

### 学习目的

1. 熟练掌握数据库技术中的数据与数据管理, 数据模型及类型, 数据库系统与数据库管理系统的区别, 熟悉数据库应用的开发流程。
2. 掌握信息与信息系统的发展及进化过程、类型以及开发方法。

### 引言

随着医院信息化的发展, 电子病历和病案的大量应用, 医疗设备和仪器的数字化, 患者的病史、诊断、检验和治疗的临床信息等都已经以电子形式进行了存储。数据库技术在数据的管理中发挥着非常重要的作用, 是构成医院信息系统的核心技术平台之一。医院数据库中存放着医院所有的公共信息资源, 是医院信息流的中心环节, 是医院中各种信息系统和决策支持系统得以正常工作的基础。而信息系统则是基于数据库技术, 为用户提供各种信息检索与分析服务平台。

## 3.1 数据库技术

随着人们对数据处理要求的不断提高, 基于计算机技术的数据管理技术也不断改进。数据库技术是近年来计算机学科发展最快和应用领域最广的技术之一, 是计算机信息系统与应用系统的核心技术和重要基础。

### 3.1.1 数据与数据管理

数据 (data) 是描述事物的符号记录, 是数据库中存储的基本对象, 也是数据库管理系统处理的基本对象。主要包括数字、文字、符号、图形、图像、声音、视频等多种表现形式。数据不等同于数字, 数字只是一种最简单的数据, 是对数据的一种传统和狭义的理解。虽然数据有多种表现形式, 但都可以经过数字化处理后存入计算机, 在计算机中存储和处理的数据均是以二进制的形式存在的。

信息 (information) 是现实世界事物的存在状态或运动方式的反映, 是经过处理的、能够反映现实世界状态及物理特性的数据。信息是数据的内涵, 也是对数据的语义解释。数据和信息之间存在着十分密切的关系。一方面, 信息是一种特殊形式的数字; 另一方面, 数据又是信息的符号表示, 是信息的现实载体。因此, 很多任务都可以概括为处理数据→获取信息→指导工作的过程。

数据处理 (data processing) 是对数据的采集、存储、检索、加工、变换和传输。数据是对事实、概念或指令的一种表达形式, 可由人工或自动化装置进行处理。数据的形式可以是数字、文

字、图形或声音等。数据经过解释并赋予一定的意义之后，便成为信息。数据处理的基本目的是从大量的、可能是杂乱无章的、难以理解的数据中抽取并推导出有价值、有意义的信息。数据处理是系统工程和自动控制的基本环节。数据处理贯穿于社会生产和社会生活的各个领域。数据处理技术的发展及其应用的广度和深度，极大地影响着人类社会发展的进程。数据处理的目的是将大量的、杂乱无章的、难以理解的数据转换成有价值的、可以理解的信息的过程，即数据+数据处理=信息。

根据处理设备的结构方式、工作方式以及数据的时间、空间分布方式的不同，数据处理有不同的方式。不同的处理方式要求不同的硬件和软件支持。每种处理方式都有自己的特点，应当根据应用问题的实际环境选择合适的处理方式。数据处理主要有四种分类方式：①根据处理设备的结构方式区分，有联机处理方式和脱机处理方式。②根据数据处理时间的分配方式区分，有批处理方式、分时处理方式和实时处理方式。③根据数据处理空间的分布方式区分，有集中式处理方式和分布式处理方式。④根据计算机中央处理器的工作方式区分，有单道作业处理方式、多道作业处理方式和交互式处理方式。

数据指数字、符号、字母和各种文字的集合。数据处理涉及的加工处理比一般的算术运算要广泛得多。

计算机数据处理主要包括 8 个方面：

- 1) 数据采集：采集所需的信息。
- 2) 数据转换：把信息转换成机器能够接收的形式。
- 3) 数据分组：指定编码，按有关信息进行有效的分组。
- 4) 数据组织：整理数据，以便进行处理。
- 5) 数据计算：进行各种算术和逻辑运算，以便得到进一步的信息。
- 6) 数据存储：将原始数据或计算的结果保存起来，供以后使用。
- 7) 数据检索：按用户的要求找出有用的信息。
- 8) 数据排序：把数据按一定要求排成次序。

数据处理系统已广泛地用于各种企业和事业单位，内容涉及薪金支付、票据收发、信贷和库存管理、生产调度、计划管理、销售分析等。它能产生操作报告、金融分析报告和统计报告等。数据处理技术涉及文卷系统、数据库管理系统、分布式数据处理系统等方面的技术。

在数据处理中，最重要的核心问题是数据管理（data management），它是对数据进行组织、分类、编码、存储、检索、维护和应用的过程。

数据管理是利用计算机硬件和软件技术对数据进行有效的收集、存储、处理和应用的过程。其目的在于充分有效地发挥数据的作用。实现数据有效管理的关键是数据组织。随着计算机技术的发展，数据管理经历了人工管理、文件系统、数据库系统三个发展阶段。在数据库系统中所建立的数据结构，更充分地描述了数据间的内在联系，便于数据修改、更新与扩充，同时保证了数据的独立性、可靠性、安全性与完整性，减少了数据冗余，提高了数据共享程度及数据管理效率。

