

万丈高楼平地起——数据库基础

导学

随着大数据、云计算、物联网、人工智能等信息技术的飞速发展,数据资源量急速增长,数据库产业对于我国现代化发展至关重要。在《大数据库看改革开放新时代》一书中采用的数据分析技术,就充分地展示了改革开放的伟大成就。这就要求大学生不断提升管理、存储、处理数据的岗位能力,以及组织、利用、优化资源的职业技能。那么如何利用数据库管理系统科学有效地组织、优化、管理、维护和共享信息系统中的海量数据呢?本单元从数据库的基本知识入手,探究数据库应用的关键技术,培养大国工匠精神和科技强国理念。

预习本单元,思考以下问题。

- (1) 什么是数据库?什么是数据库管理系统?
- (2) 数据库是由哪几部分组成的?各自有什么作用?
- (3) 什么是数据库的模式、外模式和内模式?它们各自有什么特点?

本单元的学习任务

从宏观上对数据库的概念、数据库的体系结构、数据库的主要技术进行简单而全面的了解。

- (1) 理解数据库的概念、特点及其发展;
- (2) 掌握数据库是由哪几个部分组成的;
- (3) 掌握数据库系统的体系结构。

1.1 数据库技术概述

1.1.1 数据库的基本概念

在信息时代,由于数据量的急剧增长,需要利用计算机快速准确地处理和加工大量的数据,于是产生了数据库和数据库管理系统。数据库技术是一门综合性的软件技术,随着

计算机应用的不断发展,在计算机应用领域中,数据处理越来越占据重要地位,数据库技术的应用也越来越广泛。下面先来了解一下数据库的基本概念。

1. 什么是数据

数据(Data)是数据库中存储的基本对象,是用于描述事物的符号记录。这种描述事物的符号都可以数字化形式存入计算机。

数据的表现形式有多种,可以是数字、文字、图形、图像、声音等。

2. 什么是数据库

数据库(Database, DB)是长期存放在计算机内、有组织的、可共享的数据集合。

数据库中的数据是按照一定的数据模型进行组织、描述和储存的,具有较小的冗余度、较高的数据独立性和易扩展性,并可为各种用户所共享,可以形象地理解为存储数据的仓库。数据库是数据表及相关操作对象的集合,而数据表则由一个或多个相关的数据项组成。

数据库技术具有存储、管理大量数据和高效检索的优势,广泛应用于我们日常生活中的各个领域,包括超市、银行、学校、网站、电信、航运、企业、政府机构等。

例如,在某图书超市中,有 20 多名员工,经营着数万本图书。该超市用图书销售系统进行销售管理,而该图书销售系统提供了图书的购买入库、查询、销售及汇总等功能。从管理者角度来看,该系统大大节省了管理维护图书的时间和费用,减轻了账务处理的负担;而从购书者的角度来看,也大大缩短了其寻找书籍的时间。

3. 什么是数据库管理系统

数据库管理系统(Database Management System, DBMS)是位于用户与操作系统之间的一套数据管理软件,它属于系统软件,为用户或应用程序提供访问数据库的方法,包括数据库的建立、查询、更新及各种数据控制 and 操作。

DBMS 就像一个大管家,负责数据库中所有的对内、对外操作,如图 1-1 所示。

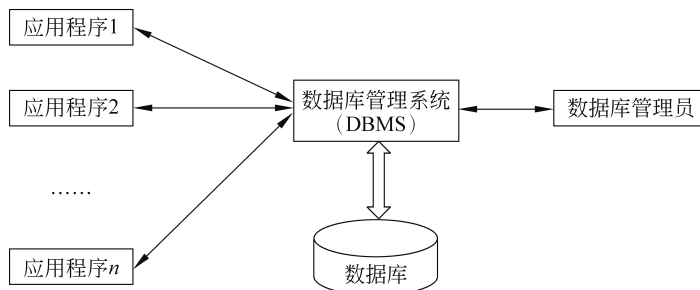


图 1-1 数据库管理系统图示

数据库管理系统具有如下功能。

(1) 数据定义功能。DBMS 提供数据定义语言(Data Definition Language, DDL),用

户通过它可以方便地对数据库的相关内容进行定义,如对数据库、基本表、视图和索引等进行定义。

(2) 数据操纵功能。DBMS 向用户提供数据操纵语言(Data Manipulation Language,DML),可实现对数据库的基本操作,如对数据库中数据的查询、插入、删除和修改。

