

数 据 库 系 统

鲁 振 德

东北工学院计算机系软件教研室

一九八四年

序

数据库是近年来发展最为迅速的电子计算机软件，是数据信息管理的最新技术，广泛地用于国民经济、文化教育、军事情报、科学研究、人工智能和计算机辅助设计等领域，为计算机应用开辟了广阔的天地。

数据处理是当今世界计算机应用最活跃、最庞大的应用领域。数据处理泛指非数值计算方面对任何型式数据（信息）的所有计算、管理和加工处理。例如企业管理、库存管理、报表统计、帐目计算、信息情报检索等都可以认为是数据处理。

早期的电子计算机数据处理系统只是一些彼此无关的子系统。如工资、子系统、库存子系统等，每个子系统所带有的数据文件也只包含部分信息。自从六十年代中期磁盘存储器问世以来，产生了管理系统的概念。然而，对大单位管理信息系统的数据文件过多，造成维护上的困难，因而出现了数据管理的最新技术——数据库。它是数据处理的核心机构。

各种不同的数据库系统按所用的数据模型而异。当前最盛行的有三种模型，它们是关系模型、层次模型和网状模型。本着全面系统地介绍数据库的基础知识，本书概述了数据库结构及系统的基本内容，介绍了各种文件的组织方法，这是数据库系统的物理基础。对三种模型及有关的问题都作了比较详细的介绍，由于关系方法是数据库的发展方向，故较全面地叙述关系方法的理论与实现。

本书在周围各位老师帮助下编写的，主要参考如下文献：

1. Ullman: Principles of Data Base systems (1980)
 2. C. J. Date: An Introduction to Date Base systems (1977)。
 3. J. Martin: Computer Data Base Organization
 4. 数据库原理与方法 国防科技大学 郑若忠 王鸿武
 5. 数据库系统基础 美国马利兰大学 姚诗斌
 6. 数据库系统讲义 中国人民大学 萨师煊
 7. 数据库基础 北京工业大学 王连瑞
- 由于编者水平有限，书中不妥之处，请批评指正。

目 录

第一章 概述	1 - 1
1.1 什么是数据库系统	1 - 1
1.2 数据库系统的特点	1 - 2
1.3 数据库系统的典型结构	1 - 6
1.4 数据语言	1 - 7
1.4.1 数据描述语言	1 - 7
1.4.2 数据操作语言	1 - 8
1.5 数据库管理系统	1 - 9
1.5.1 数据字典 (<i>Data dictionary</i>)	1 - 10
1.5.2 数据库管理员 (<i>DBA</i>)	1 - 11
1.5.3 用户访问数据库的过程	1 - 12
1.6 实体——联系方法	1 - 14
1.7 数据模型	1 - 17

第二章 存储结构	2 - 1
2.1 引言	2 - 1
2.1.1 文件的基本概念	2 - 1
2.1.2 数据库操作速度的估计	2 - 4
2.1.3 指示器 (<i>Pointer</i>)	2 - 5
2.1.4 关键字 (<i>Keys</i>)	2 - 5
2.1.5 钉定 (<i>Pinned</i>) 和未钉定的记录	2 - 6
2.1.6 文件结构概述	2 - 7
2.2 顺序文件	2 - 10
2.2.1 如何确定关键字值的顺序	2 - 10
2.2.2 顺序文件的存储组织	2 - 10
2.2.3 顺序文件的查找	2 - 11

2.3	随机结构之一——散列方法	2-12
2.3.1	散列方法的简要回顾	2-12
2.3.2	散列文件的设计	2-15
2.3.3	可扩充的散列	2-16
2.4	随机结构之二——索引结构	2-21
2.4.1	索引顺序文件	2-21
2.4.2	索引无序文件	2-21
2.4.3	索引的组织	2-22
2.4.4	索引文件的查找	2-25
2.5	<i>B</i> -树	2-25
2.5.1	二叉树	2-25
2.5.2	<i>B</i> -树	2-27
2.5.3	<i>B</i> +树	2-29
2.5.4	一个 <i>B</i> +树实例	2-31
2.6	变长记录文件	2-33
2.7	倒排文件	2-35
第三章	关系方法	3-1
3.1	关系及基本术语	3-1
3.2	关系运算	3-3
3.2.1	关系代数	3-3
3.2.2	元组关系演算	3-9
3.2.3	域关系演算	3-12
3.3	关于数据库的数据操作语言	3-14
3.3.1	基于关系代数的语言 <i>ISBL</i>	3-15
3.3.2	介于关系代数与演算之间的语言 <i>SEQUEL</i>	3-19
3.3.3	关于元组演算的语言 <i>QUEL</i>	3-27
3.3.4	基于域演算的语言 <i>QBE</i>	3-32
3.4	关系数据库的模式和子模式	3-37
3.4.1	源模式、目标模式及其物理映射	3-37
3.4.2	子模式、目标子模式及其映射	3-41

