

模块一 数据库创建

任务 1 数据库技术基础知识



任务描述

本任务将带领大家深入数据库的最基层，让大家对数据库技术基础知识有更深入的了解，从而更好地掌握数据库技术基础知识，将数据库技术应用到平时的工作、生活中，将会带来极大的方便。

数据库技术是一门综合性的软件技术，是使用计算机进行各种信息管理的必备知识。数据库技术是 20 世纪 60 年代开始兴起的一门信息管理自动化的学科，是计算机科学中的一个重要分支。随着计算机应用的不断发展，在计算机应用领域中，数据处理越来越占主导地位，数据库技术已经成为各领域各种业务数据管理的重要工具和最新技术，与计算机网络、人工智能被称为计算机技术界的三大热门技术。



知识重点

- (1) 熟悉数据、数据管理和数据库的基本概念。
- (2) 掌握数据库技术的特点、应用及发展趋势。
- (3) 熟悉数据库系统的组成及数据库的体系结构。



知识难点

- (1) DBMS 的工作模式、主要功能及其组成。
- (2) 数据模型与概念模型的理解。

子任务 1.1 数据与数据库的基本概念

【子任务描述】

随着 IT 技术的快速发展和广泛应用，数据库技术的应用已经从事务处理扩展到计算机网络服务、商务智能、计算机辅助设计和决策支持系统等新领域，各行业大量的重要数据需要经过数据库才能进行有效组织、存储、处理和共享。通过学习数据库有关技术知识，可以为今后的业务学习和就业奠定重要基础。

【子任务实施】

熟练掌握数据、数据库、数据管理与处理和数据库系统等基本概念，对数据库技术和后续知识及内容的学习极为重要。

一、信息和数据的概念

1. 信息

信息 (Information) 通常被认为是有一定含义的、经过加工的、对决策有价值的信息。例如,“2014 年全市高校新生人数为 8 万人”是一条信息,而“全市”、“2014”、“年”及“8”等只是数据。数据表示信息,而信息只有通过数据形式表示出来才能为人所理解。

2. 数据

数据 (Data) 是信息的表达方式和载体,是人们描述客观事物及其活动的抽象表示,是描述事物的符号记录。它是利用信息技术进行采集、处理、存储和传输的基本对象,数据的概念包括描述事物特性的数据内容和存储在某一种媒体上的数据形式。

数据的概念包括两方面含义:数据的含义是信息,数据的表现形式是符号。通常,数据分为数值数据和非数值数据两大类,可以是数字、文字、符号、图形、表格、图像、声音、录像、视频等形式。数据是数据库中存储与管理的基本对象,人们收集并抽取需要的大量数据之后,将其保存起来,经过进一步加工处理,从而得到有用信息。

数据库中的数据具有的特性如下。

(1) 全局性:数据库中的数据都是从全局观点出发建立的,按一定的数据模型(即结构,详见子任务 1.2)进行组织、描述、存储、管理和控制。

(2) 共享性:数据库中的数据是为多用户共享建立的,已经摆脱了具体程序的限制和制约,不同的用户可以按各自的需求和用法使用数据库中的数据。

3. 信息与数据的区分

数据与信息既有区别又互相依存。数据是信息的具体表示形式和载体,信息反映了数据的含义。数据是数据库管理的基本内容和基本对象,是信息的一种符号化表示方法,采用一定的符号表示信息,而具体用哪种形式的符号及表示方式则是人为规定的。

信息来源于数据,数据是信息的具体表现形式,信息以数据的形式存储、管理、传输和处理,数据经过处理后可得到更多有价值的信息。信息是概念性的,数据是物理性的。信息可用数据的不同形式来表示,数据的表示方式可以选择,而信息不随数据表现形式而改变。

二、数据库与数据库管理系统

1. 数据库

数据库 (Database, DB) 指的是以一定方式存储在一起、能为多个用户共享、具有尽可能小的冗余度、与应用程序彼此独立的数据集合。

数据库中的数据是从全局观点出发建立的,按照一定的数据模型进行组织、描述和存储。按数据管理类型来划分,数据库主要分为层次数据库、网状数据库和关系型数据库,目前应用最多的是关系型数据库。

2. 数据库管理系统

数据库管理系统(Database Management System, DBMS)是一种操纵和管理数据库的软件,它对数据库进行统一的管理和控制,以保证数据库的安全性和完整性。用户通过 DBMS 访问数据库中的数据,数据库管理员也通过 DBMS 进行数据库的维护工作。它可以使多个应用程序和用户用不同的方法在同时或不同时刻建立、修改和查询数据库。DBMS 是整个数据库系统的核心,对数据库中的各种数据进行统一管理、控制和共享。

数据库管理系统为用户提供以下几个主要功能。

(1) 建立数据库功能:DBMS 通过相应的操作语言实现对采集的数据的组织与存储。

(2) 数据操纵功能：根据用户的需求，对数据库中的数据进行修改、删除、插入、检索、重组等操作。

(3) 数据库的控制与维护功能：通过对数据库进行有效的控制、分析与监视，实现数据的完整性、安全性及并发控制与数据恢复。

(4) 数据的网络化：通过数据库的操作语言产生数据网页，实现数据的网络查询、修改等功能，并实现数据与其他管理系统数据格式的转换功能，最大限度地实现数据共享。

三、数据库技术的应用及特点

1. 数据库技术应用

随着 IT 技术的快速发展，数据库技术的应用从数据处理与管理，扩展到计算机辅助设计、人工智能、决策支持系统和计算机网络应用等新领域。在现代信息化社会，由于信息无处不在，所以，数据库技术的应用非常广泛，遍布各个领域、行业、业务部门和各个层面。网络数据库系统及数据库应用软件已成为信息化建设和应用中的重要支撑性软件产业。

数据库技术在以下几种行业中得到了广泛的应用。

(1) 销售业：存储、查询供应商、商品、客户及销售信息和商品的网上订购等。

(2) 金融业：用于银行客户的信息、账户、贷款及银行的交易记录。还可以用于存储股票、债券、金融票据的持有、出售和买卖等交易信息。

(3) 制造业：用于产供销存储产品的订单、产品原料的供应情况，跟踪产品的产量及仓库产品的详细清单。

(4) 电信业：用于存储通信网络的信息，存储通话记录，存储用户付费业务记录及产生每月通信账单，以及交费情况等。

(5) 航空业：用于存储、查询、网络订购国内外各种航班和票务信息。

(6) 教育系统：用于存储教职员的信息，存储工资、津贴和纳税的信息，产生工资单，存储学生信息、课程及实验信息、成绩信息和大学生科创信息等。

数据库技术是数据管理的最新技术，给广大用户的业务发展和生活带来了极大便利。例如，通过网络查询信息、预订机票、网上购物和付费等，数据库的应用更加广泛。

随着信息技术的快速发展，数据库技术也产生了一些新的应用领域，主要如下。

(1) 云数据库：将数据库应用于地理信息系统 (GIS) 和计算机辅助设计与制造 (CAD/CAM)，与地球上的空间位置相关的空间数据是 GIS 的重要组成部分，而设计数据则是 CAD/CAM 的核心。

(2) 多媒体数据库：多媒体数据库主要存储与多媒体相关的，如声音、图像和视频等数据。多媒体数据最大的特点是数据量大、数据类型多且数据类型间差距较大，因此，需要较大的存储空间和较复杂的数据管理。