(3) 数据库的运行管理。这是 DBMS 的核心部分,它包括并发控制、安全性检查、完整性约束条件的检查和执行、数据库的内部管理(如索引、数据字典的自动维护)等。所有数据库的操作都要在这些控制程序的统一管理下进行,以保证数据安全性、完整性和多个用户对数据的并发操作。

(4) 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,数据库的转储、恢复功能,数据库的重新组织功能和性能监测、分析功能等。这些功能通常是由一些实用程序完成的,它是数据库管理系统的一个重要组成部分。

(5) 数据字典。数据字典(Data Dictionary,DD)是存放与数据库各级模式结构相关的描述,也是访问数据库的接口。在大型系统中,数据字典也可以单独成为一个系统。

(6) 数据通信功能。包括与操作系统(Operating System,OS)的联机处理、分时处理和提供远程作业传输的相应接口等,这一功能对分布式数据库系统尤为重要。

4. 什么是数据库系统

数据库系统(Database System,DBS)是一个实际可运行的存储、维护和为应用系统提供数据的软件系统,是存储介质、处理对象和管理系统的集合体。

1.1.2 数据库技术的发展

数据管理是对数据进行组织、存储、加工及维护的过程。随着计算机技术的不断发展,数据管理技术经历了人工管理、文件系统和数据库系统三个阶段,目前的主流技术是数据仓库技术。

1. 人工管理阶段

在 20 世纪 50 年代中期以前,数据管理主要由人工完成,那个时候计算机中没有专用的软件对数据进行管理,也没有磁盘之类的存储设备用来存储数据。由于应用程序和数据之间是一对一关系,即一个程序对应一组数据,而在程序设计过程中,不仅需要规定数据的逻辑结构,且还要定义数据的物理结构,一旦当数据的物理组织或存储设备发生改变时,则必须重新编写应用程序,因此存在很多问题,包括程序间不能共享数据;应用程序和数据之间依赖性太强,独立性差;数据大量冗余,难以保证应用程序之间的数据一致性等。

人工管理阶段的程序与数据间的关系如图 1-2 所示。

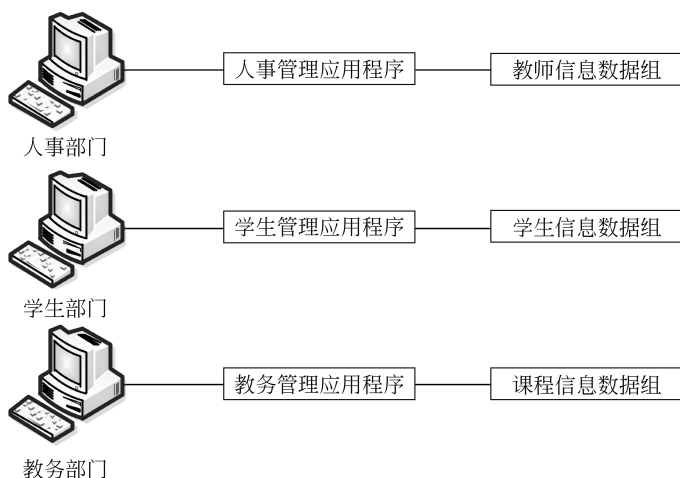


图 1-2 人工管理阶段的程序与数据间的关系

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期,随着计算机软硬件技术的飞速发展,出现了专门管理数据的软件,也就是文件系统。文件系统有多种形式,包括顺序文件、索引文件和随机文件等。

在文件系统数据管理阶段,数据会按一定的规则组织成为一个文件,应用程序通过文件系统对文件中的数据进行存取和加工。这样,程序和数据就实现了分离,数据可以长期保存在外围设备上,而且数据的逻辑结构和数据的存储结构之间有一定的独立性,实现了以文件为单位的数据共享。但是此时仍然还存在着不少问题,包括不同的应用程序很难共享同一个数据文件,数据的独立性仍然较差,数据冗余度较大,数据的一致性差等。