3.5	询问的优化·····	3-45
3.5.1	优化的一般策略·····	3-46
3.5.2	关系代数表达式的等价代换规则·····	3-47
3.5.3	关系代数表达式的优化算法·····	3-48
第四章	层次方法·····	4-1
4.1	一般概念·····	4-1
4.1.1	树·····	4-1
4.1.2	层次系统的数据模型·····	4-3
4.1.3	层次顺序与层次路径·····	4-5
4.1.4	层次系统的模式与子模式·····	4-7
4.2	IMS 系统的逻辑结构·····	4-8
4.2.1	IMS 的逻辑结构·····	4-8
4.2.2	IMS 的 DBD·····	4-9
4.2.3	IMS 的 PSB·····	4-12
4.3	IMS 的存储结构·····	4-14
4.3.1	HSAM·····	4-14
4.3.2	HISAM·····	4-15
4.3.3	HIDAM 和 HDAM·····	4-19
4.4	IMS 的数据子语言·····	4-25
4.4.1	子语言 DL/1·····	4-25
4.4.2	IMS 的应用程序·····	4-30
4.4.3	应用程序的运行·····	4-35
4.5	IMS 存储结构补充·····	4-36
4.5.1	辅数据集组·····	4-36
4.5.2	IMS 辅助索引·····	4-38
4.5.3	IMS 的逻辑数据库·····	4-40
第五章	DBTG 建议的网状模型的数据库系统·····	5-1
5.1	DBTG 系统的结构·····	5-1

5. 2	DBTG的数据模型	5 - 2
5.2.1	记录类型	5 - 2
5.2.2	络类型 (Set type)	5 - 3
5.2.3	络事件 (Set occurrence)	5 - 5
5.2.4	事物联系的DBTG表示法	5 - 7
5. 3	记录类型描述及其存储映射	5 - 10
5.3.1	DBTG句法使用的符号	5 - 10
5.3.2	记录类型的描述	5 - 11
5.3.3	记录类型的存储映射	5 - 13
5.3.4	记录类型举例	5 - 16
5. 4	络类型描述及其存储映射	5 - 17
5.4.1	络类型 (Set mode)	5 - 17
5.4.2	络次序 (Set order)	5 - 18
5.4.3	从记录类型性质的描述	5 - 23
5.4.4	络选择 (Set selection)	5 - 24
5.4.5	络类型举例	5 - 25
5. 5	模式数据描述	5 - 29
5. 6	子模式数据描述	5 - 29
5.6.1	子模式与模式的区别	5 - 30
5.6.2	子模式举例	5 - 30
5. 7	数据操作语言 (DML)	5 - 32
5.7.1	程序的运行	5 - 32
5.7.2	DML语句概述	5 - 33
5.7.3	应用程序举例	5 - 38

第六章 关系数据库的规范化理论

6. 1	关系模式的规范化概述	6 - 1
6. 2	关系数据库的设计理论	6 - 5
6.2.1	函数依赖 (Functional Dependency)	6 - 6
6.2.2	计算闭包	6 - 10

6.2.3	依赖集的复盖	6 - 12
6. 3	关系模式的分解	6 - 13
6.3.1	分解的定义	6 - 14
6.3.2	联接的不丢失性 (<i>Lossless join</i>)	6 - 14
6.3.3	联接不丢失性的检验	6 - 14
6.3.4	分解对依赖的保持	6 - 16
6. 4	关系模式的规范化	6 - 17
6. 5	结果为 <i>BCNF</i> 的联接不丢失性分解	6 - 19
6. 6	多值依赖及第四范式	6 - 22

第一章 概述

1.1 什么是数据库系统

数据库(database)是计算机软件的一个重要分支,是近十几年迅速发展起来的一门新兴学科。从本质上讲是大宗,数据的集中存贮。这些数据是被若干特定企业的应用系统所使用。