(3) 信息检索系统：信息检索系统是一种典型的联机文档管理系统，一直与数据库技术同步发展。

(4) 决策支持系统：联机分析处理 (OLAP) 是数据库系统的主要应用，它支持复杂的分析操作，侧重决策支持，并且提供直观易懂的查询结果。可使分析管理人员或执行人员从多角度对信息进行快速、一致、交互地存取，获得对数据更深入地了解。

(5) 移动数据库：移动数据库是在移动计算机系统上发展起来的，其最大特点是通过无线数字通信网络进行传输，用户可以随时随地访问和获取数据，为一些商务应用和应急情况带来了极大的便利。

2. 数据库技术的主要特点

1) 控制数据的安全性和完整性

针对数据库进行的各种操作必须根据操作者所拥有的权限进行鉴别,鉴别机制由 DBMS 提供,每个用户的操作权限设定则由数据库管理员(DBA)负责,以保障数据库应用系统及数据的安全性、机密性和完整性。

2) 数据独立冗余低

数据独立性是指数据库存储的数据与处理数据的应用程序互相独立,避免了在传统的数据库处理应用系统中,应用程序与相关业务数据关联,致使各种业务数据在多种不同的数据文件中分别存储,数据大量冗余且无法统一更新等问题的发生。数据库技术可对所有数据集中管理,并利用有效地数据共享功能,不再需要各项业务单独保存各自的数据文件,极大地减少了数据冗余。

3) 数据高度集成

数据处理应用系统中的数据源于多项业务,且数据之间相互关联。如在一个商品供销管理信息系统中,进货数据来源于供货管理、销售数据来源于售货管理、员工数据来源于人力资源管理。对这些数据进行集中管理,保持它们的正确关联,才能完成所需的综合数据处理。利用数据库技术和 DBMS 提供的数据库管理功能可实现多种数据的集成。

4) 应用程序开发与维护效率高

在应用程序开发时,数据的独立性可不必考虑软件和数据关联问题,以及所处理的数据组织等问题,减少了应用程序的开发与维护的工作量。只在应用系统开发初期,需要规划数据库、设计数据库中的各个数据集、规范数据库中相关数据间的关联。只有一个满足规范化设计要求的数据库,才能够真正实现各类业务不同的应用需求。

5) 数据广泛共享

在一个数据库应用系统中,可对集中管理的数据进行共享。如供货管理需要参考商品销售管理系统中近期的销售数据,确定进货种类与数量,确定销售单价时又需要参照最近的进货单价等,利用数据库技术通过计算机网络可实现数据广泛共享。

6) 实施统一的数据标准

数据标准是指数据库中数据项的名称、数据类型、数据格式、有效数据的判定准则及要求等数据项特征值的取值规则。

7) 保证数据一致性

数据一致性是指存储在数据库中不同数据集(表)的相同数据项必须具有相同的值。一个数据库由多种数据文件组成,数据文件之间通过公共数据项相联系,当对一个数据文件中的数据项进行更新时,相关联文件中的对应数据项也必须自动更新,才能始终保持数据库中数据的一致性和正确性。通过 DBMS 可以自动实现对数据库中数据进行追加、插入和删除等操作时的一致性问题。

四、数据管理技术的产生和发展

数据库技术随着数据应用和需求的变化而不断发展。数据处理是指对各种数据进行收集、存储、加工和传播的一系列活动的总和。数据处理的目的是从大量的、原始的数据中获得所需要的资料并提取有用的数据成分,作为行为和决策的依据。数据管理则指对数据进行分类、组织、编码、存储、检索和维护,它是数据处理的中心问题。随着电子计算机软件和硬件技术的发展,数据处理过程发生了划时代的变革,而数据库技术的发展,又使数据处理跨入了一个崭新的阶段。

数据管理技术的发展大致经历了以下 3 个阶段。

1. 人工管理阶段

世界上第一台计算机“ENIAC”的诞生初期,应用计算机面临的一个重要问题就是数据的存储。当时的计算机将数据以打孔的方式存储在纸带上,既不容易检索也不容易修改。它以电子管为主要元器件,主要依靠硬件系统,包括运算器、控制器、存储器和简单的输入输出设备,工作效率极低,只能计算并输入输出很少的数据。

人工管理阶段的计算机主要应用于科学计算,绝大部分的数据管理基本上以是手工方式,用纸卡及报表等进行一记载、存储、查询和修改。当时,外存没有磁盘等直接存取的存储设备,也没有操作系统和数据库。

人工管理阶段的数据具有以下几个特点。

(1) 数据不保存。由于当时计算机主要用于科学计算,数据保存上并不做特别要求,只是在计算某一个课题时将数据输入,对数据不做保存。

(2) 数据不独立。数据是输入程序的组成部分,即程序和数据是一个不可分割的整体,数据和程序同时提供给计算机运算使用。对数据进行管理,如现在的操作系统可以以目录、文件的形式管理数据。程序员不仅要知道数据的逻辑结构,也要规定数据的物理结构,程序员对存储结构、存取方法及输入输出的格式有绝对的控制权,要修改数据必须修改程序。

(3) 数据不共享。数据是面向应用的,一组数据对应一个程序。不同应用的数据之间是相互独立、彼此无关的,即使两个不同应用涉及相同的数据,也必须各自定义,无法互相利用、互相参照。数据不但高度冗余,而且不能共享。

(4) 由应用程序管理数据。数据没有专门的软件进行管理,需要应用程序自行进行管理,应用程序中要规定数据的逻辑结构和设计物理结构(包括存储结构、存取方法、输入/输出方式等),因此程序员负担很重。

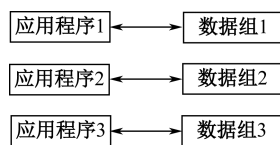


图 1.1 人工管理阶段应用和数据文件关系图

人工管理阶段应用和数据文件之间的关系如图 1.1 所示。

2. 文件管理阶段

20 世纪 50 年代,计算机以晶体管取代了运算器和控制器中的电子管,由于存储介质的更新,数据以文本文件或二进制文件的形式存储;可将成批数据单独组成文件存储到外部存储设备,出现了操作系统、汇编语言和一些高级语言。计算机不限于科学计算使用,还大量用于管理等,在操作系统中有专门的数据管理软件,称为文件系统,是数据库系统发展的初级阶段,并非真正的数据库系统。

文件管理阶段把有关的数据组织成一种文件,这种数据文件可以脱离程序而独立存在,由一个专门的文件管理系统实施统一管理。文件管理系统是一个独立的系统软件,它是应用程序与数据文件之间的一个接口。在这一管理方式下,应用程序通过文件管理系统对数据文件中的数据进行加工处理。应用程序的数据具有一定的独立性,比手工管理方式前进了一步。但是,数据文件仍高度依赖于其对应的程序,不能被多个程序通用。由于数据文件之间不能建立任何联系,因而数据的通用性仍然较差,并且冗余量大。

文件管理阶段的数据具有以下几个特点。

(1) 数据长期保留。数据可以长期保留在外存上反复处理,即可以经常有查询、修改和删除等操作,所以计算机大量用于数据处理。

(2) 数据的独立性。由于有了操作系统,可利用文件系统进行专门的数据管理,程序员可以集中精力在算法设计上,而不必过多地考虑细节。例如,要保存数据时,只需给出保存指令,

而不必所有的程序员都精心设计一套程序,控制计算机物理地实现保存数据。在读取数据时,只要给出文件名,而不必知道文件的具体存放地址。文件的逻辑结构和物理存储结构由系统进行转换,程序与数据有了一定的独立性,数据的改变不一定要引起程序的改变。

