

1

数据库概述



本章描述

数据库技术是信息社会的重要基础技术，是计算机领域中发展最为迅速的重要分支。随着计算机技术的发展，数据处理所占的比重越来越大，而数据库技术是目前数据信息处理最完善的手段之一。各种管理信息系统、决策支持系统、办公自动化系统、计算机辅助设计/制造系统都使用数据库作为收集、组织、存储、加工、抽取和传播信息的主要手段。



学习目标

- 了解数据库技术的发展史
- 了解常用的商业数据库
- 了解最新的数据库研究领域及发展
- 掌握数据库的基本概念
- 掌握数据库系统结构的三级模型两级映像

1.1 数据管理技术的发展史

任务：了解数据管理技术的发展史

任务描述

随着计算机技术的不断发展，数据处理的效率和深度大大提高，数据管理技术也得到了巨大发展，经历了人工管理、文件系统和数据库系统三大阶段，如图 1-1 所示。

相关知识

数据库技术产生的主要目的是有效地管理和存取大量的数据资源。数据库技术主要研究如何存储、使用和管理数据。近年来，数据库技术和计算机网络技术的发展相互渗透，相互促进，已成为

当今计算机领域发展迅速、应用广泛的两大领域。数据库技术不仅应用于事务处理，更进一步应用到情报检索、人工智能、专家系统、计算机辅助设计等领域。下面就来具体了解一下数据库的各个发展阶段。

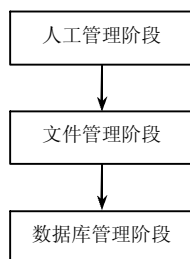


图 1-1 数据管理发展阶段

1. 人工管理阶段

早期的计算机主要用于科学计算。这个阶段的数据处理是通过手工进行的。当时计算机没有专门管理数据的软件，也没有像磁盘这样可以随机存取的外部存储设备，对数据的管理没有一定的格式，数据依附于处理它的应用程序，使得数据和应用程序一一对应。如图 1-2 所示是应用程序与数据之间的关系。

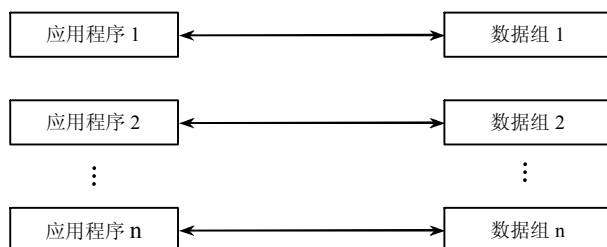


图 1-2 人工管理阶段应用程序与数据之间的关系

这种关系使程序和数据之间的依赖性强、独立性差，并且数据与数据之间存在许多重复，造成大量数据冗余。

2. 文件管理阶段

从 20 世纪 50 年代后期到 60 年代末为文件管理阶段，应用程序通过专门管理数据的软件即文件系统管理来使用数据。在文件系统中，数据按照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。如图 1-3 所示为文件管理阶段应用程序与数据之间的关系。

在文件系统管理阶段，文件系统为程序与数据之间提供了一个公共接口，使应用程序可以采用统一的存取方法来操作数据。但是，不同的应用程序很难共享同一数据，也就是说数据的独立性仍然较差，数据冗余度较大。

3. 数据库管理阶段

数据库管理阶段是 60 年代末在文件管理基础上发展起来的。随着计算机及软件技术的发展，人们克服了文件系统的不足，开发出新的数据管理软件——数据库管理系统，运用数据库管理技术

进行数据管理，将数据管理技术推向了数据库管理阶段。

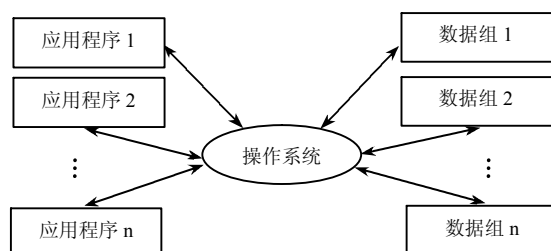


图 1-3 文件管理阶段应用程序与数据之间的关系

数据库技术使数据有了统一的结构，可以对数据进行统一、集中、独立的管理，以此实现数据的共享，并能够保证数据的完整性和安全性，提高了数据管理效率。在数据库管理阶段，应用程序与数据之间的关系如图 1-4 所示。

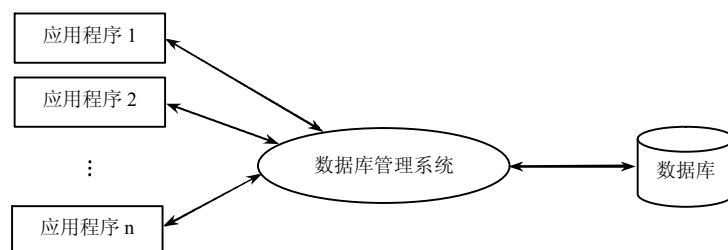


图 1-4 数据库管理阶段应用程序与数据之间的关系



数据库系统与文件系统的区别是：数据库中数据的存储是按同一结构进行的，不同的应用程序都可以直接操作使用这些数据，应用程序与数据之间保持了高度的独立性；数据库提供一套有效的管理办法，保证数据的完整性、一致性和安全性，使数据具有充分的共享性；数据库还为用户管理、操作数据提供了功能强大的操作命令，使用户直接使用命令或将命令嵌入到应用程序中，简单方便地实现了数据库的管理和控制操作。

1.2 数据库的组成及特点

任务：掌握数据库的组成及特点

任务描述

在计算机技术日益发达的今天，计算机存储容量不断扩大，处理能力不断加快，数据大都集中存放在计算机或计算机网络中，并以数据库的形式存在。因此，在详细了解数据库设计知识前，首先要掌握数据库的组成部分：数据（Data）、数据库（DB）、数据库管理系统（DBMS）、数据库系统（DBS），这些组成部分的作用层次如图 1-5 所示。

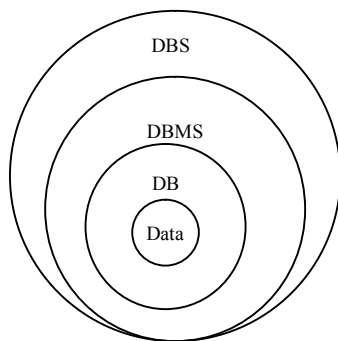


图 1-5 组成部分层次图

相关知识

一、数据库基本概念

1. 数据

数据（Data）——客观事物的反映和记录，是用来载荷信息的物理符号。

人们通常使用各种各样的物理符号来表示客观事物的特性和特征，这些符号及其组合就是数据。在计算机系统中，能被计算机存储并处理的数字、字符、图形和声音等统称为数据。也就是说数据就是指存储在某一种媒体上能够识别的物理符号。