C. J. Date 从三个方面描述数据库:①存储在磁鼓、磁盘或其他外存介质上的数据集合——这是指数据库自身;②存在以这种数据为背景运行的若干个批应用程序,对其施行通常操作(检索、修改、插入和删除,一般以检索为最通常和最重要),另外可能有一些联机用户从远程终端与数据库相互作用;③数据库是集成的,即包含许多用户的数据,每个用户只享用其中的小部分,且不同用户使用的部分以多种方式重叠——即单独的数据片能够被许多不同用户共享。

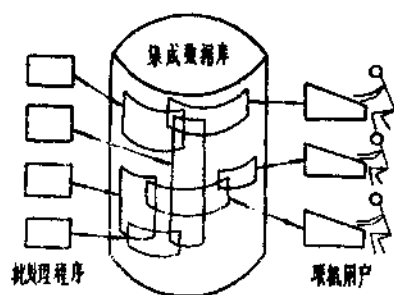


图 1-1 数据库系统示意图

数据库的概念并不是突然出现的,而是由于经验和需要逐渐发展起来的。计算机不仅用于科学计算还大量用于管理,任何企业都必须有一大批围绕其工作的数据,例如制造公司的生产管理和产品供销数据、银行帐目数据、医院的患者病例数据、学校的教学管理数据、政府部门的管理、统计和计划数据等。

设计数据库保存这些数据的目的,不仅仅是为了扩展人们的记忆,而主要是帮助人们去控制与之相关的事物。如图 1-2 所示,人们观测客观事物中得到大量信息,对这些信息进行记录、整理和归类(总称规范),然后将规范信息数据化并送入数据库中保存起来,其中一部分信息可能直接送入控制决策机构。另一方面,控制决策机构(它既可由一些人组成,也可是一个自动控制系统)向数据库发出询问,并利用数据库响应后提供的信息作出决策,再行控制客观事物。譬如,就一个教学信息控制系统,客观事物可指学生、课程、学习情况等,决策控制机构指教学领导机关,策略是提高教学质量的措施。当数据库的管理人员把观测客观事物(学生、课程、

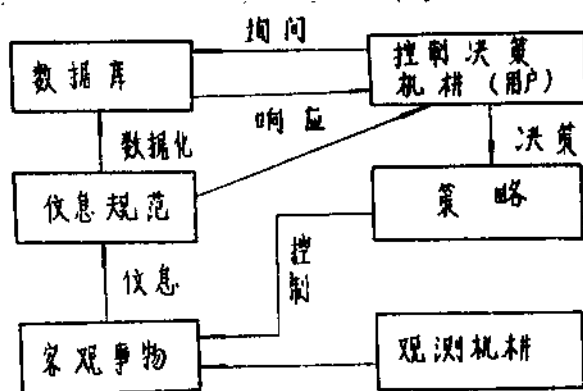


图 1-2 信息控制系统图示

学习情况)得到的信息规范化、数据化并送入数据库后,教学领导机关即可通过询问数据库迅速而准确地得到学生各门学习情况,再参考教学经验,提出改进教学的措施(策略),并在教学中付诸实行(控制客观事物)。

综上所述,数据库可以定义为:它是按一定的组织方式存贮在计算机中相互有关的数据的集合,能以最佳的方式,最少的重复,最大的独立性为多种用户或应用程序服务。而数据库管理系统是指:支持人们建立、查询、修改和删除数据库中数据的软件。数据库和数据库管理系统总和称为数据库系统,它即包括计算机软件资源,也包括硬件资源。

1.2 数据库系统的特点

六十年代以来已经有了计算机的文件系统,当数据大大增加、使用数据的用户越来越多时,早期的文件系统便不能适应更有效地使用数据的需要了,而数据库系统针对文件系统的问题,从概念上向前发展了,主要方面为:

(1) 数据的共享性,尽量减少数据的冗余度(Redundancy)

早期的文件系统是根据应用程序的需要而各自专门建立的,当不同的应用程序所需要使用的数据有许多部分相同时也必须建立各自的文件,就有大量重复的数据分别存放。如:学校中的学生管理上,教务处建立学生的学习方面的文件,医务室建立学生的健康方面的文件,……但这许多文