(3)可以实时处理。由于有了直接存取设备,也有了索引文件、链接存取文件、直接存取文件等,所以既可以采用顺序批处理方法,也可以采用实时处理方式,数据的存取以记录为基本单位。

文件管理阶段应用和数据文件之间的关系如图 1.2 所示。

文件管理阶段比人工管理阶段有了很大的改进,但随着数据管理规模的扩大,数据量急剧增加,文件系统的缺陷显现出来,主要表现为如下。

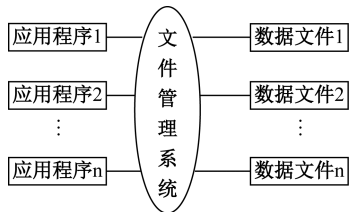


图 1.2 文件管理阶段应用和数据关系图

数据共享性差,冗余度大。

由于数据文件之间缺乏联系,造成每个应用程序都有对应的数据文件,这样相同的数据会在多个数据文件中重复存储。

数据和程序缺乏足够的独立性。

由于数据冗余,在进行更新操作时,若不注意,就容易使相同的数据在多个数据文件中不一致,如修改时遗漏或错改数据。

数据联系弱。

文件之间相互独立且缺乏必要的关联,影响了数据管理。

3. 数据库系统管理阶段

从 20 世纪 60 年代中期以来,随着计算机软硬件技术的快速发展,CPU 向超大规模集成电路发展,为存储和处理大数据量的数据库给予了极大的技术支持。同时,操作系统得到了发展,而且各种 DBMS 软件不断涌现,使得数据库管理技术不断发展和完善,成为计算机领域中最具影响力和发展潜力、应用范围最广、成果最显著的技术之一,形成了“数据库时代”。数据库系统管理阶段指对所有的数据实行统一规划管理,形成一个数据中心,构成一个数据库,数据库中的数据能够满足所有用户的不同要求,供不同用户共享。在这一管理方式下,应用程序不再只与一个孤立的数据文件相对应,而是可以取整体数据集中的某个子集作为逻辑文件与其对应,通过数据库管理系统实现逻辑文件与物理数据之间的映射。

在数据库系统管理的环境下,应用程序对数据的管理和访问灵活方便,而且数据与应用程序之间完全独立,使程序的编制质量和效率都有所提高。由于数据文件之间可以建立关联关系,因此数据的冗余量大大减少,数据共享性显著增强。

数据库管理阶段的主要特点如下。

1) 数据结构化

数据结构化是数据库系统与文件系统的根本区别。在文件系统中,相互独立的文件的记录内部是有结构的,传统文件的最简单形式是等长同格式的记录集合,这样可以节省许多存储空间。数据的结构化是数据库主要特征之一,这是数据库与文件系统的根本区别。

2) 数据共享性高,冗余小

数据库从整体的观点来看待和描述数据,数据不再面向某一应用,而是面向整个系统。这样就减小了数据的冗余,节约了存储空间,缩短了存取时间,避免数据之间的不相容和不一致。对数据库的应用可以很灵活,面向不同的应用,存取相应的数据库的子集。当应用需求改变或

增加时,只要重新选择数据子集或者加上一部分数据,便可以满足更多更新的要求,即保证了系统的易扩充性。

3) 数据独立性高

数据库提供数据的存储结构与逻辑结构之间的映像或转换功能,使得当数据的物理存储结构改变时,数据的逻辑结构可以不变,从而使程序也不用改变,这就是数据与程序的物理独立性。也就是说,程序面向逻辑数据结构,不考虑物理的数据存放形式。数据库可以保证数据的物理改变不引起逻辑结构的改变。

4) 数据统一进行管理和控制

数据库管理系统对所有数据统一进行管理和控制,保证了数据的安全性和完整性。数据库系统对访问用户身份及其操作的合法性进行检查,自动检查数据的一致性、相容性,保证数据符合完整性约束条件,以并发控制手段有效控制多用户程序同时对数据操作,保证共享及并发操作,恢复功能保障,当数据库遭到破坏时能自动恢复到正确状态。

数据库管理阶段应用和数据文件之间的关系如图 1.3 所示。

4. 高级数据库管理阶段

从 20 世纪 80 年代以后,数据库技术在商业领域取得巨大成功,激发了其他领域对其需求的快速增长。数据库技术新的应用领域的研究,极大地推动了数据库技术,特别是面向对象数据库系统的研究和发展。同时,它不断与其他技术结合,向高级数据库技术发展。

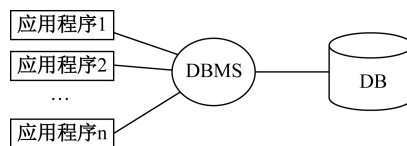


图 1.3 数据库管理阶段应用和数据文件关系图

1) 分布式数据库系统

分布式数据库系统是在集中式数据库系统成熟技术的基础上发展起来的,但不是简单地把集中式数据库分散地实现,它是具有自己的性质和特征的系统。集中式数据库系统的许多概念和技术,如数据独立性、数据共享和减少冗余度、并发控制、完整性、安全性和恢复等在分布式数据库系统中都有了不同之处及更加丰富的内涵。分布式数据库系统具有以下几个特点。

(1) 数据独立性。

数据独立性是数据库方法追求的主要目标之一。在集中式数据库中,数据独立性包括两方面:数据的逻辑独立性与数据的物理独立性,其含义是用户程序与数据的全局逻辑结构及数据的存储结构无关。

在分布式数据库中,数据独立性这一特性更加重要,并具有更多的内容。除了数据的逻辑独立性与物理独立性外,还有数据分布独立性(亦称分布透明性)。用户不必关心数据的逻辑分区、数据物理位置分布的细节,也不必关心局部场地上数据库支持哪种数据模型。当数据从一个场地移到另一个场地时不必改写应用程序,当增加某些数据的重复副本时也不必改写应用程序。数据分布的信息由系统存储在数据字典中,用户对非本地数据的访问请求由系统根据数据字典予以解释、转换、传送。

(2) 集中与自治相结合的控制结构。

在集中式数据库中,为了保证数据库的安全性和完整性,对共享数据库的控制是集中的,并设有 DBA 负责监督和维护系统的正常运行。在分布式数据库中,数据的共享有两个层次:一是局部共享,即在局部数据库中存储局部场地上各用户的共享数据;二是全局共享,即在分布式数据库的各个场地中也存储可供网中其他场地的用户共享的数据,支持系统中的全局应

用。因此,分布式数据库系统常常采用集中和自治相结合的控制结构,各局部的 DBMS 可以独立地管理局部数据库,具有自治的功能。同时,系统又设有集中控制机制,协调各局部 DBMS 的工作,执行全局应用。

当然,不同的系统集中和自治的程度不尽相同,有些系统高度自治,有些系统则集中控制程度较高,场地自治功能较弱。

(3) 适当增加数据冗余度。

在集中式数据库中,尽量减少冗余度是系统目标之一,其原因是冗余数据浪费存储空间,而且容易造成各副本之间的一致性。而为了保证数据的一致性,系统要付出一定的维护代价,减少冗余度的目标是用数据共享来达到的。而在分布式数据库中却希望增加冗余数据,在不同的场地存储同一数据的多个副本,其原因如下。