数据的概念包括两个方面，即数据内容和数据形式。数据内容用于描述客观事物特性，也就是通常所说的数据的“值”；数据形式是指数据内容存储在媒体上的形式，也就是通常所说的数据的“类型”。

在计算机系统中，数据不等于数字。数据是指所有能输入到计算机并被计算机程序处理的符号的介质的总称，是用于输入电子计算机进行处理，具有一定意义的数字、字母、符号和模拟量等的通称。数据不仅包括数字、字母、文字和其他符号组成的文本形式的数据，而且还包括图形、图像、动画、影像、声音等多媒体数据。数据包括两大类：数值型数据和非数值型数据。



例如，学校为了管理学生信息，可能会将学生的学号、姓名、性别、年龄、出生年月、籍贯、所在系、入学时间等数据组织起来，下面就是一个学生有关这些方面的数据集：

（20012502033，李明，男，19，1982，江苏，计算机系，2001）

2. 数据库

数据库（DataBase，DB）是一个按数据结构来存储和管理数据的计算机软件系统，是数据库系统中以一定的组织方式将相关数据组织在一起，存储在外部存储介质设备上所形成的、能为多个用户共享的、与应用程序相互独立的相关数据集合。

数据库的概念实际上包含两层意思：

（1）数据库是一个实体，它是能够合理保管数据的“仓库”，用户在该“仓库”中存放要管理的事务数据。

（2）数据库是数据管理的新方法和新技术，它能够更合理地组织数据，更方便地维护数据，

更严密地控制数据和更有效地利用数据。



数据库，顾名思义，是存放数据的仓库。只不过这个仓库是在计算机存储设备上，而且是按一定的格式存放的。

3. 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大型软件，用于建立、使用和维护数据库。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。它提供多种功能，可使多个应用程序和用户用不同的方法在同时或不同时刻去建立、修改和访问数据库。它使用户能方便地定义和操纵数据，维护数据的安全性和完整性，以及进行多用户下的并发控制和恢复数据库。数据库管理系统的基本工作模式如图 1-6 所示。

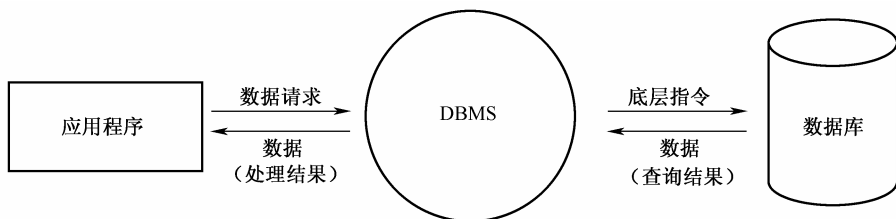


图 1-6 数据库管理系统的工作模式

一般来说，数据库管理系统包含如下几个功能：

（1）数据定义。DBMS 提供数据定义语言（DDL）来描述和定义数据库中的各种数据和数据之间的联系。

（2）数据操纵。DBMS 提供数据操纵语言（DML），实现对数据库数据的基本存取操作：检索、插入、修改和删除。

（3）数据控制。DBMS 提供数据控制功能，即数据的安全性、完整性和并发控制等，对数据库运行进行有效的控制和管理，以确保数据正确有效。

（4）数据组织、存储与管理。DBMS 要分类组织、存储和管理各种数据，包括数据字典、用户数据、存取路径等，需要确定以何种文件结构和存取方式在存储级上组织这些数据，如何实现数据之间的联系。数据组织和存储的基本目标是提高存储空间利用率，选择合适的存取方法提高存取效率。

（5）数据库的保护。数据库中的数据是信息社会的战略资源，所以数据的保护至关重要。DBMS 对数据库的保护通过 4 个方面来实现：数据库的恢复、数据库的并发控制、数据库的完整性控制、数据库的安全性控制。DBMS 的其他保护功能还有系统缓冲区的管理和数据存储的某些自适应调节机制等。

（6）数据库的维护。这一部分包括数据库的数据载入、转换、转储、数据库的重组/重构、性能监控等功能，这些功能分别由各个应用程序来完成。

（7）通信。DBMS 具有与操作系统联机处理、分时系统及远程作业输入的相关接口，负责处理数据的传送。对网络环境下的数据库系统，还应该包括 DBMS 与网络中其他软件系统的通信功

能以及数据库之间的互操作功能。



关系数据库领域中典型的 DBMS 系统有：SQL Server、MySQL、Oracle、DB2、Sybase、FoxPro 等。在后面针对典型的 DBMS 产品会进行详细介绍。

4. 数据库系统

数据库系统 (DataBase Systems, DBS) 实际上是一个应用系统, 由数据库 (DB)、数据库管理系统 (DBMS)、用户和计算机系统等组成, 如图 1-7 所示。

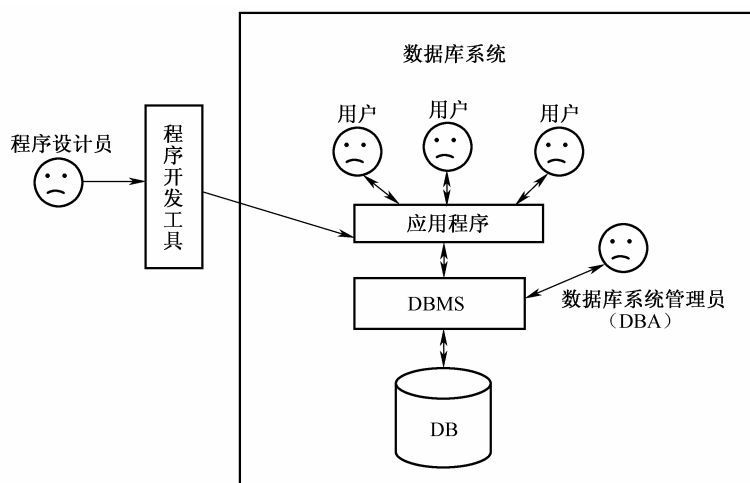


图 1-7 数据库系统展示

数据库系统是为适应数据处理的需要而发展起来的一种较为理想的数据处理的核心机构, 是一个实际可运行的存储、维护和应用数据的软件系统, 是存储介质、处理对象和管理系统的集合体。数据库系统一般由以下四部分组成:

(1) 数据库: 是数据库系统操作的对象。数据库中的数据具有集中性和共享性。所谓集中性是指数据库可以被看做是性质不同的数据文件的集合, 其中的数据冗余很小。所谓共享性是指多个不同用户, 使用不同的语言, 为了不同的应用目的可以同时存取数据库中的数据。

(2) 数据库管理系统: 是数据库系统中负责对数据库进行管理的软件系统。它对数据库中的数据资源进行统一的管理和控制, 把用户程序和数据库数据进行隔离。

(3) 用户: 是指使用数据库的人员。数据库系统中的用户有终端用户、应用程序员和数据库管理员三类。

(4) 计算机系统: 是指存储数据库及运行 DBMS 的软硬件资源的系统, 如操作系统和磁盘等。

如图 1-8 所示是数据库管理系统与计算机硬件及其他软件的层次关系, 外层应用依赖于内层资源的支持。

综上所述, 数据库中包含的数据是存储在存储介质上的数据文件的集合; 每个用户均可使用其中的部分数据, 不同用户使用的数据可以重叠, 同一组数据可以为多个用户共享; DBMS 为用户提供对数据的存储组织和操作管理的功能, 用户通过 DBMS 和应用程序实现数据库系统的操作与应用。

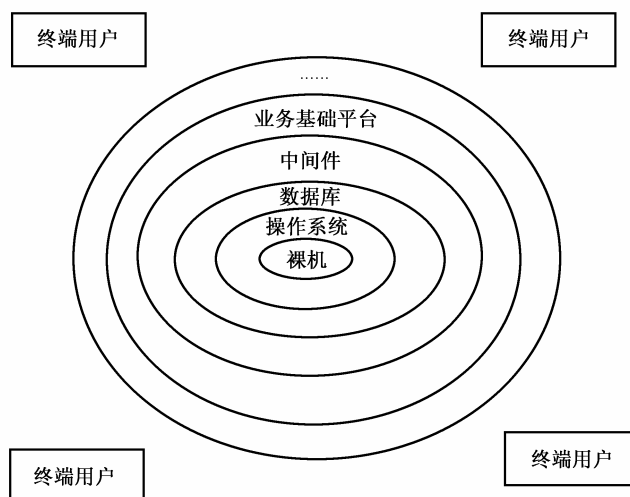


图 1-8 软硬件的层次关系

二、数据库系统的特点

数据库系统的出现是计算机数据处理技术的重大进步。数据库技术主要解决数据处理的非数值计算问题，例如仓库管理、档案管理、图书资料管理等；数据库处理的主要内容包括数据的存储、查询、修改、分类排序、决策支持等；具有数据共享、数据结构化、数据独立性、可控数据冗余度和统一数据控制等特点。

1. 数据结构化

数据结构化是数据库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。在文件系统中，尽管记录内部已有了某些结构，但记录之间没有联系。数据库系统实现了整体数据的结构化，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，数据不再针对某一应用，而是面向全组织，具有整体的结构化。不仅数据是结构化的，而且存取数据的方式也很灵活，可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录。而在文件系统中，数据的最小存取单位是记录。

2. 数据的共享性高，冗余低，易扩展

数据库系统从整体角度描述数据，数据不再面向某个应用而是面向整个系统，因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余，节约存储空间。数据共享还能够避免数据之间的不相容性与不一致性。所谓数据的不一致性，是指同一数据不同拷贝的值不一样。采用人工管理或文件系统管理时，由于数据被重复存储，当不同的应用使用和修改不同的拷贝时就很容易造成数据的不一致。在数据库中数据共享，减少了由于数据冗余造成的不一致现象。由于数据面向整个系统，是有结构的数据，不仅可以被多个应用共享使用，而且容易增加新的应用，这使得数据库系统弹性大、易于扩充，可以适应各种用户要求。可以取整体数据的各种子集应用于不同的应用系统，当应用需求改变或增加时，只要重新选取不同的子集或加上一部分数据便可以满足新的需求。

3. 数据的独立性高

数据的独立性是数据库领域中的一个常用术语，包括数据的物理独立性和数据的逻辑独立性。

物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的。也就是说,数据在磁盘上的数据库中的存储是由 DBMS 管理的,用户程序不需要了解,应用程序要处理的只是数据的逻辑结构,这样当数据的物理存储改变时,应用程序却不用改变。逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的,也就是说,数据的逻辑结构改变了,用户程序也可以不变。数据与程序的独立,把数据的定义从程序中分离出去,加上数据的存取又由 DBMS 负责,从而简化了应用程序的编制,大大减少了应用程序的维护和修改。

4. 数据由 DBMS 统一管理和控制,增强了数据安全性和完整性保护

数据库的共享是并发的共享,即多个用户可以同时存取数据库中的数据,甚至可以同时存取数据库中的同一数据。为此,DBMS 还必须提供以下几方面的数据控制功能:

(1) 数据的安全性 (Security) 保护。

数据的安全性是指保护数据以防止不合法的使用造成数据的泄密和破坏,使每个用户只能按规定对某些数据以某些方式进行使用和处理。

(2) 数据的完整性 (Integrity) 检查。

数据的完整性是指数据的正确性、有效性和相容性。完整性检查将数据控制在有效的范围内,或保证数据之间满足一定的关系。

(3) 并发 (Concurrency) 控制。

当多个用户的并发进程同时存取、修改数据库时,可能会发生相互干扰而得到错误的结果,或使得数据库的完整性遭到破坏,因此必须对多用户的并发操作加以控制和协调。

(4) 数据库恢复 (Recovery)。

计算机系统的硬件故障、软件故障、操作员的失误,以及故意的破坏会影响数据库中数据的正确性,甚至造成数据库部分或全部数据丢失。DBMS 必须具有将数据库从错误状态恢复到某一已知的正确状态(亦称为完整状态或一致状态)的能力,这就是数据库的恢复功能。数据库是长期存储在计算机内有组织的、大量的、共享的数据集合。它可以供各种用户共享,具有最小冗余度和较高的数据独立性。DBMS 在数据库建立、运用和维护时对数据库进行统一控制,以保证数据的完整性、安全性,并在多用户同时使用数据库时进行并发控制,在发生故障后对系统进行恢复。数据库系统的出现使信息系统从以加工数据的程序为中心转向围绕共享的数据库为中心的新阶段。这样既便于数据的集中管理,又有利于应用程序的研制和维护,提高了数据的利用率和相容性,提高了决策的可靠性。目前,数据库已经成为现代信息系统的不可分离的重要组成部分。具有数百万甚至数十亿字节信息的数据库已经普遍存在于科学技术、工业、农业、商业、服务业和政府部门的信息系统中。

