

数据库系统

及其微型机实用技术

上海电子计算机厂

目 录

第一章 数据库引论	(1)
§ 1. 数据库技术发展概述	(1)
§ 2. 基本概念	(12)
第二章 数据库的组织与技术	(21)
§ 1. 数据信息的处理领域	(21)
§ 2. 数据的独立性	(22)
§ 3. 操作系统的文件管理	(22)
§ 4. 数据库的文件组织	(24)
§ 5. 顺序文件	(25)
§ 6. 索引文件	(27)
§ 7. 散列文件	(28)
§ 8. 树结构	(31)
第三章 层次型数据库及实现技术	(45)
§ 1. 层次型数据库IMS的简述	(45)
§ 2. IMS系统的数据存取	(49)
§ 3. IMS数据语言DL/I	(59)
第四章 网络型数据库及实现技术	(62)
§ 1. CODASYL概述	(62)
§ 2. CODASYL数据库的数据结构	(63)
§ 3. CODASYL的逻辑结构	(65)
§ 4. CODASYL数据描述语言	(68)
§ 5. CODASYL数据库实现技术	(73)

第五章 分布式数据库系统及实现技术.....(78)

§ 1. 集中式数据库系统和分布式数据库系统的特点(78)

§ 2. 分布式数据库系统的分类.....(81)

§ 3. 分布式数据库管理系统.....(83)

§ 4. 分布式数据库模式的构成技术.....(86)

§ 5. 分布式数据库网络结构技术.....(88)

§ 6. 分布系统的查询处理.....(91)

§ 7. 分布存取控制技术.....(92)

第六章 关系型数据库系统及实现技术.....(98)

§ 1. 关系型数据库系统的概述.....(98)

§ 2. 关系型的数学概念与方法.....(102)

§ 3. 关系型数据库的模式和子模式.....(110)

§ 4. 关系型数据库的数据操作语言.....(113)

§ 5. 关系型数据库的实现.....(127)

§ 6. 小结.....(144)

第七章 微型机数据库管理程序 MDBMP 的分析与解

释.....(145)

§ 1. 主控清单(菜单)程序base'.com.....(145)

§ 2. 格式定义程序 base'.1(148)

§ 3. 数据输入程序 base'.2(162)

§ 4. 数据输出程序 base'.3(175)

§ 5. 图形输出程序 base'.4(185)

§ 6. 数据检索程序 base'.5(190)

§ 7. 分类程序 base'.6(199)

§ 8. 运算程序 base'.7(209)

§ 9. 文件操作程序 base'.8(217)

第八章 微型机数据库管理程序 (MDBMP) 的特点

与使用.....(223)

§ 1. MDBMP的设计概要.....(223)

§ 2. MDBMP数据库的文件结构.....(225)

§ 3. MDBMP系统的支持环境及移植.....(228)

§ 4. MDBMP的功能特点与使用方法.....(229)

第九章 微型机关系数据库系统 DBASE II(246)

§ 1. DBASE II的发展.....(246)

§ 2. DBASE II的性能规范.....(247)

§ 3. DBASE II中的表达式.....(249)

§ 4. DBASE II命令的分类与功能.....(260)

§ 5. DBASE II的文件.....(267)

§ 6. 命令符号的意义.....(271)

§ 7. 全屏幕操作.....(274)

§ 8. DBASE II的使用方法.....(276)

附录：主要支持dBASE II的国产微型计算机简介

.....(280)

第一章 数据库引论

本章要点：①回顾数据库的发展过程，从数据库技术的特点上介绍数据库发展的各阶段概况，同时简要介绍数据库的现状与展望。②比较了文件系统和数据库的特点。③介绍数据库中一些基本概念，为学习掌握后面的内容做一些准备工作。

读者应通过本章，着重分析数据库发展中各时期的特点及今后的趋势。正确理解和掌握基本概念的意义。

§1 数据库技术发展概述

随着计算机技术的发展，计算机系统从用于军事和科学计算扩展到应用数据处理各个领域。1945年，第一台电子计算机ENIAC的诞生，是出于军事目的，主要用来计算弹道轨迹。那时注重于计算速度，而对数据存储没有更多的要求。1950年由J. P. Eckert和J. W. Mauckley 教授设计的UNIVAC—I电子计算机系统配置了磁带机这个快速大容量外部存储器，为计算机数据处理创造了条件。当时美国人口调查局用它进行人口调查数据处理。从此，计算机数据处理逐渐引起人们的关注。但50年代数据处理主要是独立设计的简单子系统，到60年代，以共享数据资源为目标的数据集中存储IDS (Integrated Data Store) 与数据库概念才出现。