提高系统的可靠性、可用性:当某一场地出现故障时,系统可以对另一场地上的相同副本进行操作,不会因一处故障而造成整个系统的瘫痪。

提高系统性能:系统可以根据距离选择离用户最近的数据副本进行操作,减少通信代价,改善整个系统的性能。

(4) 全局的一致性、可串行性和可恢复性。

分布式数据库中各局部数据库应满足集中式数据库的一致性、可串行性和可恢复性。此外,还应保证数据库的全局一致性、并行操作的可串行性和系统的全局可恢复性。因为全局应用要涉及两个以上节点的数据,因此在分布式数据库系统中一个业务可能由不同场地上的多个操作组成。

分布式数据库系统兼顾分布、集中管理两项任务,因而具有良好的性能,其具体结构如图 1.4 所示。

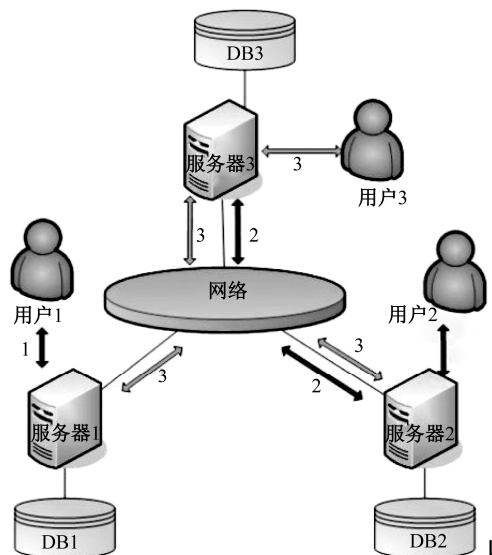


图 1.4 分布式数据库系统

2) 面向对象数据库技术

在数据库中提供面向对象的技术是为了满足特定应用的需要。随着许多基本设计应用(如 MACD 和 ECAD)中的数据库向面向对象数据库的过渡,面向对象思想也逐渐延伸到其他涉及复杂数据的应用中,其中包括辅助软件工程(CASE)、计算机辅助印刷(CAP)和材料需求计划(MRP)。这些应用如同设计应用一样在程序设计方面和数据类型方面都是数据密集型的,

它们需要识别于类型关系的存储技术，并能对相近数据备份进行调整。面向对象数据库技术主要有两个特点。

对象数据模型能完整地描述现实世界的数据结构，表达数据间嵌套、递归的联系。

具有面向对象技术的封装性和继承性的特点，提高了软件的可重用性。

3) 面向应用领域的数据库技术

数据库技术经过几十年的发展，形成完善的理论体系和实用技术。为了适应应用多元化的需求，结合各应用领域的特点，将数据库技术应用到特定领域，产生了工程数据库、地理数据库、统计数据库、科学数据库、空间数据库等多种数据库，也出现了数据仓库和数据挖掘等技术，使数据库领域中的新技术不断涌现。

最新的 SQL Server 2012 实现了一个为云做好准备的信息平台，可以有效解决日益增加的数据量带来的挑战，以帮助用户管理任意大小、本地或云端中的任何数据。通过提供的数据平台和工具，用户可以提取更有价值的信息，从而做出有效决策。

SQL Server 2012 与以前版本相比，更具扩展性、更可靠，提供了更高的性能。另外，它还包含了 Power View 功能，微软称“这是行业领先的商业智能功能”，该功能通过强大的交互能力，可以将用户对于任何地方、任何数据的探索，转变成一种更加自然、轻松的体验。此外，通过 Windows Azure 中的一项基于 Apache Hadoop 的服务，可以连接 SQL Server 并整合商业智能工具。

5. 数据库技术的发展趋势

随着经济的快速发展和信息网络日新月异的进步，功能强大的数据库系统在电子商务、商业智能、在线交易处理、内容管理中发挥着重要作用。数据库技术已经表现出明显的发展趋势，更多的理解功能赋予数据库系统，并具备对冷、热两种存储中的物理数据管理的能力。此外，数据库技术还将向着实用性更强、可扩展性更强、更加人性化、智能化的方向发展。

自数据库技术从产生至今，在社会对其研究的不断加深，以及实践应用的逐渐广泛，不论是在理论基础和实际应用，还是在商业产品方面，都已形成了较为完善的体系。同时，数据库技术也因其自身的强大功能，逐渐成为一项广泛研究的技术。但是，伴随着各种新技术和新方法的不断涌现，也给数据库技术带来了一定的冲击，为确保数据库技术的地位，研究人员不断对其进行优化和创新，研发了很多创造性的数据模型，使数据库技术不断多样化。面向专业应用领域的数据库技术，提出了适合应用领域的数据库技术，如工程数据库、统计数据库、科学数据库、空间库、地理数据库等。这类数据库在原理上没有太多的变化，但与应用相结合，把理论切实用到了实践当中，从而加强了系统对有关应用的支撑能力，尤其表现在数据模型、语言、查询等方面。

子任务 1.2 数据模型

【子任务描述】

数据需要经过认识、理解、整理、规范和加工才能存储到数据库中。在数据库中，数据以一定的数据模型进行组织、描述和存储，供用户共享。数据模型是数据库系统的核心和基础，是用于抽象、表示和处理现实世界中的数据和信息的模拟方法。

【子任务实施】

现实世界中的客观事物是相互联系的。一方面，某一事物内部的诸因素和诸属性根据一定的组织原则相互具有联系，构成一个相对独立的系统。另一方面，某一事物同时也作为一个更

大系统的一个因素或一种属性而存在，并与系统的其他因素或属性发生联系。

客观事物的这种普遍联系性决定了作为事物属性记录符号的数据与数据之间也存在一定的联系性。具有联系性的相关数据总是按照一定的组织关系排列的，从而构成一定的结构，对这种结构的描述就是数据模型。

1. 数据模型

数据模型 (Data Model) 是一种表示数据特征的抽象模型，是数据处理的关键和基础。它是专门用于抽象、表示和处理现实世界中的数据 (信息) 的工具，DBMS 的实现都是建立在某种数据模型基础上的。数据模型通常由数据结构、数据操作和完整性约束 (数据的约束条件) 3 个基本部分组成，称为数据模型的三要素。不同的数据模型实际上是提供给我们模型化数据和信息不同工具。数据模型应满足以下 3 方面的要求。

(1) 数据模型应能够比较真实地模拟现实世界。只有数据模型精确表达了真实的世界，才能正确地在计算机中存储数据信息。例如，利用数据模型正确地表达学生、教师与课程的关系。

(2) 数据模型应容易为人所理解。当设计人员构建数据模型表达客观世界时，必须先调查用户的实际需求，借助数据模型抽象用户需求，并通过不断反复的协商，与用户达成共识。因此数据模型不但要被设计人员理解，还要被用户理解。

(3) 便于在计算机上实现。由于计算机不能直接处理现实世界中的客观事物，所以必须通过一定的规则，将客观事物转化成可以存储在计算机中的数据，并有序地存储、管理这些数据，用户利用这些数据能够查询所需的信息。

因此，现实世界中客观对象的抽象过程，是一个将现实世界转变成信息世界、将信息世界转变成机器世界的过程，如图 1.5 所示。

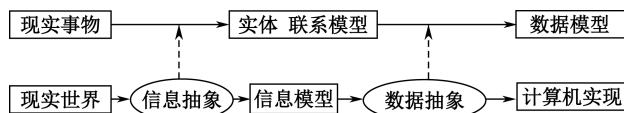


图 1.5 模型的抽象

2. 数据模型的组成因素

通常情况下，数据模型由数据结构、数据操作和完整性约束 3 部分组成。

1) 数据结构

数据结构是对系统静态特征的描述，它规定了如何把基本的数据项组织成较大的数据单位，以描述数据的类型、内容、性质和数据之间的相互关系。

2) 数据操作

数据操作是对系统动态特征的描述，它是指一组用于数据结构的所有有效的操作或推导规则。例如，从数据集中查询数据，根据现实世界的变化，修改、删除和更新数据。数据模型要给出这些操作的操作规则，如操作对象是谁，操作结果是什么，实现操作的方法是什么等。