1.3 常用商业数据库

任务: 了解常用数据库

任务描述

当前数据库管理的主流产品均是采用关系模型的数据库管理系统。按其规模大致可以分为大型、中型、小型、桌面式四种。每种规模的数据库管理系统都有一些具有代表性的产品,如图 1-9 所示。

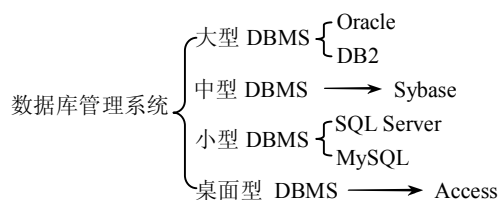


图 1-9 DBMS 的分类

相关知识

常用数据库管理系统介绍

1. SQL Server

SQL Server 是由微软开发的数据库管理系统，是 Web 上最流行的用于存储数据的数据库，它已广泛用于电子商务、银行、保险、电力等与数据库有关的行业，目前最新版本是 SQL Server 2012。

SQL Server 提供了众多的 Web 和电子商务功能，如对 XML 和 Internet 标准的丰富支持，通过 Web 对数据进行轻松安全的访问，具有强大的、灵活的、基于 Web 的和安全的应用程序管理等。而且，由于易操作性及友好的操作界面，其深受广大用户的喜爱。缺点是，它只能在 Windows 上运行，操作系统的稳定性对数据库十分重要，并行实施和共存模型并不成熟，很难处理日益增多的用户数和数据卷，伸缩性有限。操作界面如图 1-10 所示。

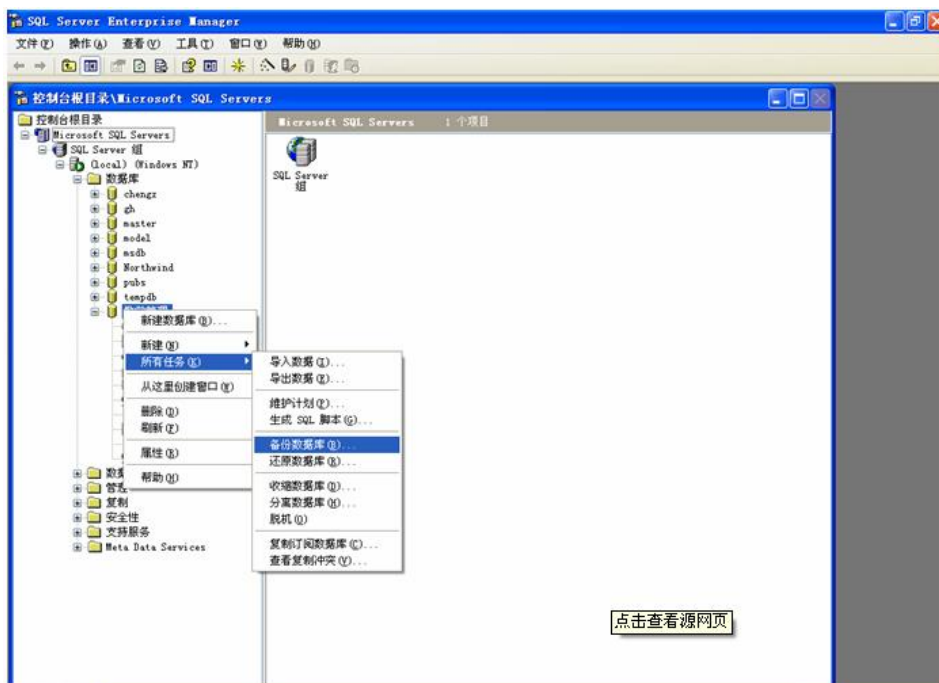


图 1-10 常用商业数据库 SQL Server 操作界面

2. MySQL

MySQL 是最受欢迎的开源 SQL 数据库管理系统，由 MySQL AB 开发、发布和支持。MySQL AB 是一家基于 MySQL 开发人员的商业公司，是一家使用了一种成功的商业模式来结合开源价值和方法论的第二代开源公司。MySQL 是 MySQL AB 的注册商标。与其他数据库管理系统相比，MySQL

具有以下优势：

- (1) MySQL 是一个关系数据库管理系统。
- (2) MySQL 是开源的。
- (3) MySQL 服务器是一个快速的、可靠的和易于使用的数据库服务器。
- (4) MySQL 服务器工作在客户/服务器或嵌入式系统中。
- (5) 有大量的 MySQL 软件可以使用。

MySQL 是一个快速的、多线程、多用户和健壮的 SQL 数据库服务器。MySQL 服务器支持关键任务、重负载生产系统的使用，也可以将它嵌入到一个大配置（mass-deployed）的软件中去。操作界面如图 1-11 所示。