文件系统阶段程序与数据间的关系如图 1-3 所示。

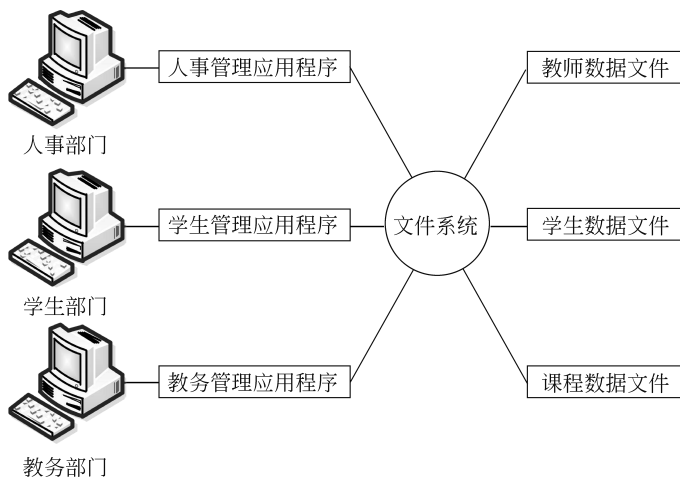


图 1-3 文件系统阶段的程序与数据间的关系

3. 数据库系统阶段

20 世纪 60 年代后期,为了满足巨大的信息流和数据流的需要,数据库系统出现了。数据库是在数据库管理系统的集中控制之下,按一定的组织方式存储起来的、相互关联的数据集合。在数据库系统中,应用程序具有高度的独立性,同时,数据库系统对保证数据的完整性、唯一性和安全性都提供了一套有效的管理手段,从而减少了开发和维护应用程序的费用。另外,数据库系统还提供管理和控制数据的各种简单操作命令,用户在编写程序时会更容易上手。总之,数据库系统阶段具有如下特点:

- (1) 数据的共享程度高,冗余度低;
- (2) 数据真正实现了结构化。

数据库系统阶段的程序与数据间的关系如图 1-4 所示。

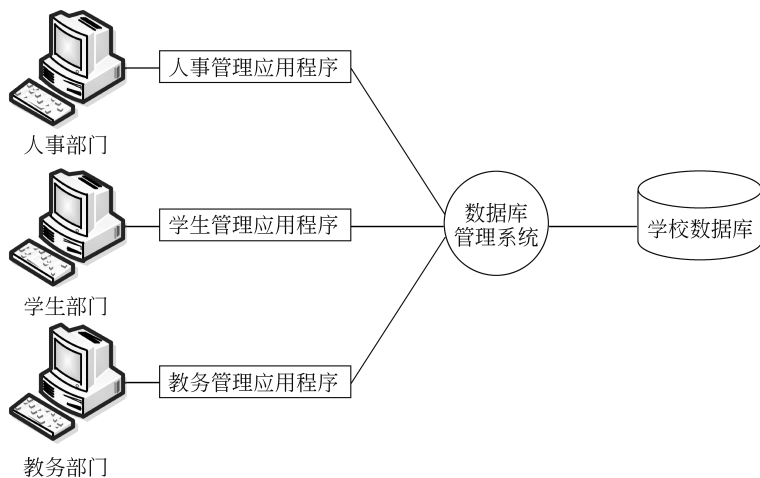


图 1-4 数据库系统阶段的程序与数据间的关系

1.1.3 数据库系统的特点

数据库系统不仅实现了多用户共享同一数据的功能,还解决了由于数据共享而带来的数据完整性、安全性及并发性控制等一系列问题。同时,数据库系统也克服了文件系统中存在的数据冗余大和数据独立性差等缺陷,而且实现了数据与程序之间的独立。

1. 查询迅速且准确

与手工操作相比,数据库系统在查询数据时迅速且准确,而且可以节省大量纸张。以一个大型仓库管理系统为例,用手工操作,如果要查找“某厂家生产的商品的名称,规格,数量”时,就可能要翻阅大量账本,费时且费力。而在使用数据库系统时,由于数据是由 DBMS 按一定的结构组织存放在计算机中,因此用户可以迅速查找所要求的数据,而且很少出现错误。