3) 完整性约束

完整性约束是完整性规则的集合。它定义了给定数据模型中数据及其联系具有的制约和依存规则。只有在满足给定约束规则的前提下，才允许对数据库进行更新、删除等操作。

3. 数据模型的类型

数据模型是对现实世界的模拟，需要满足一方面要求，较真实地模拟现实世界、容易理解、易在计算机上实现。针对不同使用对象和应用目的，可采用不同的数据模型。

不同的数据模型用于模型化数据和信息不同工具。根据模型的不同应用，数据模型可分为 3 类：层次模型、网状模型和关系模型。其中，层次模型以“树结构”表示数据之间的联系，网状模型以“图结构”来表示数据之间的联系，关系模型以“二维表”来表示数据之间的联系。

1) 层次模型

层次模型 (Hierarchical Model) 是数据库系统最早使用的一种模型, 它的数据结构是一棵“有向树”。根节点在最上端, 层次最高, 子节点在下, 逐层排列。层次模型的优点是存取方便且速度快, 结构清晰, 易于理解, 数据修改和数据库扩展容易实现; 缺点是结构呆板, 缺乏灵活性, 数据冗余大。

层次模型的特点在于这种模型的数据库系统中, 要定义和保存每个节点的记录型及其所有值和每个父子联系; 对数据进行操作, 需要给出从树根节点开始的完整路径。用层次模型表示概念模型时, 对于一对一和一对多的联系可直接转换成层次模型中的父子联系, 而对于多对多联系则不能直接转换过来, 通常需要分解为一对多的联系来实现。用层次模型表示概念模型不方便, 因此产生了网状模型。层次模型的案例如图 1.6 所示。

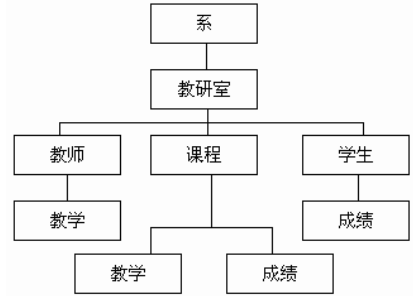


图 1.6 层次模型案例

2) 网状模型

网状模型 (Network Model) 以网状结构表示实体与实体之间的联系。网中的每一个节点代表一个记录类型, 联系用链接指针来实现。网状模型是层次模型的扩展, 可以表示多个从属关系的联系, 也可以表示数据间的交叉关系, 即数据间的横向关系和纵向关系。网状模型案例如图 1.7 所示。

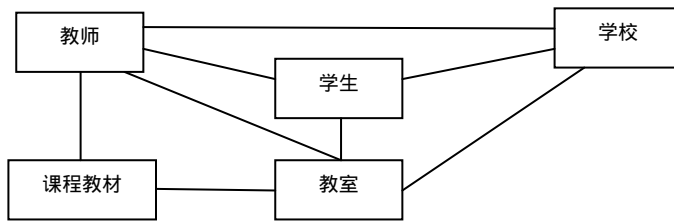


图 1.7 网状模型案例

网状模型的特点在于型和值的区别, 且需要经常变化。型是抽象的、静态的、相对稳定不变的, 值是具体的、动态的。由于经常需要对数据库中的业务数据 (值) 进行插入、删除和修改等实际操作, 改变具体实际的数据值。而逻辑数据结构模型一经建立后一般不会被轻易修改。在以网状数据模型实现的数据库系统中, 同样需要建立和保存所有节点的记录型、父子联系型, 以及所有数据值。

在数据库的查询和更新方面, 网状模型比层次模型灵活, 既允许按给定路径查询和更新数据, 又允许直接按节点的数据值查询和更新数据, 并可从子节点向父节点查询。网状模型和层次模型统称为非关系模型, 其本质相同, 网状包含层次, 适应范围更广。数据的操作方式都是过程式的, 即按照所给路径访问一个记录, 若要同时访问多条记录, 则必须通过用户程序中的循环过程来实现。网状模型表示数据之间多对多联系仍不简便, 也需要设法转换成一对多的联系, 而且, 存取数据仍是过程式的, 还需在程序中给出存取路径和具体方法, 增加了编程的负担, 程序和数据没有完全独立。

3) 关系模型

关系模型 (Relational Model) 是目前最流行的数据库模型。关系模型以二维表结构来表示实体与实体之间的联系, 操作的对象和结果都是二维表。该模型不分层也没有指针, 是建立空间数据和属性数据之间关系的一种非常有效的数据组织方法。关系模型的优点是结构灵活, 能搜索、组合和比较不同类型的数据, 数据增删方便, 数据独立性和安全保密性高; 缺点是数据库大时, 查找满足特定关系的数据费时, 无法满足空间关系。关系模型案例如图 1.8 所示。

学号	姓名	性别	专业	
0511011401	陈琳	女	国际贸易	01/01/88
0511011402	李兰兰	女	国际贸易	
0511011403	孙阳光	男	国际贸易	02/06/89
0511011501	吴云芳	女	服装设计	09/24/88
0511011502	关温丹	男	服装设计	
0511011601	秦春霞	女	艺术设计	01/30/87

图 1.8 关系模型案例

采用关系模型建立数据库系统具有以下优点。

(1) 理论基础扎实。关系模型与非关系模型不同，关系模型一开始就注重理论研究，建立在严格的数学概念基础上。关系系统研究逐渐完善，也促进了其他软件分支（如软件工程）的发展。

(2) 数据结构简单。在关系模型中，实体或实体之间的联系都用关系表示，关系不仅表示数据的存储，还表示数据之间的联系。从用户角度来看，关系模型中数据的逻辑结构是一张二维表，数据及其定义都以二维表（关系）的结构形式表示，符合人们使用数据的习惯，也便于计算机实现，每个关系（表）可以作为一个文件被保存到外存，由 DBMS 和操作系统共同管理。

(3) 查询处理方便，存取路径清晰。关系操作采用集合操作方式，关系模型中常用的关系操作包括选择、投影、连接、除、并、交、差等查询操作和插入、删除、修改等更新操作两部分，其中查询是最基本的操作。操作对象和结果都是集合，一次可操作所有满足条件的记录。

(4) 关系的完整性好。关系模型的完整性规则是对数据的约束，关系模型提供了 3 类完整性规则：实体完整性规则、参照完整性规则和用户自定义完整性规则。

(5) 数据独立性高。在关系模型中，对数据的操作不涉及数据的物理存储位置，而只需给出数据所在的表、属性等有关数据自身的特性，具有较高的数据独立性。

关系模型存在的不足之处如下。

(1) 查询效率低。关系模型的 DBMS 提供了较高的数据独立性和非过程化的查询功能，因此系统的负担很重，直接影响了查询速度和效率。

(2) RDBMS（关系数据库管理系统）实现较困难。由于关系数据库管理系统的效率较低，因此必须对关系模型的查询进行优化，这一工作相当复杂，实现难度较大。

子任务 1.3 概念模型

【子任务描述】

对现实世界认识和抽象时，需要先通过概念模型进行分析，再考虑数据结构形式中数据本身的结构和相互之间的内在联系，最后由计算机具体实现。

【子任务实施】

1. 概念模型

概念模型（Conceptual Model）是对真实世界中问题域内的事物的描述，不是对软件设计的描述。概念的描述包括记号、内涵、外延，其中记号和内涵是其最具实际意义的。概念数据模型也称信息模型，位于客观现实世界与机器世界之间。它只是用于描述某个特定机构所关心的数据结构，实现数据在计算机中表示的转换，是一种独立于计算机系统的数据模型。概念模型按用户的观点对数据进行建模，强调其语义表达能力。概念应该简单、清晰、易于用户理解，