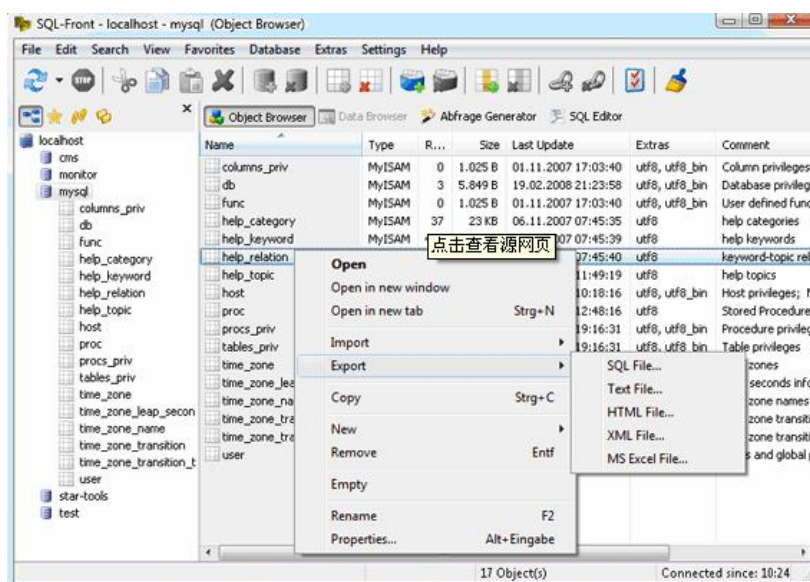


图 1-11 常用商业数据库 MySQL 操作界面

3. Oracle

说起数据库，第一个想到的公司，一般都会是 Oracle（甲骨文）。该公司成立于 1977 年，最初是一家专门开发数据库的公司。Oracle 数据库系统是美国 Oracle 公司提供的以分布式数据库为核心的的一组软件产品，是目前最流行的客户/服务器（Client/Server）或 B/S 体系结构的数据库之一。比如 SilverStream 就是基于数据库的一种中间件。Oracle 数据库是目前世界上使用最为广泛的数据库管理系统，作为一个通用的数据库系统，它具有完整的数据管理功能；作为一个关系数据库，它是一个完备关系的产品；作为分布式数据库，它实现了分布式处理功能。但它的知识只要在一种机型上学习了，便能在各种类型的机器上使用它。目前，Oracle 产品覆盖了大、中、小型机等几十种机型，Oracle 数据库成为世界上使用最广泛的关系数据系统之一。Oracle 的操作界面如图 1-12 所示。

Oracle 数据库产品具有以下优良特性：

- (1) 兼容性。

Oracle 产品采用标准 SQL，并经过美国国家标准技术所（NIST）测试。与 IBM SQL/DS、DB2、INGRES、IDMS/R 等兼容。

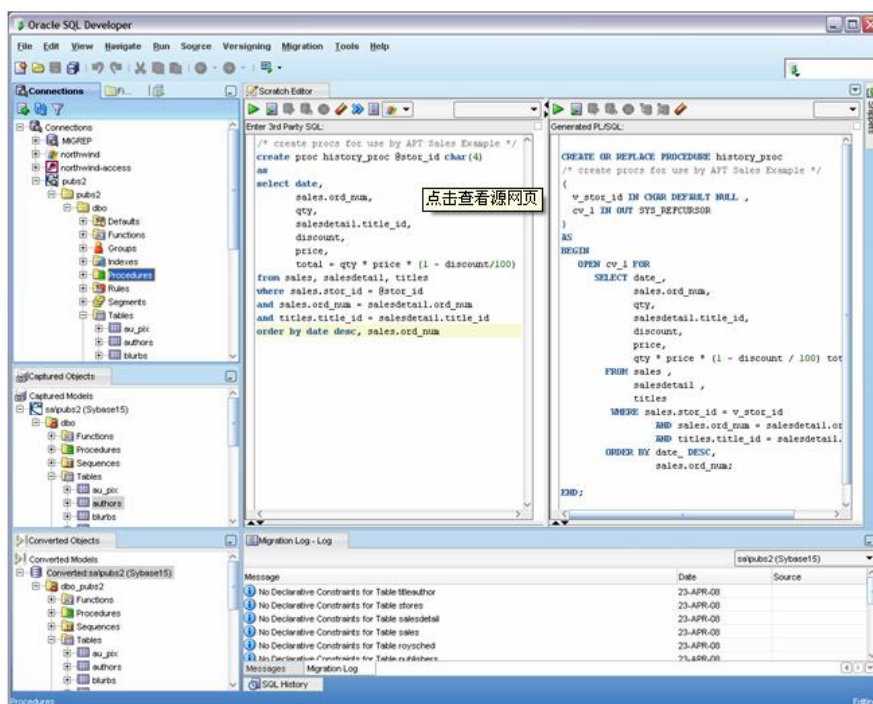


图 1-12 常用商业数据库 Oracle 操作界面

(2) 可移植性。

Oracle 的产品可运行于很宽范围的硬件与操作系统平台上。可以安装在 70 种以上不同的大、中、小型机上；可在 VMS、DOS、UNIX、Windows 等多种操作系统下工作。

(3) 可联结性。

Oracle 能与多种通讯网络相连，支持各种协议（TCP/IP、DECnet、LU6.2 等）。

(4) 高生产率。

Oracle 产品提供了多种开发工具，能极大地方便用户进行进一步的开发。

(5) 开放性。

Oracle 良好的兼容性、可移植性、可连接性和高生产率使 Oracle RDBMS 具有良好的开放性。

其他数据库如 Sybase、DB2，包括微软的 Access 和 FoxPro 等，这里就不一一介绍了，有兴趣的读者可以查找相关资料学习。

1.4 数据库系统结构

任务：掌握数据库系统结构

任务描述

虽然数据库管理系统众多，支持的数据模型各有不同，使用不一致的数据库语言，针对于不同的操作系统，存储结构及规模也不太相同。但是对于数据库系统内部体系结构来说是一致的。数据

库系统的典型结构是“三级模式、两级映象”结构。如图 1-13 所示表示数据抽象各个层次中记录的联系。

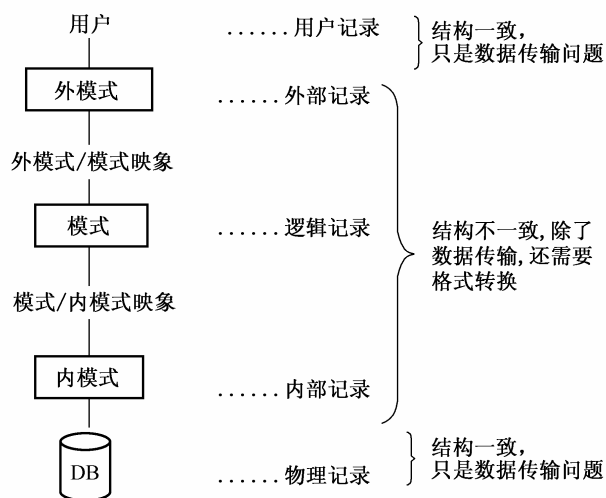


图 1-13 数据库系统的三级模式、两级映象结构

相关知识

对数据库中全体数据的逻辑结构和特征所作的描述就称为数据库系统的模式（Schema）。这种描述只刻画了数据集合的“型”，不涉及具体的数据“值”。和型相对应的值有若干个，每一个值称为模式的一个实例（Instance）。

模式反映了数据库系统中数据的结构和相互联系，是一种相对稳定的描述；实例反映了数据库系统在某一时刻的状态，是一种动态的数据集。如图 1-14 所示为数据访问的过程。