2. 数据结构化且统一管理

数据库中的数据是有结构的,并且由 DBMS 统一管理。DBMS 既管理数据的物理结构,也管理数据的逻辑结构,即考虑数据之间以及文件之间的联系,可见 DBMS 管理的是结构化的数据。

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本质区别。在数据库系统中数据的结构用“数据模型”概念加以描述,而数据的最小存取单位是“数据项”。

3. 数据冗余度小

在文件系统中,当一个应用程序在处理专用的一个或几个数据文件时,会有许多重复的数据,即产生了数据冗余。这是由于文件系统中的数据重复存储,且由不同的应用程序使用和修改所造成的。数据库系统是从整体上即从全局角度看待和描述数据,数据不仅面向某个应用,还面向整体应用,从而大大减少了数据冗余,节省了存储空间,避免了数据之间的不一致性。所谓数据之间的不一致是指同一数据在不同存储位置的值不一样。

4. 具有较高的数据独立性

数据独立性是指用于应用程序与存储在磁盘上的数据库中数据的相互独立性,也就是说,数据在磁盘上的数据库中的存储是由 DBMS 管理的,所以数据对应用程序的依赖程度大大降低,数据和程序之间具有较高的独立性。

数据库中的数据独立性一般可以分为两级。

1) 数据的物理独立性

应用程序要处理的只是数据的逻辑结构也就是数据库表中的数据,这样当数据在计算机存储设备上的物理存储发生改变时,应用程序可以不必改变,而由 DBMS 来处理这种改变,称为物理独立性。

2) 数据的逻辑独立性

有的 DBMS 还提供一些功能,使某些程序上数据库的逻辑结构虽然发生了改变,但用户程序可以不改变,称为逻辑独立性。

数据独立性是数据库的一种特征和优点,它有利于当数据库结构出现变动时应用程序可以尽可能地不做改变或少做改变,这样就大大减少了应用程序开发人员的工作量。

5. 数据的共享性好

在数据库应用中,数据是共享的,这不仅使某些应用程序的编写更加方便,而且冗余度小,节省了存储空间,避免了数据间的不一致性,系统易维护、易扩充。

6. 数据控制功能强

为了适应共享数据的环境,DBMS 还提供了数据控制功能。数据库系统提供了四个方面的数据控制功能,即数据库的并发控制、数据库的恢复、数据的安全性和数据的完整性。

1) 数据库的并发控制(concurrency)

良好的并发控制可以使在多用户同时存取或修改数据库时,防止相互干扰而提供给用户不正确的数据,并避免数据库受到破坏。

2) 数据库的恢复(recovery)

当数据库被破坏或数据不可靠时,系统有能力将数据库从错误状态恢复到最近某一时刻的正确状态。

3) 数据的安全性(security)

所谓数据的安全性是指保护数据以防止不合法的使用所造成数据的泄露和破坏。例如,系统提供口令检查或其他手段来验证用户身份,或对数据的存取权限进行限制等。

4) 数据的完整性(integrity)

所谓数据的完整性是指数据的正确性、有效性和相容性。

(1) 数据的正确性是指数据的合法性。

(2) 数据的有效性是指数据是否在其定义的有效范围。

(3) 数据的相容性是指表示同一事实的两个数据应相同。

1.2 数据库系统的组成



1.2 数据库系统组成

数据库系统是一种按照数据方式存储、管理数据并向用户或应用系统提供数据支持的计算机应用系统,它通常包括支持数据库系统的计算机硬件、存储数据的数据库、操纵数据的应用软件以及开发者、管理者和使用者。

图 1-5 所示为数据库系统结构示例。

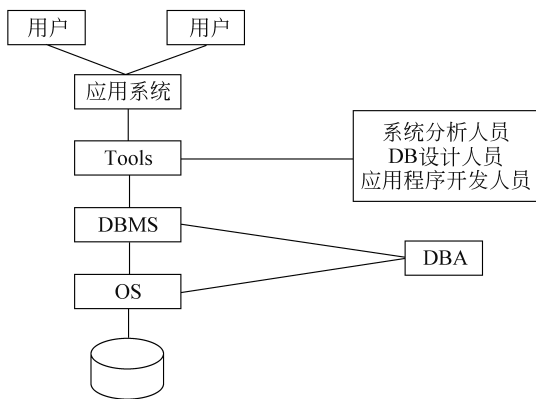


图 1-5 数据库系统结构示例

1. 硬件

计算机硬件是数据库系统建立的基础,数据库系统在必需的硬件资源支持下才能工作。计算机硬件是存储和运行数据库系统的硬件设备,主要包括 CPU、内存、大容量的存储设备、外部设备等。