对于20多年来数据库的发展有人把它分成三个阶段或时期。

1—1 数据库的萌芽期

这一时期为60年代初到70年代初,是数据库概念的产生,探讨、研究时期。初期,对数据库(data base)这一概念,人们的理解并不十分确切。由于base一词在军事术语中是基地的意思,所以当时人们把data base理解为“数据的提供基地”。直到60年代中期,数据处理一直是以共享文件方式(如管理信息系统MIS)和集成数据存储系统(IDS),这也是现代数据库的前身。其特征是包罗了诸文件形式的特点,具有调用控制等数据控制概念。

1967年以研究计算机语言为主名为CODASYL的国际组织成立了数据库任务组(DBTG),从事数据库的研究工作。该组织的数据库研究工作,是扩充COBOL程序设计语言的结果,通常把这种以一种程序设计语言为基础,来构成系统的方式称为主语言(host Language)型。另一方面,为了使用户不必了解内部的存储结构,就可以方便地管理大量的数据。采取了不用以前程序设计语言的形式,而以新的独立的非顺序系统为方向,即独立语言(Belf-Contained),这种数据库系统多为在线应用。

在萌芽期末, MIS (Management Information System)比较盛行,主要以信息检索为方面;对于数据库萌芽时期的主要事件可归纳为如下几点:

(1) 用DOS方式在磁盘上实现了COBOL结构,并以共享多个COBOL程序的数据库系统为目标,其最初成果被称为“DBTG69”的语言设计方案。

(2) IBM1969年推出了依据Mc Gee层次结构的MIS。至

今还在使用的象MARKIV、TOTAL等商用程序包，就是这个时期产生的。

(3) 在出现了在线系统的同时，开发了共享以文件为主的信息检索目录的TDMS和SYSTEM2000，并开始研制AD-ABAS系统。

(4) 在理论研究方面，1962年就出现了信息代数(Information Algebra)，从1968年D.Chilels的STDS来看，这方面工作没有进一步展开，缺乏对数据库基础的研究。

1—2 数据库的成长期

在70年代，随着计算机硬系统的性能价格比的不断提高，各计算机还开始配置了CODASYL方式数据库需要的编译程序，这样，数据库技术得到了急速发展。另一方面，商用程序包得到了更广泛的应用，以关系模式为中心的基础研究也大有进展，数据库成了流行系统，在与信息检索和数据存储关联的各领域都开始引用，这个时期数据库成为信息系统不可缺少的基础技术。

这个时期数据库技术有以下几个特点：

(1) CODASYL方式

CODASYL方式是以1969年DBTG69为基础，经修改充实1971年提出了DBTG71语言设计方案，决定了以COBOL为语言的数据库系统通用语言。

CODASYL方式是面向程序设计的，不太适应一般的用户。DBTG也发现对用户来说需要简便的问答式及自然语言为使用工具。因此1980年CODASYL发表了适于一般用户的功能语言设计方案。

(2) 关系方式的研究

数据库成长时期，是关系方式的发展时期。在1970年IMB的E.F.Codd提出了以数据的关系模式为基础研究数据库的理论后，引起了人们对关系方式的重视。实践证明，把关系方式作为研究数据库系统的基础是极为有效的。

关系方式的数据库系统对用户来讲更方便，就象用户从使用汇编语言到使用高级程序设计语言，程序设计从依赖计算机物理结构到脱离计算机的物理结构。使用关系方式的数据库系统，就不再需要考虑如何编制和调用文件了。

1—3 数据库的成熟期

1981年以后，数据库系统逐渐成熟起来。商品化的实用数据库系统不断增多，各个领域普遍开始应用数据库系统。关系数据库系统作为主要应用的数据库系统开始流行。各种专用数据库系统(如CAD、CAM数据库系统等)广泛应用。