是对现实世界的第一层抽象，是用户和数据库设计人员之间进行交流的工具，这类模型中最常用的是“实体联系模型”。

概念模型中的基本术语包括以下几种。

1) 联系

联系 (Relationship) 是指实体之间的相互关系，通常表示与实体有关的一种活动，如一张订单、一个讲座、一场比赛、一次选课等。

2) 键

键 (Key) 是区别实体的唯一标识，如学号、身份证号、工号、电话号码等，一个实体可以存在多个键。键可能是实体中的一个或一组属性，特别在联系实体中常为一组属性。实体中用于键的属性称为主属性 (Main Attribute)，否则称为非主属性。

3) 域

域 (Domain) 是实体中相应属性的取值范围，如“性别”的属性域为 (男, 女)。

2. 概念模型中实体的联系分类

联系分类 (Relationship Classify)

联系分类是指两个实体型 (含联系型在内) 之间的联系的类别。按照一个实体型中的实体个数与另一个实体型中的实体个数的对应关系，可分类为一对一联系、一对多联系、多对多联系等 3 种情况。

1) 一对一联系

若一个实体型中的一个实体至多与另一个实体型中的一个实体发生关系，同样，另一个实体型中的一个实体至多与该实体型中的一个实体发生关系，则这两个实体型之间的联系被定义为一对一联系，简记为 1 : 1。

2) 一对多联系

若一个实体型中的一个实体与另一个实体型中的任意多个实体 (含 0 个) 发生关系，而另一个实体型中的一个实体至多与该实体型中的一个实体发生关系，则这两个实体型之间的联系被定义为一对多联系，简记为 1 : n。

3) 多对多联系

若一个实体型中的一个实体与另一个实体型中的任意多个实体 (含 0 个) 发生关系，或另一个实体型中的一个实体与该实体型中的多个实体 (含 0 个) 实体发生关系，则这两个实体型之间的联系被定义为多对多联系，简记为 m : n。3 种实体间的联系如图 1.9 所示。

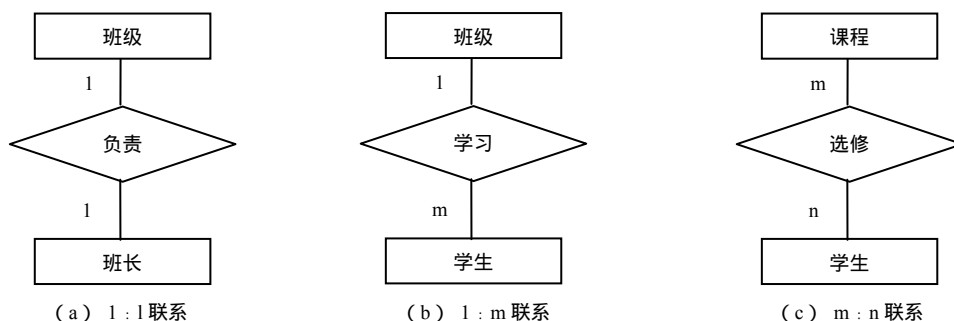


图 1.9 实体间联系图

3. 概念模型的代表方法

1976 年，美籍华人陈平山提出了实体联系模型 (Entity Relationship Model)，也称 E-R 模型或 E-R 图。这种模型是用 E-R 图描述事物及其联系的概念模型，是数据库应用系统设计者与普通用户进行数据建模和交流沟通的常用工具，非常直观易懂、简单易用。进行数据库应用系

统设计时,先要根据用户需求建立 E-R 模型,再建立与 DBMS 相适应的逻辑数据模型和物理数据模型,最后在计算机系统上建立、调试和运行数据库。

1) 绘制 E-R 图所需的图形工具

E-R 模型是一种用图形表示数据及其联系的方法,所使用的图形构件包括矩形、菱形、椭圆形和连接线。矩形表示实体类型(考虑问题的对象),矩形框内写上实体名。菱形表示联系类型(实体间联系),菱形框内写上联系名。椭圆形表示实体类型和联系类型的属性,椭圆形框内写上属性名。连接线表示实体、联系与属性之间的所属关系或实体与联系之间的相连关系。

2) E-R 图的 3 个主要部分

实体集:在 E-R 图中用矩形来表示实体集,实体是实体集的成员。

联系:在 E-R 图中用菱形来表示联系,联系与其涉及的实体集之间以直线连接,并在直线端部标上联系的种类(1:1, 1:n, m:n)。

属性:在 E-R 图中用椭圆形来表示实体集和联系的属性,对于主键码的属性,在属性名下画一条横线。

3) E-R 模型应用案例

使用 E-R 模型建立数据及其联系,首先要将应用系统中涉及的所有数据分类整理划分为若干个相互独立的实体,然后通过数据之间实际存在的各种关系建立起各独立实体之间的相互联系,最后形成统一的 E-R 图。

下面以网上商城系统为例绘制局部 E-R 模型图,并转换成关系模式。网上商城系统致力于提供产品展示及订购为核心的网上购物服务,商家通过该系统宣传自己的产品,并将产品展现给客户,让客户通过网上浏览便能自由地选择购买产品。该局部中存在 3 个实体集,一是“用户”实体集,属性有会员 ID、用户名、密码、详细地址等;二是“商品”实体集,属性有商品 ID、商品名、商品重量、商品描述等;三是“订单”实体集,属性有订单 ID、订单日期、总金额、运费等。