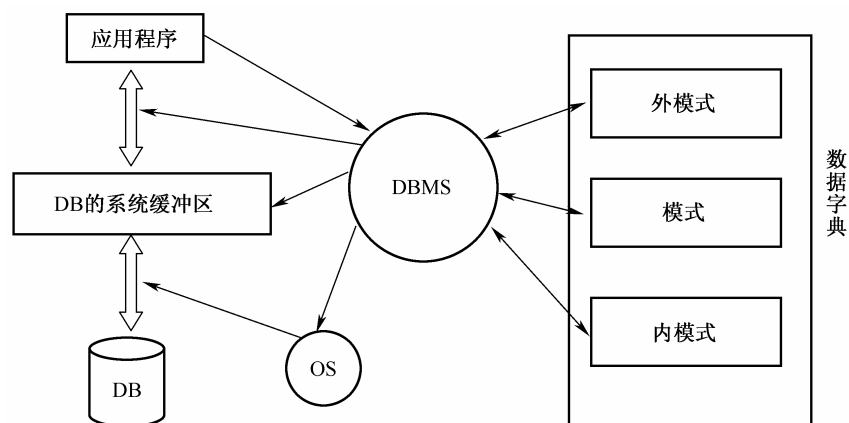


图 1-14 用户访问数据的过程

一、数据库系统的三级模式结构

数据库系统的三级模式结构由外模式、模式、内模式以及外模式/模式映象、模式/内模式映象组成。

1. 模式（Schema）

模式也称为逻辑模式，是 DB 中全体数据的逻辑结构描述，是所有用户的公共数据视图。模式不仅要描述数据的逻辑结构，还要定义与数据有关的安全性、完整性要求，同时还要描述数据项之间以及记录之间的联系。



一个数据库只有一个模式；是数据库数据在逻辑级上的视图；数据库模式以某一种数据模型为基础；定义模式时不仅要定义数据的逻辑结构（如数据记录由哪些数据项构成，数据项的名字、类型、取值范围等），而且要定义与数据有关的安全性、完整性要求，定义这些数据之间的联系。

2. 外模式（External Schema）

外模式也称为子模式（Subschema）或用户模式，是某用户看到的数据视图，即是对某应用有关的数据的逻辑描述。子模式一般是模式的一个子集，不同用户的外模式可以互相覆盖，一个外模式可以被多个应用使用，但一个应用只能使用一个外模式。



对数据库中的同一数据，在模式和外模式的描述中是可以有差别的，以适应不同的需要。外模式由数据库管理系统提供的子模式 DDL（数据定义语言）决定实现。

3. 内模式（Internal Schema）

内模式也称存储模式（Storage Schema），是对数据的物理结构和存储方式所作的描述。如数据存放的原则、索引方式、数据是否压缩/加密、数据的存储记录结构说明等。例如，记录的存储方式是顺序存储、按照 B 树结构存储还是按 Hash 方法存储；索引按照什么方式组织；数据是否压缩存储，是否加密；数据的存储记录结构有何规定。



一个数据库只有一个内模式，一个表可能由多个文件组成，如数据文件、索引文件，它是数据库管理系统对数据库中数据进行有效组织和管理的方法，其目的有：①减少数据冗余，实现数据共享；②提高存取效率，改善性能。

下面来看一下外模式、模式、内模式三级模式在具体数据库应用中的具体表现，如图 1-15 所示是销售系统数据库的三级模式实例。

4. 三级模式的比较

数据库系统三级模式结构概念的比较如表 1-1 所示。

表 1-1 数据库系统三级模式结构概念的比较

	外模式	模式	内模式
名称	也称子模式或用户模式，还称用户级模式	也称逻辑模式，还称概念级模式	也称存储模式，还称物理级模式
定义	是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述	是数据库中全体数据的逻辑结构和特征的描述，它包括：数据的逻辑结构、数据之间的联系和与数据有关的安全性、完整性要求	它是数据物理结构和存储方式的描述

续表

	外模式	模式	内模式
特点 1	是各个具体用户所看到的数据视图，是用户与 DB 的结构	是所有用户的公共数据视图。一般只有 DBA 能看到全部	是数据在数据库内部的表示方式
特点 2	可以有多个外模式	只有一个模式	只有一个内模式
特点 3	针对不同的用户，有不同的外模式描述。每个用户只能看见和访问所对应的外模式中的数据，数据库中的其余数据是不可见的。所以外模式是保证数据库安全性的一个有力措施	数据库模式以某一种数据模型（层状、网状、关系）为基础，统一综合地考虑所有用户的需求，并将这些需求有机地结合成一个逻辑整体	以前由 DBA 定义，现在基本由 DBMS 定义
特点 4	面向应用程序或最终用户	由 DBA 定义与管理	由 DBA 定义或由 DBMS 预先设置
DDL	DBMS 提供 3 种模式的描述语言（DDL）来严格定义 3 种模式，如子模式 DDL、模式 DDL 和内模式 DDL。子模式 DDL 和用户所用的程序设计语言具有相容的语法，如 Cobol 子模式 DDL。关系数据库 3 种模式的描述语言统一于 SQL 语言中		