### 3.1.2 数据管理技术发展历程

随着计算机硬件和软件的发展，应用中的数据管理需求逐步发生转变，相应地，数据管理技术的发展也经历了人工管理、文件系统和数据库系统三个主要阶段。

1) 人工管理阶段（20 世纪 50 年代中期以前）：此时的计算机主要用于科学计算，对数据保存的需求尚不迫切。在硬件方面，可用的外存只有卡片、纸带和磁带，而没有磁盘等直接存取设

备, 存储量非常小。在软件方面, 可用的编程语言只有汇编语言, 没有操作系统和管理数据的软件, 应用程序不仅需要考虑数据的逻辑结构, 还要考虑数据的存储结构、存取方法和输入、输出方式, 程序对数据的依赖程度非常强, 如果数据的类型、格式或输入、输出方式等逻辑结构或物理结构发生变化, 应用程序也需要做出相应的修改。

2) 文件系统阶段(20世纪50年代后期至60年代中期): 此时的计算机已不仅仅局限于科学计算, 还大量地应用于信息管理。该阶段是利用计算机进行数据管理的初始阶段, 已经出现了对大量数据进行存储、检索和维护的需求。在硬件方面, 有了磁盘、磁鼓等直接存储设备, 数据可以长期保存在磁盘上。在软件方面, 出现了高级语言和操作系统。并且在操作系统中有了专门管理数据的软件, 称为文件管理系统。

3) 数据库系统阶段(20世纪60年代后期开始): 此时的计算机已经广泛的应用于规模更加庞大, 且数量急剧增长的信息管理。在硬件方面, 出现了大容量磁盘, 使计算机联机存取大量数据成为可能; 同时, 计算机硬件的价格逐渐下降, 而软件的价格却逐渐上升, 系统软件的开发和维护成本逐渐增加。基于文件系统的数据库管理方法已无法满足应用系统开发中的多用户、多应用程序共享数据的需求, 出现了统一管理数据的专门软件系统, 即数据库管理系统(database management system, DBMS)。数据库管理系统的出现与完善, 将程序员从数据管理编程的琐碎工作中解脱出来, 不需要考虑数据改动造成的不一致, 也不用担心应用功能的扩充是否会导致程序重写或者数据结构的变动。

数据管理技术发展的三个阶段是数据库技术从无到有的发展过程, 如果说从人工管理到文件系统, 是计算机开始应用于数据的实质进步, 那么从文件系统到数据库系统, 标志着数据管理技术质的飞跃。20世纪80年代后, 不仅在大、中型计算机上实现并应用了数据管理的数据库技术, 如 Oracle、DB2、Sybase、Informix 等, 在微型计算机上也可使用数据库管理软件, 如常见的 Access、MS-SQL Server、FoxPro 等软件, 数据库技术得到了广泛应用和普及。三个阶段各自的特点如表 3-1 所示。

表 3-1 数据管理三个阶段的对比

人工管理	文件系统	数据库系统
数据不保存在机器中	数据可长期保存在磁盘上	数据结构化
没有管理数据的软件	文件系统管理数据	数据由 DBMS 统一控制
数据无共享	数据共享性差, 冗余大	高共享, 低冗余
数据不具有独立性	数据独立性差	数据独立性高

### 3.1.3 数据库的数据模型

数据模型(data model)是对客观事物特征的数据抽象, 是严格定义的一组概念的集合, 反映了事物与事物之间的联系。按照应用层次的不同, 可以将数据模型分成概念数据模型、逻辑数据模型和物理数据模型三类。

**1. 概念数据模型**(conceptual data model) 简称概念模型, 有时也称之为信息模型, 是面向数据库用户的观点或认识的现实世界的模型, 主要用来描述世界的概念化结构, 是一种用户容易理解的、对现实世界特征的数据抽象描述, 是数据库设计员与用户之间进行交流的语言。应该满足三个方面的要求: ①能够比较真实地模拟现实世界; ②容易为人所理解; ③便于计算机实现。

**2. 逻辑数据模型**(logical data model) 简称逻辑模型, 有时也称之为结构模型, 是用户从数

数据库中看到的模型,是具体的数据库管理系统所支持的数据模型,如层次数据模型、网状数据模型、关系数据模型和面向对象数据模型等。因此,该模型既要面向用户,又要面向系统,主要用于数据库管理系统的实现。逻辑数据模型也是一种图形化的展现方式,一般采用面向对象的设计方法,有效组织来源多样的各种业务数据,使用统一的逻辑语言描述业务。借助相对抽象、逻辑统一且结构稳健的结构,实现数据仓库系统所要求的数据存储目标,支持大量的分析应用,是实现业务智能的重要基础,同时也是数据管理分析的工具和交流的有效手段。