用户与商品间存在“购买”联系,每个用户可选购多种商品。用户与订单之间存在“产生”联系。网上商城系统用户、商品和订单之间的局部 E-R 图如图 1.10 所示。

关系模式如下。

用户(用户编号,用户账户,用户密码,联系方式)。

商品(商品编号,商品名称,商品内容,生产厂家)。

订单(订单编号,下单时间,订单状态,买方编号)。

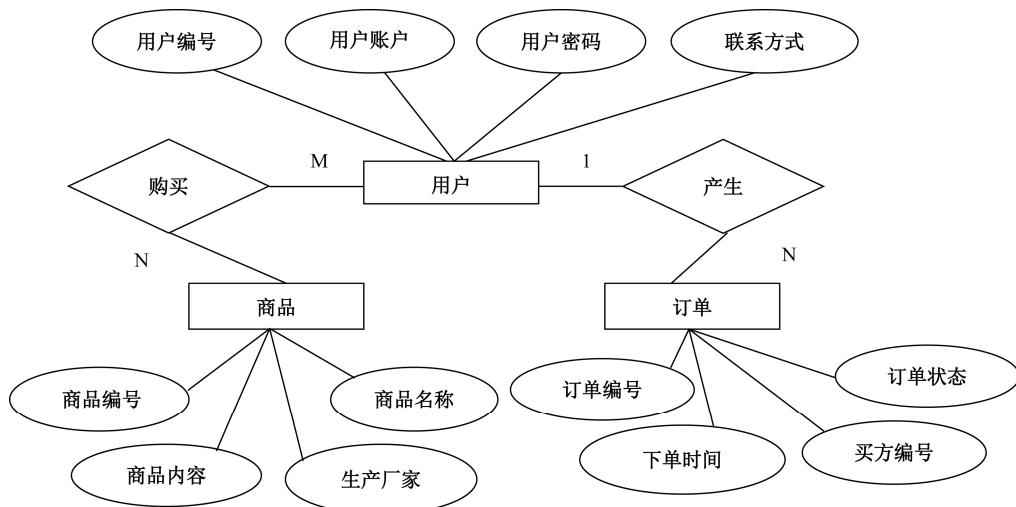


图 1.10 用户、商品和订单之间的局部 E-R 图

子任务 1.4 主流数据库系统及数据库语言

【子任务描述】

目前,数据库产品以关系数据库为主流,但各大数据库厂商也在开发面向对象的关系型数据库,实现高级程序设计语言和数据库的无缝连接,以发挥各自的优点。下面介绍几款代表性的数据库产品,用户可以结合相应的特点,选择适合自己的数据库产品。

【子任务实施】

目前,市场上主流的数据库产品包括 Oracle 公司的 Oracle、IBM 公司的 DB2 以及 Microsoft 公司的 SQL Server 和 Access 等。在一定意义上,这些产品的特征反映了当前数据库产业界的最高水平和发展趋势。因此,分析这些主流产品的发展状况,是了解数据库技术发展的一个重要方面。

一、数据库系统

1. Oracle

Oracle 是 1983 年推出的世界上第一个开放式商品化关系型数据库管理系统,Oracle 数据库被认为是业界目前比较成功的关系型数据库管理系统。它采用标准的结构化查询语言,支持多种数据类型,提供面向对象存储的数据支持,被认为是运行稳定、功能齐全、性能超群的产品。这一方面反映了它在技术方面的领先,另一方面反映了它在定位上更着重于大型的企业数据库领域。对于数据量大、事务处理繁忙、安全性要求高的企业,Oracle 无疑是比较理想的选择(当然,用户必须在费用方面做出充足的考虑,因为 Oracle 数据库在同类产品中是比较贵的)。随着 Internet 的普及,带动了网络经济的发展,Oracle 适时地将自己的产品紧密地和网络计算结合起来,成为在 Internet 应用领域数据库厂商的佼佼者。

Oracle 公司是目前全球最大的数据库软件公司,也是近年业务增长极为迅速的软件提供与服务商。IDC (Internet Data Center) 2007 年度统计数据显示数据库市场总量份额如下:Oracle 44.1%、IBM 21.3%、Microsoft 18.3%、Teradata 3.4%、Sybase 3.4%,但从使用情况看,BZ Research 的 2007 年度数据库与数据存取的综合研究报告表明 76.4% 的公司使用了 SQL Server,高端领域仍然以 Oracle、IBM、Teradata 为主。

2007 年 7 月 12 日,甲骨文公司在美国纽约宣布推出数据库 Oracle 11g,这是 Oracle 数据库的最新版本。Oracle 公司介绍说,Oracle 11g 有 400 多项功能,经过了 1500 万个小时的测试,开发工作量达到了 3.6 万人/月。Oracle 11g 在安全、XML DB、备份等方面得到了很大提升。Oracle 数据库可以运行在 UNIX、Windows 等主流操作系统平台,完全支持所有的工业标准,并获得最高级别的 ISO 标准安全性认证。Oracle 采用完全开放策略,可以使客户选择最适合的解决方案,同时对开发商提供全力支持。

2010 年 9 月份,Oracle 再次发布了数据库版本的重要更新,推出了 11.2.0.2.0,这也是 11g R2 的第一个 Patchset。同时从 11.2.0.2.0 开始,今后 Oracle 数据库所有的 Patchset 将以完整安装包的形式发布。Oracle 在 11.2.0.2 补丁版本中主要增强了 Oracle Automatic Storage Management Cluster File System (ACFS) 功能、Quality of Service Management 功能、Database Replay 功能和 Management 功能。

2. DB2

DB2 是 IBM 公司研制的一种关系型数据库系统,主要应用于大型应用系统,具有较好的可伸缩性,可支持从大型机到单用户等环境,应用于 OS/2、Windows 等平台下。DB2 提供了

高层次的数据利用性、完整性、安全性、可恢复性,以及小规模到大规模应用程序的执行能力,具有与平台无关的基本功能和 SQL 命令。其功能足以满足大中型公司的需要,并可灵活地服务于中小型电子商务解决方案,DB2 系统在企业级的应用中十分广泛。

1968 年,IBM 公司推出的信息管理系统(Information Management System,IMS)是层次数据库系统的典型代表,是第一个大型的商用数据库管理系统。1970 年,IBM 公司的研究员首次提出了数据库系统的关系模型,开创了数据库关系方法和关系数据理论的研究,为数据库技术奠定了基础。20 世纪 80 年代初期,DB2 的重点放在大型的主机平台上,到 20 世纪 90 年代初,DB2 发展到中型机、小型机及微机平台。2001 年,IBM 公司兼并了世界排名第四的著名数据库公司 Informix,并将其所拥有的先进特性融入到 DB2 中,使 DB2 系统的性能和功能有了进一步提高。2006 年 7 月 14 日,IBM 全球同步发布了一款具有划时代意义的数据库产品——DB2 9。这款新品最大特点即是率先实现了可扩展标记语言(XML)和关系数据间的无缝交互,而无需考虑数据的格式、平台或位置。

DB2 采用了数据分级技术,能够使大型机中的数据很方便地下载到 LAN 数据库服务器中,使得客户机/服务器用户和基于 LAN 的应用程序可以访问大型机数据,并使数据库本地化及远程连接透明化。它以拥有一个非常完备的查询优化器而著称,其外部连接改善了查询性能,并支持多任务并行查询。DB2 具有很好的网络支持能力,每个子系统可以连接十几个分布式用户,可同时激活上千个活动线程,对大型分布式应用系统尤为适用。

自 2006 年以来,IBM 增加了 2000 多名具有相关技能的从业人员,并继续扩展其全球的技术队伍,以便在客户部署信息基础架构环境时给予充分的帮助。同时,为了配合信息随需应变计划,IBM 宣布将在 3 年中投入 10 亿美元,并将其掌握相关技能的从业人员的数量增加 65%。IBM 还将在新的研究工作、产品开发、战略收购、客户服务、销售队伍和新的合作伙伴计划中进行持续不断的投资。

3. SQL Server

SQL Server 是一个关系数据库管理系统。它最初是由 Microsoft、Sybase 和 Ashton-Tate 3 家公司共同开发的,于 1988 年推出了第一个 OS/2 版本。在 Windows NT 推出后,Microsoft 与 Sybase 在 SQL Server 的开发上“分道扬镳”,Microsoft 将 SQL Server 移植到 Windows NT 系统上,专注于开发推广 SQL Server 的 Windows NT 版本。Sybase 则较专注于 SQL Server 在 UNIX 操作系统上的应用。