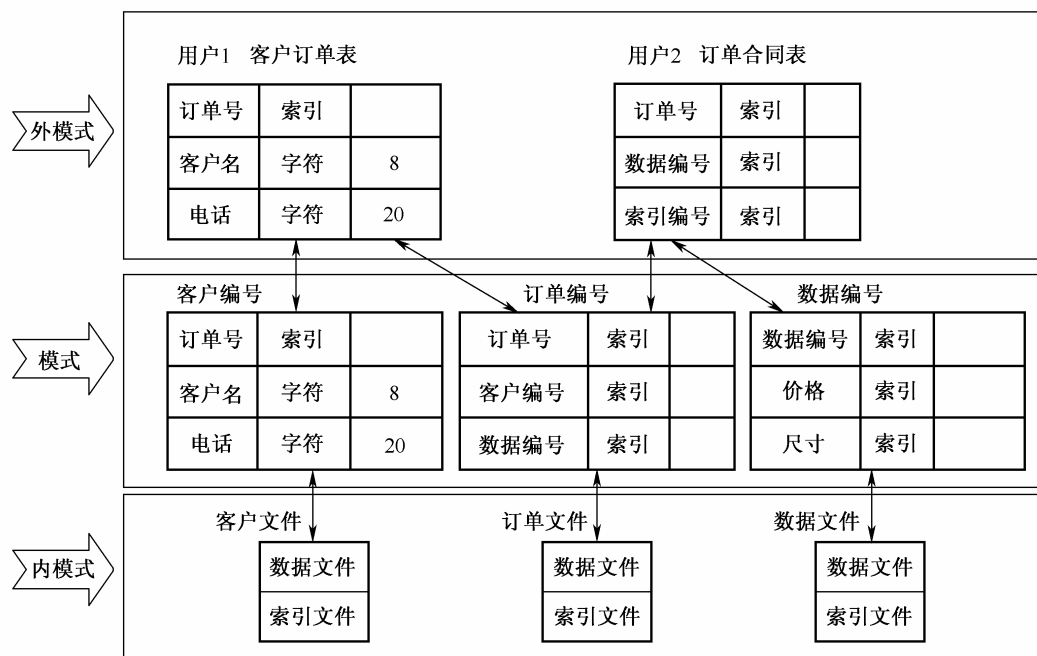


图 1-15 销售系统数据库的三级模式实例

二、数据库的二级映象功能与数据独立性

1. 外模式/模式映象

在模式和每一个外模式之间，均有一个由 DBMS 维护的映象，这个映象定义了该外模式和模式之间的对应关系，称为外模式/模式映象。这一级映象保证了程序与数据的逻辑独立性。



数据独立性包括数据的物理独立性和逻辑独立性。逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的，即当数据的逻辑结构改变时，用户程序也可以不变。

2. 模式/内模式映象

数据库系统的模式和内模式之间也存在一级映象，称为模式/内模式映象。它定义了数据库的全局逻辑结构和存储结构之间的对应关系。这个映象保证了程序与数据之间的物理独立性。

物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的，即数据在磁盘上怎样存储由 DBMS 管理，用户程序不需要了解，应用程序要处理的只是数据的逻辑结构，这样当数据的物理存储改变时，应用程序不用改变。

1.5 数据库技术的研究领域及其发展

任务：了解数据库技术的发展

任务描述

随着数据库领域的不断扩张和信息量的急剧增长，占主导地位的数据库系统已不能满足新的应用领域的需求，如计算机辅助设计、计算机辅助制造、办公自动化、地理信息系统和知识库系统等，都需要数据库新技术的支持。数据库研究领域主要在分布式数据库、面向对象数据库、数据仓库三个方面，如图 1-16 所示。

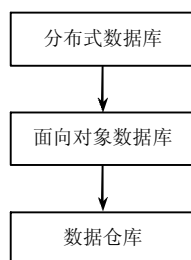


图 1-16 数据库技术发展阶段

相关知识

从 20 世纪 60 年代中期以来，数据库技术已从第一代的网状/层次数据库系统、第二代的关系数据库系统，发展到第三代以面向对象模型为主要特征的数据库系统，并且数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透，互相结合，成为当前数据库技术发展的主要特征。

数据库技术与其他技术相结合，出现了数据库的许多新分支，如与网络技术结合出现了网络数据库；与分布处理技术结合出现了分布式数据库；与面向对象结合出现了面向对象数据库；与人工智能技术相结合出现了知识库等。

一、分布式数据库

分布式数据库（Distributed DataBase System，DDBS）是在集中式数据库系统的基础上发展而

来的,是数据库技术与网络技术结合的产物,是一种数据分布在计算机网络中的各台计算机上的数据库。分布式数据库系统通常使用较小的计算机系统,每台计算机可单独放在一个地方,每台计算机中都有 DBMS 的一份完整的拷贝副本,并具有自己局部的数据库,位于不同地点的许多计算机通过网络互相连接,共同组成一个完整的、全局的大型数据库。



分布式数据库系统有两种:一种是物理上分布的,但逻辑上却是集中的,这种分布式数据库只适宜用途比较单一的、不大的单位或部门;另一种分布式数据库系统在物理上和逻辑上都是分布的,也就是所谓的联邦式分布数据库系统。由于组成联邦的各个子数据库系统是相对“自治”的,这种系统可以容纳多种不同用途的、差异较大的数据库,比较适宜于大范围内数据库的集成。

分布式数据库组织数据库的方法克服了物理中心数据库组织的弱点。首先,降低了数据传送代价,因为大多数对数据库的访问操作都是针对局部数据库的,而不是对其他位置的数据库访问;其次,系统的可靠性提高了很多,因为当网络出现故障时,仍然允许对局部数据库的操作,而且一个位置的故障不影响其他位置的处理工作,只有当访问出现故障位置的数据时,才在某种程度上受到影响;第三,便于系统的扩充,增加一个新的局部数据库或在某个位置扩充一台适当的小型计算机都很容易实现。然而有些功能要付出更高的代价。例如,为了调配在几个位置上的活动,事务管理的性能比在中心数据库时花费更高,而且甚至抵消许多其他的优点。

二、面向对象数据库

随着计算机技术的发展和应用的普及,人们要求数据库系统不仅能处理简单的数据类型,还要处理如图形、图像、音频、视频等更加复杂的信息。新一代面向对象的数据库系统是解决上述问题的有效途径。

面向对象数据库(Object-Oriented DataBase System, OODBS)是将面向对象技术应用于关系数据库中。将面向对象技术应用到关系数据库中可以有两种方式:一种方式是把面向对象技术中的对象作为关系数据库系统中的一种新的数据模型,关系表中的属性值包含对象指针,对象数据的操作在关系数据库之外进行,把面向对象数据模型(ODM)和关系数据模型(RDM)结合起来,对关系数据库管理系统进行扩充,但对象查询功能受到一定的限制;另一种方式是把面向对象接口添加在关系数据库中,在关系数据库系统中增加一个对象至关系转换器,将上层的面向对象模式转化为关系存储模式,存放到关系数据库中。

面向对象数据库系统的优点:

(1) 能有效地表达客观世界和有效地查询信息:面向对象的方法符合一般人的思维规律,即将现实世界分解成明确的对象,这些对象具有属性和行为。

(2) 可维护性好:在耦合性和内聚性方面,面向对象数据库的性能尤为突出。这使得数据库设计者可在尽可能少影响现存代码和数据的前提下修改数据库结构。这种先进的耦合性和内聚性也简化了在异种硬件平台网络上的分布式数据库的运行。

(3) 能很好地解决 Impedance mismatch 问题。

面向对象数据库系统的缺点:

(1) 技术还不成熟。面向对象数据库技术的根本缺点是这项技术还不成熟,还不广为人知。ODBMS 如今还存在着标准化问题,由于缺乏标准化,许多不同的 ODBMS 之间不能通用。此外,