件中都有学生的：姓名、学号、年龄、出身……，这些数据没有很好共享，且浪费大量的存储空间。更严重的是，存在存储数据的不一致性，即有学生的某方面数据进行修改，同样的这个数据若存在多个文件中，修改起来很困难，且会产生有的文件中此数据没有修改，这是很有害的，大大降低了数据的正确性。

随着用户对数据访问的需要，用户不是以记录或数据项为单位访问数据，而是以文件为单位共享数据。数据库系统对数据在统一的控制之下，同样的物理数据可以导出多个不同的逻辑文件，可以来自多个物理文件，以最少的数据冗余，为尽可能多的用户服务。

(2) 数据的独立性 (data independence)

一般的文件管理系统中，数据 and 应用程序过分地相互依赖，这是因为文件系统完全是根据具体的应用程序的要求而建立的，数据的逻辑结构是对该应用程序优化的，如满足访问速度最快，而采用 hash 存取方法；考虑有大量的顺序访问，文件采取顺序文件等等。文件管理系统中文件的逻辑结构和物理结构是一致的，要想对现有的数据进行新的应用是相当困难的，一旦数据的结构要修改，（如文件由顺序的变为 hash 的，则应用程序也必须相应地进行修改。反之，应用程序改变也将影响到数据结构的改变。（如数据由十进制变为二进制的，怎样存储也要改变），而数据库系统使应用程序与数据结构隔离开，改变数据的物理位置和存储结构不必修改或重写应用程序，只修改模式。用户逻辑数据与他们的物理存储之间的转换由数据管理的软件完成。

(3) 数据的完整性 (Integrity)

完整性问题是保证数据库仅包含正确的数据，数据库系统对存入的数据都进行描述，规定它的类型及值域，对操作数据规定正确性检查。当数据进行修改时，保证数据的特性符合于原来的定义，当输入数据时，检查输入的数据是否符合各项规定，即数据生效性得到保证，集中控制后，不会出现一个职工每周工作 200 小时，人的年龄大于 200 岁，月份是 1-12 之外的整数。

(4) 保密性和安全性 (Privacy and security)

保密性是指用户个人或团体有指定他们自己信息的密级的权力。数据库中的文件及记录都有他们各自的密级，防止数据有意或无意的泄露。避

免数据正确性的破坏。

安全性是指对数据的保护能力，即对数据库中的数据实现存取控制的技术手段。防范利用计算机犯罪。

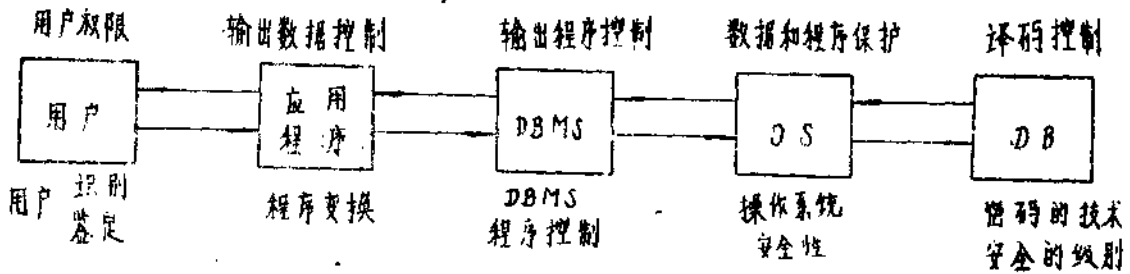


图 1-3 数据安全的模型和技术

计算机安全性的模型和技术如上图。

在五个模型中，我们可以一层层设“防”以实现安全性。

第一块中由“用户识别和鉴定”检查用户的合法使用权。一般用名字号码或用多长时间来辨别，具体办法有：

(1) 用户鉴别：对号码。用机时，用户先报一个与系统约定的号码。经鉴别后，一方面使他取得使用权。另一方面根据号码也便于系统向他结算机费。

(2) 用户证实：号码对了，人对不对？往往窃取一个约定号码不甚困难。所以还要进一步证实是不是这个人？

- ① 口令 (Password)
- ② 一次一个口令，用随机数产生
- ③ 提问题的系统。因为口令易忘，当你上机时问你，这类问题，只有那个人能记住。

④ 符号标记，用一张密码卡片往机器上一放就履行了证实。

⑤ 手指印，指印证实是警察局常用的，这类证实涉及到模式识别。

⑥ 声音辨别，有的机器能对声音，回声音。

用户权限是指某些用户可用那些文件及怎样使用。

① 基本存取特权

a. 只读 (read only)