数据库系统数据量大,数据结构复杂,软件内容多,因此要求计算机硬件的数据存储容量大,数据处理速度和数据输入输出速度要快。

2. 数据库

数据库是与应用程序彼此独立并以一定组织方式存储在一起彼此相互关联具有较少冗余且能被多个用户共享的数据集合。它存储和运行在计算机硬件基础上。现实中的数据库需要通过数据库管理系统来创建和管理。

3. 软件

数据库系统中的软件包括一系列的管理和开发软件。

1) 操作系统

操作系统是所有计算机软件的基础,在数据库系统中起着支持 DBMS 和主语言系统工作的作用。

2) 数据库管理系统

数据库管理系统是实现对数据库的定义、建立、维护、使用及控制的系统软件,是整个数据库系统的核心。通常把数据库管理系统直接称为数据库,常用的大型 DBMS 有 Oracle、SQL 等,小型 DBMS 有 Access 等。

3) 主语言系统

主语言系统是为应用程序提供的诸如程序控制、数据输入输出、功能函数、图形处理、计算方法等数据处理功能的系统软件。

4) 应用程序

数据库应用程序是指通过 DBMS 访问数据库中的数据并向用户提供服务的程序,简单地说,它是允许用户插入、删除、修改和访问数据库中数据的程序。应用程序由程序员通过程序设计语言或某些软件开发工具(如 Visual C++、Java 等),并按照用户的要求编写的。

5) 应用开发工具

应用开发工具是 DBMS 系统为应用开发人员和最终用户提供的高效率、多功能的应用生成器,包括第四代计算机语言等各种软件工具,如报表生成器、表单生成器、查询和视图设计器等。这些开发工具为数据库系统的开发和使用提供了良好的环境和帮助。

硬件、数据库与软件的层次结构如图 1-6 所示。



图 1-6 硬件、数据库与软件的层次结构

4. 相关人员

数据库系统的人员在整个数据库系统中扮演着系统开发者、管理者和使用者的角色,包括系统分析员、数据库设计人员、数据库管理员、应用程序员和用户等。

(1) 系统分析员(System Analyst,SA)。系统分析员是在大型、复杂的信息系统建设任务中,承担分析、设计和领导实施的领军人物。系统分析员的职责包括应用系统的需求分析和规格说明;与用户以及 DBA 配合,确定系统的软件和硬件配置;参与数据库的概要设计,规划系统的数据库模型等。

(2) 数据库设计人员。数据库设计人员要按照需求分析和总体设计的框架,合理、有效、科学、安全的设计数据库结构,定义各个表结构、存储过程、触发器等,确定数据库中的数据,并设计系统数据库的各级模式。

(3) 数据库管理员(Database Management,DBA)。数据库管理员的核心目标是保证数据库管理系统的稳定性、安全性、完整性和高性能。其职责覆盖产品从需求设计、测试到交付上线的整个生命周期,具体职责包括决定数据库中的信息内容和结构;决定数据库的存储结构和存储策略;定义数据的安全性要求和完整性约束条件;监控数据库的运行情况和使用情况;以及进行数据库的改进和重组等。

(4) 应用程序员。应用程序员负责编写系统主程序,实现系统功能,并负责应用程序的安装和调试。

(5) 用户。用户是使用数据库系统的人员。

1.3 数据库系统体系结构

1.3.1 数据库系统三级模式

从 DBMS 的角度看,数据库系统有一个严谨的体系结构,从而保证其功能得以实现。根据 ANSI/SPARC(美国标准化协会和标准计划与需求委员会)提出的建议,数据库系统是三级模式和两级映像结构的,通过数据库三级模式的划分可以使不同类型的数据库系统人员以不同的视图看待数据库中的数据,如图 1-7 所示。