1—4 数据库系统与文件系统的比较

人们在没有计算机的年代里就采用文件形式处理大量的数据信息，当人们借助计算机处理数据时，首先采用的方式就是文件形式，即建立计算机文件系统。数据库就是从文件系统演变发展而成。文件系统与数据库系统相比，在数据量迅速增多，共享数据资源的用户越来越多时，就不能有效地使用数据了，这主要是因为：

(1) 随着数据的增多，数据的冗余度(Redundancy)也增大。这是由于文件系统多是根据应用程序的需要而独自建立的，即使不同的应用程序需要使用的数据有许多相同的部分，也必须重复建立在各自的文件中。大量的重复数据，影响了数据的共享。而且不仅浪费了大量的存储空间，也使得数据修改变得十分困难，常常造成数据的不一致，从而降低

了数据的可靠性和正确性。这是文件系统的主要问题。

(2) 文件系统中的数据缺乏独立性(independency)。因为文件系统主要依据于具体应用程序的要求而建立的。数据的逻辑结构,是在具体应用程序的限制下进行优化的;而且,数据存储的物理结构与其逻辑结构是一致的。因此要想改变数据的某些形式或某些应用意义是相当困难的。一旦要修改某些数据的结构,则应用程序也必须进行相应的修改。反之,应用程序的某些改变,也将影响到相应数据结构的变化。这样使得数据与应用程序相互依赖。

(3) 文件系统对数据的控制缺乏规格化的模式和统一的方法。对数据的正确性,安全性、保密性等缺乏有效的管理手段。而且,编制相应的应用程序十分繁琐,需要有相当水平的程序员来完成。

在不断改进和完善文件系统过程中,产生了数据库。数据库出现的主要目的就是实现数据在最大范围内的共享,即使数据在统一的控制下,为尽可能多的应用项目 和 用户 服务。在数据库中,数据相对于应用程序有比较大的独立性,使得数据的重复现象大幅度减少。同时,应用程序尽可能地不依赖数据的物理结构和存储介质的种类,对数据结构的影响也比较小。而且,修改数据时,也不要求应用程序有很大的改变。

在数据库技术中,由数据库管理系统(DBMS)进行对数据的定义,建立、检索、修改等操作,并完成对数据的完整性、安全性、保密性的统一控制和管理,使用户更有效地使用数据库中的各种数据。

与文件系统相比较,数据库由于进行统一管理和各子模

式的转换等方面，CPU(或主机)的开销(overhead)要大一些。因此，在小规模数据处理或有固定格式的数据处理时，文件系统又显得高效而简便。

1—5 数据库发展中的新课题

数据库的发展需要在现有的基础上取得新的突破，换句话说就是有一些新课题急待解决。

1. 基础理论的研究

早期数据库的开发，不是依靠可靠的理论基础；主要是经验的结果。数据库中的各种关联关系与概念在设计时，靠图示来构造和描述。如图1.1-1。

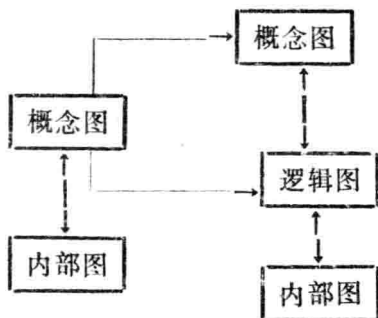


图1.1-1 数据库设计图示

概念图是对客观世界的描述；

逻辑图是决定的逻辑结构；

内部图是对特定系统的内部逻辑设计。

这样描述是不严密的，没有可靠的理论保证。

关系式数据库系统只建立在严格的关系数学基础上，具有严密的理论基础。但是，如何用关系模型来描述客观世界，决定数据结构，进行逻辑设计是需要研究的一大课题。

2. 数据库机器

数据库机器是指具有专门应用于数据库的结构 的计算机。这种用硬件(或专机)实现数据库管理系统构成数据库机器的研究,是又一研究课题。硬件的联想存储部件,为关系式数据库机器创造了有利条件。因为对于关系式数据库来说,表元素的序号,行和列的项目与位置是无关紧要的,对于用户的所有信息都可作为表的值记录下来,给出关系名、属性名,就可取到对应的实际数据。随着大规模集成电路技术的发展,以关系模型为特征的数据库机器形成了新的研究领域。

目前,正在研究和已经开发的数据库机器可以分成两类:

① 双机型。用通信线路把同一结构的计算机与通用主计算机连接起来,主计算机把所有与数据库有关的操作都交给专用机去完成。→ADABAS机器就这一是类型,如图1.1-2所示。它是使用了 两台相同结构的计算机,即使其中一台发生故障,另一台也能代替进行工作。对此,硬件上不更多地花费工夫,而期望软件性能不断提高。

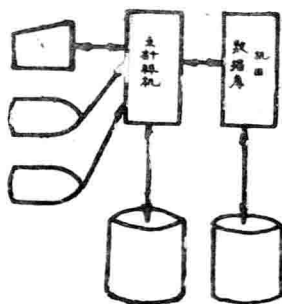
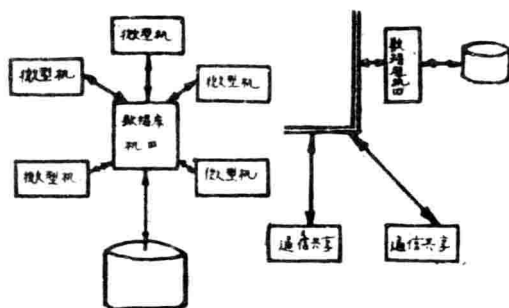


图1.1-2 双机型数据库机器

② 数据库共享型。在构成计算机网路中，取代了以前文件共享的功能。象IDM 500(200)的Shared Data base System及IDBP就是这一类。请见图1.1-3。



(a) IDM200

(b) IDBP

图1.1-3 数据库共享型数据库机器

这一类数据库机器一般是多个微型机共享数据库资源，与在线系统相似。但是，其中核心计算机（数据库机器）是数据库专用机，其终端是平常使用的微型计算机。

3. 分布式数据库系统

分散的计算机资源由计算机网络把其从物理上结合起来，而分布式数据库系统是在把物理结合的计算机网路的数据，逻辑地结合成数据库。

对于大规模的计算机网络，建立数据库是十分重要的。分布式数据库系统的主重课题是，在构成的计算机网上，系统的组成方式。商用数据库机器也是朝此方面发展。但是，微型计算机如何承担数据库共享的功能，是至今没有解决的课题。

4. 数据辞典

数据辞典 (data dictionary) 是指与数据库中数据相

关的数据。也就是说，把数据元整理成易于人们理解的形式。在操作系统内与数据索引表相对应。特别是数据库的规模更大，其属性增多，这一能力是必不可少的。在关系数据库系统中，数据辞典作为关系表成为系统的一部分。

在数据辞典中，主要记载了数据名、属性、数据间的相互关系、数据的使用者及使用数据的程序等，另外，还应考虑，数据的一致性（意义）的限制，分散数据的位置等。

现在，美国的规格协会在1980年设置了X 3 H 4 (Information Resource Dictionary System)，着手进行数据辞典的标准化工作。

5. 处理数据的多样化

目前，研究和开发的数据库，主要是用于事务处理方面。其操作对象多是文字数据，即使有一些图象和声音方面的操作，也都是经过人们用文字加以描述的代码。这些二次信息无疑对人们来说很不方便，而且更重要的是程度不同地破坏了原信息的真实度。因此，将来的数据库系统要能直接进行图象、声音和数值等操作，即处理数据的多样化。

6. 知识库系统

知识库系统在一定形式上是结构化的，对于存储大量称作知识信息元的知识库，以利用这些信息元完成问题解决的装置——推理机构为基本功能模块构成系统。

现在从事知识库系统研究的组织和个人，在国外已有不少。随着人工智能和第五代计算机的研究开发，知识库系统将在理论和实践上成为现实。

§2 基本概念

为了读者易于理解后面的内容，本节就几个常见的基本概念加以讲解。

2—1 数据库的定义

数据库(data base)是六十年代以后发展起来的一门新兴学科，是计算机技术的一部分，是计算机应用的有效手段和方法，是计算机软件工程中的一个重要分支。目前数据库的概念、原理和方法还处于研究、发展和不断变化的阶段，正由工程实践向理论化过渡。因此，对数据库的定义也不尽相同。下面就常见的几种定义归结如下。