- b. 读/写
- c. 只能执行(execute only)

② 数据检索和处理限制

- a. 检索(retrieval)
- b. 更新(update)
- c. 删去(delete)
- d. 插入(insert)
- e. 锁(lock)
- f. 解锁(unlock)
- g. 保护(protection)

系统对用户的权限加以控制，如用户无权删除数据，而它的程序中有 *Delete*，进入第二块时，程序变换将它的 *Delete* 数据去掉，使它在限制范围内变换执行。又如：

<pre>Select NAME SALARY FROM EMPLOYEE WHERE STATUS>20 AND EMP* >500</pre>	$\Rightarrow$	<pre>SELECT NAME SALARY FROM EMPLOYEE WHERE STATUS>20 AND EMP* >500 AND SALARY≤50000</pre>
---	---------------	--

原来程序要求选取资格级别超过 20，工号在 500 以后的职员的名字和工资。经变换后，加了一条限制——将原来选取的数据中工资超过 50000 都剔除。经变换后，使某些数据不被访问。第三块中“DBMS 程序控制”不让资格不够的程序用 DBMS 程序。第四块“OS 安全性”拒绝它进入存取子程序。第五块采用“密码技术”把存的资料搞乱，只有掌握键的人才能恢复原状。

类似的过程，当数据一层层由里往外送时，也有种种控制和保护，避免不该拿到的人拿到数据。比如，在终端，我们可将终端分类，资料能否到达终端要看其类别。象一个军事部门有将军用的终端，也有上尉的终端以及小兵的终端。由于一分类别，即使小兵取用将军的资料也不可能在他用的终端上显示出来。因为数据传送在终端上受到控制。

1. 3 数据库系统的典型结构

美国标准委员会 ANSI 的数据库管理系统研究组，在 1975 年公布的研究报告，把数据库分为三级：外模式(external schema)、概念模式(conceptual schema)、内模式(internal schema)。我们把数据库分为用户级、概念级和物理级。其相互关系的描述见图 1-4。

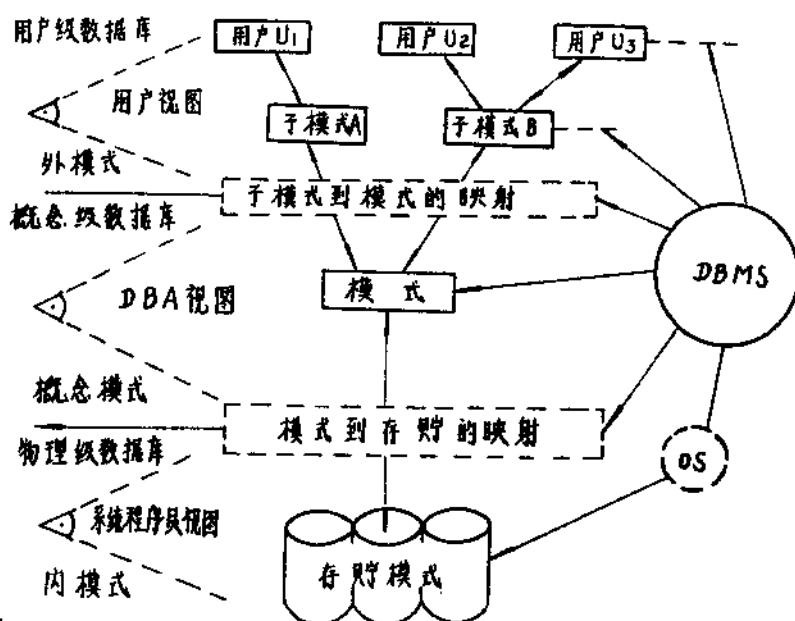


图 1-4 数据库的分级结构

用户级数据库对应于外模式，是用户看到和使用的数据库，因此也称用户视图，在许多文献中又称为子模式，也就是用户看到并获准使用那部分数据的逻辑结构，用户根据系统给他的子模式，用询问语言或应用程序去操作数据库中的数据。