SQL Server 的功能全、效率高,可以作为大中型企业或单位的数据库平台。SQL Server 在可伸缩性与可靠性方面做了很多工作,近年来在许多企业的高端服务器上得到了广泛的应用。同时,该产品继承了微软产品界面友好、易学易用的特点,与其他大型数据库产品相比,在操作性和交互性方面独树一帜。SQL Server 可以与 Windows 操作系统紧密集成,这种安排使 SQL Server 能充分利用操作系统提供的特性,不论是应用程序开发速度还是系统事务处理运行速度,都能得到较大的提升。另外,SQL Server 可以借助浏览器实现数据库查询功能,并支持内容丰富的扩展标记语言,提供了全面支持 Web 功能的数据库解决方案。对于在 Windows 平台上开发的各种企业级信息管理系统来说,不论是 C/S(客户机/服务器)架构还是 B/S(浏览器/服务器)架构,SQL Server 都是一个很好的选择。SQL Server 的许多良好的性能和可靠性优势是通过与 Windows NT Server 平台的紧密集成来获得的。在微软平台上,SQL Server 能提供各方面较好的性能,并针对该平台进行了一些专门的设计。

4. Access

Access 是微软 Office 中一个重要成员,是在 Windows 操作系统下工作的关系型数据库管

理系统。它采用了 Windows 程序设计理念，以 Windows 特有的技术设计查询、用户界面、报表等数据对象，内嵌了 VBA (Visual Basic Application) 程序设计语言，具有集成的开发环境。Access 提供了图形化的查询工具和屏幕、报表生成器，当用户建立复杂的报表、界面时，无需编程和了解 SQL，Access 会自动生成 SQL 代码。

Access 被集成到 Office 中，具有 Office 系列软件的特点，如菜单、工具栏等。与其他数据库管理系统软件相比，更加简单、易学，一个普通的计算机用户，没有程序语言基础，仍然可以快速地掌握和使用它。更重要的一点是，Access 的功能比较强大，足以应付一般的数据管理及处理需要，适用于中小型企业数据管理的需求。当然，在数据定义、数据安全可靠、数据有效控制等方面，它比前面几种数据库产品要逊色不少。Access 数据库与 SQL Server 数据库之间的区别如表 1.1 所示。

表 1.1 Access 与 SQL Server 的区别

内 容	Access	SQL Server
版本	桌面版	网络版，可支持跨界的集团公司异地使用数据库的要求
节点	节点少，要锁定	节点多，支持多重路由器
管理权限	无	管理权限划分细致，对内安全性高
防黑客能力	无	数据库划分细致，对外防黑客能力高
并发处理能力	100 人或稍多	同时支持万人在线提交
导出 XML 格式	需要单编程序	可导出为 XML 格式，与 Oracle 和 DB2 通用，减少开发成本
数据处理能力	一般	快
是否被优化过	否	是

5. 大数据

“大数据”通常用来形容一个公司创造的大量非结构化数据和半结构化数据。“大数据”的意义不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理。从技术上看，“大数据”与“云计算”的关系就像一枚硬币的正反面一样密不可分。“大数据”必然无法用单台的计算机进行处理，必须采用分布式架构。它的特色在于对海量数据进行分布式数据挖掘，但它必须依托“云计算”的分布式处理、分布式数据库和云存储、虚拟化技术。

大数据分析相比于传统的数据仓库应用，具有数据量大、查询分析复杂等特点。业界将其归纳为 4 个“V”，即 Volume (数据体量大)、Variety (数据类型繁多)、Velocity (处理速度快)、Value (价值密度低)。从某种程度上说，大数据是数据分析的前沿技术。大数据最核心的价值就在于对海量数据进行存储和分析。相比于现有的其他技术而言，大数据的“廉价、迅速、优化”这 3 方面的综合成本是最优的。

二、数据库语言

1. VB

VB (Visual Basic) 是以 BASIC 语言作为其基本语言的一种可视化编程工具，它在众多编程工具中占据着非常重要的地位。VB 作为一种较早出现的开发程序以其容易学习，开发效率较高，具有完善的帮助系统等优点曾影响了几代编程人员。但是由于 VB 不具备跨平台的特性，从而决定了 VB 在未来的软件开发中将会逐渐地退出其历史舞台。它对组件技术的支持基于 COM 和 ActiveX，对于组件技术不断完善发展的今天，也显现出它的落后性。同时，VB 在进行系统底层开发的时候相对复杂，调用 API 函数不方便，不能进行 DDK 编程，不能深入 Ring0

编程，不能嵌套汇编。此外，它面向对象的特性差，网络功能和数据库功能也没有非常突出的表现，综上所述，VB 作为一种可视化的开发工具由于其本身的局限性，导致了它在未来软件开发中逐步被其他工具代替。

2. PB

PB (Power Builder) 是开发 MIS 和各类数据库跨平台的首选，使用简单，容易学习，容易掌握，在代码执行效率上也有相当出色的表现。PB 是一种真正的 4GL 语言（第四代语言），可随意直接嵌套 SQL 语句返回值被赋值到语句的变量中，支持语句级游标、存储过程和数据库函数，是一种类似 SQL 的规范，数据访问中具有无可比拟的灵活性。但是它在系统底层开发中与 VB 具有相同的缺陷，调用 API 函数需声明，调用不方便，不能进行 DDK 编程，不可能深入 Ring0 编程，不能嵌套汇编。在网络开发中提供了较多动态 Web 页面的用户对象和服务对象，非常适合编写服务端动态 Web 应用，有利于商业逻辑的封装。但是用于网络通信的支持不足，静态页面定制支持有限，使得 PB 在网络方面的应用也不能非常广泛。

3. C++Builder/Delphi

它们都是基于 VCL 库的可视化开发工具，在组件技术的支持、数据库支持、系统底层开发支持、网络开发支持、面向对象特性等方面都有相当不错的表现，并且学习使用较为容易，充分体现了所见即所得的可视化开发方法，开发效率高。由于两者都是 Borland 公司的产品，自然继承了该公司代码执行效率高的优良传统。但是，它们并不是毫无缺点，它们的不足之处是帮助系统在众多的编程工具中较差。C++Builder 的 VCL 库是基于 Object Pascal（面向对象 Pascal），使得 C++Builder 在程序的调试执行上落后于其他编程工具。而 Delphi 则有语言不够广泛、开发系统软件功能不足两个比较大的缺点。

4. Visual C++

Visual C++ 是基于 MFC 库的可视化的开发工具，从总体上说它是一个功能强大但是不便使用的工具。它在网络开发和多媒体开发中都具有不俗的表现，帮助系统也做得非常不错。但是，作为基本语言的 C++ 在面向对象特性上却不够好，主要是为了兼容 C 的程序，结果顾此失彼。在组件支持方面，虽然它支持 COM、ActiveX、CORBA，但是没有任何 IDE 支持，最大的问题是开发效率不高。

5. Java

Java 是一种可以撰写跨平台应用程序的面向对象的程序设计语言，是由 Sun Microsystems 公司于 1995 年 5 月推出的 Java 程序设计语言和 Java 平台（即 JavaSE，JavaEE，JavaME）的总称。Java 技术具有卓越的通用性、高效性、平台移植性和安全性，广泛应用于个人 PC、数据中心、游戏控制台、科学超级计算机、移动电话和互联网，同时拥有全球最大的开发者专业社群。在全球云计算和移动互联网产业环境下，Java 具备了更显著的优势和广阔前景。

课堂练习

一、选择题

1. 在数据库中存储的是（ ）。
A. 数据
B. 数据模型
C. 数据及数据之间的联系
D. 信息
2. 存储在计算机内有结构的数据的集合是（ ）。
A. 数据库系统
B. 数据库
C. 数据库管理系统
D. 数据结构