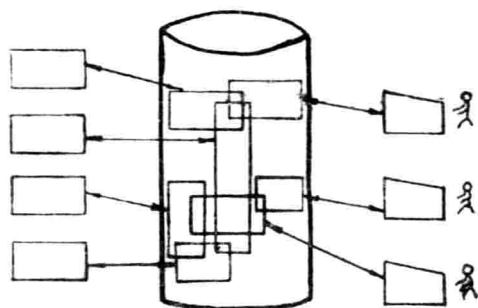
(1) 数据库任务组DBTG的定义：数据库是由一个特定模式(Schema)控制的所有记录，网络(Set)和区域组成。在有多个数据库的情况下，每一个数据库都必须有独立的模式，并假定不同数据库的内容是无关的。

(2) C.J. Date 从三个方面描述了数据库：

① 存储在外部存储介质上的数据集合，即数据库本身；

② 存在使用这些数据的若干个应用程序，可灵活地进行常规操作(检索、修改、插入和删除等)，还可供联机用户从远程终端与数据库相互作用；

③ 数据库是由许多用户数据组成的，每个用户都可用数据库中的数据，而且不同用户可以用多种方式重叠使用数据，即单独的数据片能够为许多不同的用户共享(如图 1.2-1)。



批处理程序

联机用户

图1.2-1 数据库系统简图

(3) J. Martin给出的定义:数据库是存储在一起的相关数据的集合,其中没有危害性数据并尽量削除不必要的冗余数据,能满足多种应用的需要;数据不依赖于使用它的程序而独立存在;可以按一种公用的可控制的方法对数据库进行操作(如插入,修改和检索数据等);数据已经被规格化和结构化。若在某系统中,存在结构上完全分离的多数据库时,则称该系统为一个数据库集合。

上述定义从不同角度说明了数据库的组成和特点,但这些解释重点还不够明确。我们体会数据库应包括如下几项主要内容:

- ①信息高度共享的相关数据的集合;
- ②具有可以控制和利用的合理存储结构和多种数据流路径;
- ③数据的存储独立于使用数据的方式和要求。

2—2 逻辑记录与物理记录

首先请读者看一个小例子。如表1.2—1所示是一本书的几项主要特征，即关键词。只要知道其中一项或几项特征，就可找到那类或那本书籍。

表 1.2—1 图书的主要特征表

书名	数据库设计
作者名	金 旭
出版日期	1983.3
出版社名	科学出版社
书号	TP 36
类别	计算机软件

其中每一项数据(如书名)都可以且仅直接表示事物(书)的一个特征，这样的数据常称为数据项。赋予书名并与一个事物(如图书)关联的若干个数据项(如书名，作者名，出版日期等)的组合，通常叫做逻辑记录，或简称为记录。如表1.2—2所示。在数据组织中，对于文件一级或在数据库的逻辑结构中，逻辑记录是基本数据单位。若用树结构来表示表1.2—2的记录例子，可以得到图1.2—2。

一个逻辑记录在实际存储的介质中，大多不是占据一系列连续的物理存储空间。而且，主存与外设交换数据的长度，一般是由系统程序员根据具体计算机系统的主外存储器应用情况来决定，这样每次把主、外存交换数据的基本单位称为一个物理记录。例如，外存是磁盘时，控制部件可寻址

表1.2—2

一个记录示意

记 录 内 容	注 释
记录名	图书 A
书 号	TP36
书 名	数据库设计
作者名	金 旭
出版社	科学出版社
出版日期	1981.5
版 本	初 版
单 价	单价/精 5.30元
	单价/简 3.75元
版本(再版)	修订版
再版日期	1983.3
单 价	精 5.80元
	简 4.15元

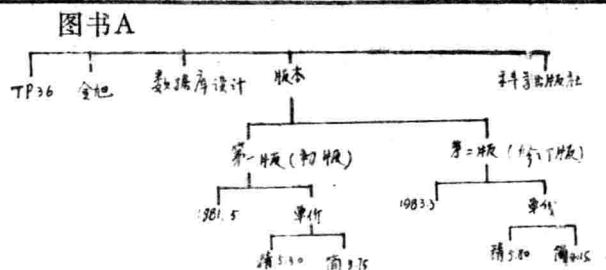


图1.2-2 用树结构表示的记录例子

(存取)的最小单位为块(即物理记录), 块的长度一般为256、512、1024、2048、4096 个字节等多种。在进行数据存取