概念级数据库对应于概念模式，简称模式，它是对数据库的整体逻辑描述，即数据库管理员看到的数据库，通常称之为 DBA 视图。模式常是所有用户视图的一个最小并集。设立概念级的目的是为了把用户视图有机地结合成一个逻辑整体，统一地考虑所有用户的要求，它涉及的仍然是数据库中所有的逻辑关系，而不是他们的物理情况。

物理级数据库对应于内模式，又称为存贮模式。它包含数据库的全部存贮数据，这些被存贮在内存、外存介质上的数据也称为原料数据，是用户操

作的对象。这些数据是用一定的文件组织方法组织起来的一个个物理文件。

对一个数据库系统来说，实际上存在的只是物理级数据库，它是数据访问的基础。概念级数据库是物理数据库的一种逻辑描述，用户级数据库是用户与数据库的接口。用户根据子模式进行的操作，通过子模式到模式的映射与概念级联系起来，又通过模式到存贮的映射与物理级联系起来，数据库管理系统(DBMS)的主要工作就是完成三级数据库之间的转换，把用户对数据库的操作转化到物理级去执行。

模式一词没有数据处理方面的涵义，它是数据库的整体逻辑描述，通常也包含有寻址方法、访问控制、保密定义、安全完整量度等内容，但模式的主体是描述数据库的数据模型。模式的表示有图示与语言描述，前者用图示方法画出模式图，非常直观对设计数据库中数据之间的联系很有好处，通常在模式设计第一阶段使用。详细情况后面讨论。而语言描述严格，可以陈述许多细节，通常只在模式设计的第二阶段使用。用语言书写的模式称为源模式，机器不能直接使用，要变为机器使用的模式，称为目标模式。目标模式通常设计成表格形式、树结构或网状结构。

1. 4 数据语言

数据语言包括数据描述语言(DDL—Data Description Language)和数据操作语言(DML—Data Manipulation Language)两大部分，前者负责描述和定义数据的各种特性；后者说明对数据进行的操作。

1. 4. 1 数据描述语言

数据描述语言是数据库设计的重要部分，它描述数据库中各种对象的特征，但最主要的是描述数据。一般地，数据描述语言应具备以下四个方面的功能。

(1) 描述数据的逻辑结构。因为数据库的逻辑描述限于概念级和用户级，所以在数据库设计中又分为模式数据描述语言(Schema DDL)和子模式数据描述语言(Subschema DDL)。前者描述模式，它必须具有定义数据模型的功能和容易阅读的优点，在CODASYL数据库系统中，模式DDL是统一的，可参照它自行设计模式DDL。后者描述子模式。它往往是用户使用的程序语言(如COBOL、PL/1、FORTRAN等)的扩充。在模式

与子模式 DDL 描述数据的逻辑组织中，一般包括如下内容：

① 描述数据模型各个部分的特征：a. 给出数据各个逻辑单位的无二义性命名，如数据模型名、记录类型名、数据项名等；b. 说明各个逻辑单位的数据特征，如数据类型是字母、数字还是字符，数据长度及取值范围等；c. 说明数据单位的自然含意，如数据项是表示姓名还是城市名。

② 描述各数据逻辑单位之间的联系。即数据表示的对象之间和对象内部的联系，通常说明：a. 数据的一个逻辑单位按照什么规则包含哪些更小的逻辑单位，如一个记录类型依次包含那些数据项；b. 那个或那些数据项组合作主关键字使用；c. 数据项之间的完整约束条件；d. 各逻辑单位按什么规则形成一个整体的数据结构，如记录类型怎样构成树或网络。

(2) 描述数据的物理特征。描述的内容包括：a. 系统建立了哪些文件；b. 说明每个物理文件的文件组织方法，如顺序、索引、hash、树结构、倒排结构等；c. 说明每个物理文件的文件名、文件类型、文件的存储设备等。

(3) 描述逻辑数据到物理数据的映射。它说明：a. 每个逻辑单位的数据存放在那个文件中，如那个记录类型对应于那个文件，存放在那个区域；b. 逻辑数据到物理数据的转换，如十进制变二进制等。

