立计算机信息系统时，系统分析员需要应用 E-R 模型把现实世界中要描述的事物抽象为一个概念模型，这个模型往往是一个图示化表示的 E-R 模型。把上面例子中的班级、班主任、学生、课程放在一起描述，可以得到如图 1-9 所示的 E-R 模型。

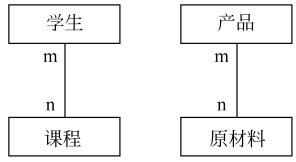


图 1-8　实体间多对多的联系

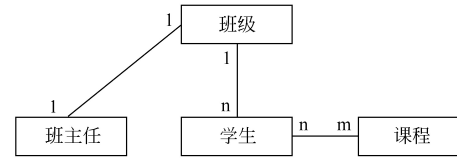


图 1-9　E-R 模型实例

1.3.4　关系模型

当前数据库系统中所支持的数据模型主要是关系模型（relational model），自 20 世纪 80 年代以来，计算机厂商推出的 DBMS 产品几乎都是支持关系模型的。如目前常见的大型数据库管理系统 Oracle、Sybase、DB2、Informix、Microsoft SQL Server 和微机数据库系统软件 Access、FoxPro、dBase 等均以关系模型为理论基础。

1. 关系模型的数据结构

关系模型是建立在数学概念基础上的。关系模型用于计算机处理，是关系数据库系统的基础，因此在信息系统建设时，完成描述客观世界的 E-R 模型后，需要把 E-R 模型转换为关系模型。在关系模型中，数据的逻辑结构是一张二维表，E-R 模型中的实体及实体的属性就表示为一张二维表，实体间一对一、一对多的联系用二维表之间的属性引用表述，实体间多对多的联系则需要再建立一张二维表表示。

把上面关于职工、部门的 E-R 模型转换为用于计算机处理的关系模型，得到如表 1-1 和表 1-2 所示的两张表。

下面介绍关系模型中的主要术语：

(1) 关系

一个关系对应于一张二维表。

(2) 元组

表中的一行称为一个元组，即一条记录。

表 1-1　职工信息表

身份证号	姓名	性别	部门编号	年龄	工资
330106xxxxxxxxx	张华	女	01	36	1 200
330201yyyyyyyyy	李维	男	02	56	1 250
210901zzzzzzzzz	王祥	男	01	24	800

表 1-2　部门信息表

部门编号	部门名称	职工人数	部门负责人身份证号
01	财务部	12	330106xxxxxxxxx
02	销售部	36	330201yyyyyyyyy
03	稽核部	3	320101zzzzzzzzz

(3) **属性**

表中的一列称为一个属性，给每一列起一个名称即属性名（如职工年龄）。

(4) **主码（键）**

表中的某个属性组，它们的值惟一地标识一个元组（如职工的身份证号，它惟一地标识一条职工的记录）。

(5) **域**

属性的取值范围（如职工的工资金额在 0 **到** 10 000 **之间，这就是一个域）。**

(6) **关系模式**

对关系的描述，用关系名（属性名 1**，属性名** 2**，…，属性名** n**）来表示。**

2. **关系模型的数据操作**

关系可以进行各种运算，就像算术运算，关系的运算方法称为关系代数，常用的关系操作有两类：

(1) **传统的集合操作**

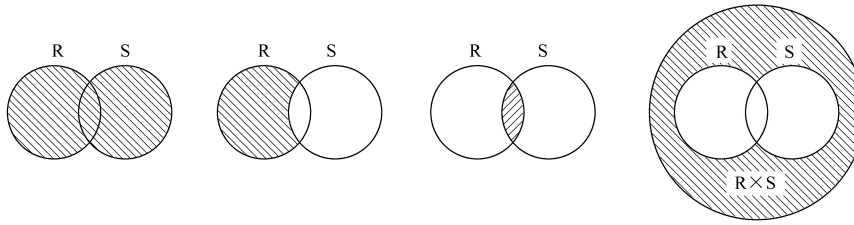
传统的集合操作有并、交、差、广义笛卡儿积。这类操作将关系看成元组的集合，其操作是从关系的“水平”方向，即行的角度来进行的。