是修改 SQL 以适应面向对象的程序，还是用新的对象查询语言来代替它，目前还没有解决，这些因素表明随着标准化的出现，ODBMS 还会变化。

(2) 面向对象技术需要一定的训练时间：人们还需要学习一套新的开发方法使之与现有的技术相结合。此外，面向对象系统开发的有关原理才开始具有雏形，还需要一段时间在可靠性、成本等方面令人可接受。

(3) 理论还需完善：从正规的计算机科学方面看，还需要设计出坚实的演算或理论方法来支持 ODBMS 的产品。此外，既不存在一套数据库设计方法学，也没有关于面向对象分析的一套清晰的概念模型，怎样设计独立于物理存储的信息还不明确。

三、数据仓库

随着 90 年代后期 Internet 的兴起与飞速发展，我们进入了一个新的时代，大量的信息和数据迎面而来，用科学的方法去整理数据，从而从不同的视角对企业经营各方面的信息进行精确分析和准确判断比以往更为迫切，实施商业行为的有效性也比以往更受关注。

数据仓库 (Data Warehouse, DW 或 DWH) 是决策支持系统 (DSS) 和联机分析应用数据源的结构化数据环境。数据仓库研究和解决从数据库中获取信息的问题。数据仓库的特征在于面向主题、集成性、稳定性和时变性。



使用这些技术建设的信息系统我们称为数据仓库系统。随着数据仓库技术应用的不断深入，近几年数据仓库技术得到了长足的发展。典型的数据仓库系统有经营分析系统、决策支持系统等。随着数据仓库系统带来的良好效果，各行各业已经能很好地接受“整合数据、从数据中找知识、运用数据知识、用数据说话”等新的关系到改良生产活动各环节、提高生产效率、发展生产力的理念。

数据仓库是面向主题的、集成的、与时间相关的、不可修改的数据集合。这是数据仓库技术特征的定位。其特点是：

(1) 在底层的数据库中保存了大量的事务级细节数据。这些数据是整个 DSS 系统的数据来源。

(2) 数据仓库对底层数据库中的事务级数据进行集成、转换、综合，重新组织成面向全局的数据视图，为 DSS 提供数据存储和组织的基础。

(3) OLAP 从数据仓库中的集成数据出发，构建面向分析的多维数据模型，再使用多维分析方法从多个不同的视角对多维数据进行分析、比较，分析活动从以前的方法驱动转向了数据驱动，分析方法和数据结构实现了分离。

(4) 数据挖掘以数据仓库和多维数据库中的大量数据为基础，自动地发现数据中的潜在模式，并以这些模式为基础自动地作出预测。数据挖掘表明知识就隐藏在日常积累下来的大量数据之中，仅靠复杂的算法和推理并不能发现知识，数据才是知识的真正源泉。数据挖掘技术指出了一条新的发展道路。

本章小结

本章介绍了数据库技术的发展历程、数据库技术的相关基本概念、3 种常用商业数据库及

数据库系统结构, 以及数据库技术研究领域的新技术, 旨在掌握数据库技术的基本概念和数据库系统结构。

习题一

一、选择题

1. 数据库的三级模式之间存在的映象关系正确的是 ()。
A. 外模式/内模式
B. 外模式/模式
C. 外模式/外模式
D. 模式/模式
2. 数据库三级结构从内到外的三个层次为 ()。
A. 外模式、模式、内模式
B. 内模式、模式、外模式
C. 模式、外模式、内模式
D. 内模式、外模式、模式
3. 下列关于数据库系统的正确叙述是 ()。
A. 数据库系统减少了数据冗余
B. 数据库系统避免了一切冗余
C. 数据库系统中数据的一致性是指数据类型一致
D. 数据库系统比文件系统能管理更多的数据
4. 数据库系统和文件系统的主要区别是 ()。
A. 数据库系统复杂, 而文件系统简单
B. 文件系统不能解决数据冗余和数据独立性问题, 而数据库系统能够解决
C. 文件系统只能管理文件, 而数据库系统还能管理其他类型的数据
D. 文件系统只能用于小型机、微型机, 而数据库系统还能用于大型机
5. 数据库三级模式中真正存在的是 ()。
A. 外模式
B. 子模式
C. 模式
D. 内模式
6. 关系数据库中的关键字是指 ()。
A. 能唯一决定关系的字段
B. 不可改动的专用保留字
C. 关键的很重要的字段
D. 能唯一标识元组的属性或属性集合
7. 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
8. 数据库的概念模型独立于 ()。
A. 具体的机器和 DBMS
B. E-R 图
C. 信息世界
D. 现实世界
9. 在数据库系统阶段, 数据是 ()。
A. 有结构的
B. 无结构的
C. 整体无结构, 记录内有结构
D. 整体结构化的
10. () 属于信息世界的模型, 实际上是现实世界到机器世界的一个中间层次。
A. 数据模型
B. 概念模型
C. E-R 图
D. 关系模型

11. 数据库系统的数据独立性是指 ()。
- A. 不会因为数据的变化而影响应用程序
 - B. 不会因为系统数据存储结构与数据逻辑结构的变化而影响应用程序
 - C. 不会因为存储策略的变化而影响存储结构
 - D. 不会因为某些存储结构的变化而影响其他的存储结构
12. 在数据库中,产生数据不一致的根本原因是 ()。
- A. 数据存储空间太大
 - B. 没有严格保护数据
 - C. 未对数据进行完整性控制
 - D. 数据冗余

二、填空题

1. 数据管理技术经历了_____、_____和_____三个阶段。
2. 数据库系统一般是由_____、_____、_____、_____和_____组成。
3. 数据库是长期存储在计算机内、有_____的、可_____的数据集合。
4. DBMS 是指_____, 它是位于_____和_____之间的一层管理软件。
5. 由_____负责全面管理和控制数据库系统。
6. 数据库系统与文件系统的本质区别在于_____。
7. 数据库系统阶段的最大改进是_____。
8. 数据独立性又可分为_____和_____。
9. 根据数据模型的应用目的不同,数据模型分为_____、_____和_____。
10. 数据模型是由_____、_____和_____三部分组成的。
11. 按照数据结构的类型来命名,逻辑模型分为_____、_____和_____。
12. 数据库的模式有_____和_____两方面,前者直接与操作系统或硬件联系,后者是数据库数据的完整表示。
13. 外模式是_____的子集。

三、问答题

1. 试述数据库系统管理数据的主要特点。
2. 试述以下概念:数据、数据模型、数据库、数据库管理系统、数据库系统。
3. 试述关系模型的优缺点。
4. 试述数据库系统的三级模式结构,这种结构的优点是什么?
5. 数据库管理系统有哪些主要功能?