(4) 描述访问规则。它包括：a. 用户与子模式的对应关系；b. 用户身份检验，如用户口令、指纹的核对；c. 对用户的授权；d. 数据的保密锁与密码。

上述内容是一些基本内容，各个数据库系统都根据自己情况决定包含那些内容。

1. 4. 2 数据操作语言

数据操作语言是用户与数据库的接口之一，是用户操作数据库中数据的工具。在设计数据操作语言时，一般要做到准确，无二义性；功能齐全，容易掌握，使用方便。一般来说，操作语言不是一种完整、独立的语言，而是一些操作语句组成的集合。大致分为两种：一种是过程式的，操作数据时，不但要指出“做什么”，而且要指出“怎样做”，如嵌入到主语言 COBOL、FORTRAN、PL/1 的一些语句：OPEN、CLOSE、FIND、GET、STORE、INSERT 等；另一种是非过程式的自立式的，操作数据时只须说明“做什么”，至于“怎样做”则留待系统去决定，如 QUEL、SEQUEL 等。二者

功能基本相似，通常包括下几方面操作：

(1) 从数据库中检索数据。这是最重要、最经常使用的一类操作。

(2) 向数据库中添加数据。因为不断记入新数据，有必要把它们存贮到数据库中去。

(3) 删除数据库中某些已经过时的、没有价值的原有数据。

(4) 修改某些属性发生了变化的数据项的值，使之能确切反映变化后的情况。

(5) 用于并发访问控制的操作。

1. 5 数据库管理系统 (DBMS)

DBMS是由许多系统程序所组成，其主要功能如下：

(1) 语言处理方面：① 模式DDL翻译程序，把模式DDL源形式翻译成机器可读的目标形式（一组模式表格）。

② 子模式DDL翻译程序，把子模式DDL源形式翻译成目标形式。

③ DML处理程序，把应用程序的DML语句转换成主语言的一个过程调用语句。

④ 终端询问解释程序，解释每个控制命令的含义，决定怎样执行。

(2) 系统运行控制方面：

① 系统总控程序，它是DBMS的神经中枢，控制、协调DBMS各个程序的活动，使其有条不紊地运行。

② 访问控制程序，其内容核对用户标识、口令，对授权表检验访问的合法性等，决定一个访问是否能够进入数据库。

③ 并发控制程序，在许多用户同时访问数据库时，协调各个用户的访问。例如，按优先级安排访问队列，封锁某些访问或某些数据，撤消某种封锁，允许某个访问或撤消某个事务等。

④ 保密控制程序，在执行操作之前，核对保密规定。例如检查保密级处理密码等。

⑤ 数据完整性控制程序，在执行操作前或后，核对数据库完整约束条件，从而决定是否允许操作执行，或清除已执行操作的影响。

⑥ 数据访问程序，根据用户访问要求，实施对数据的访问。从物理文件中查找数据。执行插入、删除、修改等操作。

⑦ 通讯控制程序，实现用户程序与 DBMS 之间的通讯。

(3) 系统建立、维护方面①数据装入程序，用于把大批原料数据按某种文件组织方法（顺序、索引、hash 等）存贮到内存介质上，完成数据库的装入。

② 工作日志程序，负责记载进入数据库的所有访问。其内容包括：用户名称、进入系统时间、进行何种操作、数据对象、数据改变情况等，使每个访问都留下踪迹。

③ 性能监督程序，监督操作执行的时间与存贮空间占用情况，作出系统性能估算，以决定数据库是否需要重新组织。

④ 重新组织程序，当数据库系统性能变坏时（例如查找时间超过规定值），需要对数据重新进行物理组织。或者按原组织方法重新装入，或者改变原组织方法，采用新的结构。一般来说，重新组织是数据库系统的一种周期性活动。

⑤ 系统恢复程序，当软、硬设备遭到破坏时，该程序把数据系统恢复到可用状态。

1. 5. 1 数据字典 (Data dictionary)

数据库系统运行时所涉及的各种对象：

数据库——一个系统可能容纳多个数据库；

模式——一个数据库可能有多个模式；

子模式——一个模式可能有多个子模式；

记录类型——一个模式可能有多个记录类型；

数据项——一个记录类型可能有多个数据项；

物理文件——一个数据库可能包含多个物理文件；

用户——一个系统有多种用户；

事务——数据处理单位或一件数据处理工作，一个系统包含许多事务，只有最经常使用的少数几个事务才登记在数据字典中；

程序——数据字典中只登记经常使用的少数特定功能的程序；

终端——一个数据库系统可能带有若干联机终端。

数据字典的任务是：

(1) 描述数据库系统的所有对象，确定其属性。如：一个模式包含的记录类型与一个记录类型包含的数据项；一个用户的标识、口令；一个