1.3 数据库系统结构

1. 模式

模式(Schema)也称逻辑模式(Logical Schema),是对数据库全局逻辑结构的描述,是所有用户的公共数据视图即全局视图,又称概念模式或概念视图。

一个数据库只有一个模式,它是由数据库设计者综合所有用户数据,按照统一的观点构造而成的。例如,对于一个学生的记录可定义为(学号、班级、姓名、性别、出生日期、家庭住址、联系电话、电子邮箱),称为记录型,而('12010101','网络 2001','赵丽','女','2001-02-03','河北省保定市','13565412300','zhl@126.com')则是该记录型的一个记录值。模式只是对记录型的描述,与具体值无关。

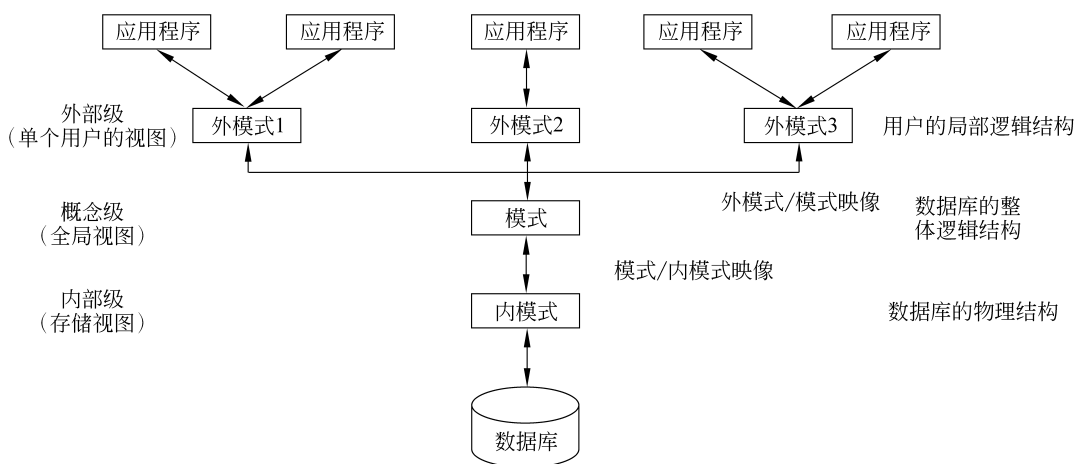


图 1-7 数据库系统的三级模式与两级映像结构

DBMS 提供了模式描述语言(Data Description Language, DDL)来定义模式。在定义模式时,不仅要定义数据的逻辑结构,而且要定义数据之间的联系,定义与数据有关的安全性、完整性要求等。

模式是数据库系统模式结构的中间层,既不涉及数据库物理存储细节和硬件环境,也与具体的应用程序以及所使用的程序设计语言或应用开发工具无关,它由 DBA 统一管理,故又称 DBA 视图。

2. 外模式

外模式(External Schema)也称子模式(Subschema)、用户模式(User Schema)或外视图,是用户观念下局部数据结构的逻辑描述,常把外视图称为用户数据视图。

外模式通常是模式的子集,一个数据库可以有多个外模式。外模式是指完全按照用户对数据的需要,站在局部的角度进行数据库设计。即使是来自模式中的同样的数据,在外模式中的结构、类型、长度、保密级别等都可以有所不同。另外,同一外模式也可以为某一用户的多个应用系统所使用,但一个应用程序只能使用一个外模式。

外模式使用外模式描述语言进行定义,该定义主要涉及对外模式的数据结构、数据域、数据构造规则及数据的安全性和完整性等的描述。

3. 内模式

内模式也称存储模式(Storage Schema)、物理模式(Physical Schema)或内视图,是数据物理结构和存储方式的底层描述,是数据在数据库内部的表示形式。

一个数据库只能有一个内模式。从形式上看,一个数据库就是存放在外存储器上的许多物理文件的集合。例如,记录的存储方式是堆存储,还是按照某个属性的升序或降序存储,或是按照属性值聚簇存储,数据是否压缩、是否加密等。

内模式使用内模式描述语言定义。内模式描述语言不仅能够定义数据的数据项、记