传统的集合操作示意图如图 1-10 **所示。表** 1-3 **是传统的集合操作的例子。**

(2) **专门的关系操作**

专门的关系操作有选择、投影、连接。这些操作不仅涉及行，而且涉及列。

关系代数的操作对象是关系，操作结果也是关系。下面介绍几种操作：



(a) 并 (R ∪ S)　(b) 差 (R−S)　(c) 交 (R ∩ S)　(d) 笛卡儿积

图 1-10　传统的集合操作

表 1-3　传统集合操作举例

R

姓名	职称	基本工资
张三	教授	3 009.95
李四	高工	2 980.30
王五	助工	1 900.43

(a)

S

工号	职称	基本工资
李四	高工	2 980.30
赵六	副教授	2 700.00
王五	助工	1 900.43

(b)

R∪S

姓名	职称	基本工资
张三	教授	3 009.95
李四	高工	2 980.30
王五	助工	1 900.43
赵六	副教授	2 700.00

(c)

R∩S

工号	职称	基本工资
李四	高工	2 980.30
王五	助工	1 900.43

(d)

R－S

姓名	职称	基本工资
张三	教授	3 009.95

(e)

R×S

姓名	职称	基本工资	姓名	职称	基本工资
张三	教授	3 009.95	李四	高工	2 980.30
张三	教授	3 009.95	赵六	副教授	2 700.00
张三	教授	3 009.95	王五	助工	1 900.43
李四	高工	2 980.30	李四	高工	2 980.30
李四	高工	2 980.30	赵六	副教授	2 700.00
李四	高工	2 980.30	王五	助工	1 900.43
王五	助工	1 900.43	李四	高工	2 980.30
王五	助工	1 900.43	赵六	副教授	2 700.00
王五	助工	1 900.43	王五	助工	1 900.43

(f)

设关系 R 和关系 S 具有相同数目的属性列（n），并且相应的属性取自同一域。则可以定义下列 7 种操作：

① 并（union）

关系 R 和关系 S 的并，由属于 R 或属于 S 的元组组成。其结果仍为 n 列属性的关系。

② 交（intersection）

关系 R 与关系 S 的交，由既属于 R 又属于 S 的元组组成。其结果仍为 n 列属性的关系。

③ 差（difference）

关系 R 与关系 S 的差，由属于 R 而不属于 S 的元组组成。其结果仍为 n 列属性的关系。

④ 广义笛卡儿积（extended cartesian product）

两个分为 n 列和 m 列的关系 R 和 S 的广义笛卡儿积是一个（n＋m）列元组的集合，元组的前 n 列是 R 的一个元组，后 m 列是 S 的一个元组。若 R 有 k1 个元组，S 有 k2 个元组，则关系 R 和 S 的广义笛卡儿积有 k1×k2 个元组。

⑤ 选择（selection）

选择操作就是在关系中选择满足某些条件的元组。在实际关系数据库产品中，选择操作是在表中选择满足某些条件的行。例如：要“找出基本工资大于 3 000 元的教职工”，就可以在职工基本工资表中做选择操作，条件是“基本工资金额大于 3 000”。

图 1-11 是关系的选择操作示意图。

基本工资表

工号	职称	…	基本工资
1	教授		3 009.95
2	高工		2 980.30
3	助工		1 900.43
4	副教授		2 980.50
5	教授		4 000.00
6	工程师		2 000.43
7	工程师		1 250.34
…			

选择

结 果 表

工号	职称	…	基本工资
1	教授		3 009.95
5	教授		4 000.00

图 1-11 关系的选择操作

⑥ 投影（projection）

投影操作是指在关系中选择某些属性列。例如，要“找出所有职工的姓名，年龄”，则可以对职工基本信息表做投影操作，选择出表中的“姓名”、“年龄”列。

图 1-12 是关系的投影操作示意图。