对于企业,逻辑数据模型就是企业基础数据的一部分,它是企业数据资产的全面的、准确的描述,是数据整合的核心或目的。数据整合就是将不同来源的数据整合到一个统一定义、统一形式的逻辑数据模型中。

**3. 物理数据模型** (physical data model) 简称物理模型,是面向计算机物理表示的模型,描述了数据在储存介质上的组织结构,它不但与具体的数据库管理系统有关,而且还与使用的操作系统和计算机硬件有关,是物理层次的模型。每一种逻辑数据模型在实现时都有对应的物理数据模型。数据库管理系统为了保证数据的独立性与可移植性,大部分物理数据模型的实现工作由系统自动完成,而数据库应用的设计者只需要设计索引、聚集等特殊结构。

物理数据模型,提供了系统初始设计所需要的基础元素,以及相关元素之间的关系。即用于存储结构和访问机制的更高层描述,描述数据是如何在计算机中存储的,如何表达记录结构、记录顺序和访问路径等信息。使用物理数据模型,可以在系统层实现数据库。数据库的物理设计阶段必须在此基础上进行详细的后台设计,包括数据库的存储过程、操作、触发、视图和索引表等;至 2013 年,还没有太多的物理数据模型,最常用的物理数据模型分为统一模型和框架存储模型。具有如下功能:

- 1) 可以将数据库的物理设计结果从一种数据库移植到另一种数据库;
- 2) 可以通过反向工程将已经存在的数据库物理结构重新生成物理模型或概念模型;
- 3) 可以定制生成标准的模型报告;
- 4) 可以转换为 OOM (object-oriented method);
- 5) 多种数据库的详细物理设计 (涵盖常用的各种数据库的 DBMS), 并生成数据库对象的 .sql 脚本。

### 3.1.4 数据库的类型

逻辑数据模型决定了数据库管理系统的实现形式,因而按照逻辑数据模型的不同,可以将数据库划分为层次数据库、网状数据库、关系数据库、面向对象数据库、XML 数据库和 NoSQL 数据库等几大类,下面分别加以介绍。

**1. 层次数据库** 层次数据库是一种以层次数据模型为基础的数据库。在层次数据模型中,采用树形结构模型来描述和表示现实世界中的实体及其相互之间的联系。在树形结构中,每一个节点表示一个记录,而各个节点之间的连线表示各个记录之间的关联。这种关联在层次模型中的双亲节点(上层节点)和子女节点(下层节点)之间只能是一对一和一对多的关系。每个记录用于描述实体集中的一个实体,它可以包含若干字段,用于描述实体的属性。在现实世界中,许多实体之间的联系本来就存在一定的层次关系。因此,使用层次模型描述现实世界中的数据关系是比较自然也比较直观的,并且易于理解和表示。

层次数据库的主要优点包括:①数据存取方便,且存取速度快;②数据结构清晰,便于理解;③数据修改和数据库扩展容易实现;④检索关键属性十分方便。主要缺点包括:①结构呆

板, 缺乏灵活性; ②同一属性数据要存储多次, 数据冗余大(如公共边); ③不适合拓扑空间数据的组织。

**2. 网状数据库** 网状数据库是一种以网状数据模型为基础的数据库。在网状数据模型中, 采用图模型来描述和表示现实世界中的实体及其相互之间的联系。在网状结构中, 每个节点表示一个记录, 而各个节点之间的连线表示的是各个记录之间的关联。这种关联在网状模型中的双亲节点(上层节点)和子女节点(下层节点)之间可以是多对多的关系。此时, 一个上层节点可以有多个下层的子女节点, 而一个下层节点也可以有多个上层的双亲节点。每个记录用于描述实体, 它可以包含若干字段, 用于描述实体的属性。与层次数据模型相比, 网状数据模型的结构更加灵活, 表达的关系也更加丰富, 但是同时也带来了处理上的复杂性。

网状数据库的主要优点包括: ①能明确而方便地表示数据间的复杂关系; ②数据冗余小。主要缺点包括: ①网状结构复杂, 增加了用户查询和定位的困难; ②需要存储数据间联系的指针, 使得数据量增大; ③数据的修改不方便(指针必须修改)。

**3. 关系型数据库** 关系型数据库是一种以关系数据模型为基础的数据库。在关系数据模型中, 采用一系列由行和列组成的二维表(关系)来描述和表示现实世界中的实体及其相互之间的联系。一个关系可以用来描述一个实体及其属性, 也用来描述实体之间的联系。每张关系表内部的各项数据项之间可以存在一些关联关系, 不同关系表的数据项之间也可以存在一定的关联关系。在关系数据模型中, 一个关系就是一个二维表(但不是任意一个二维表都能表示一个关系), 二维表名就是关系名。二维表中的每一行称为一个记录或元组; 二维表中的每一列称为一个字段或属性; 属性的取值范围称为域。

关系型数据库的主要优点包括: ①建立在严格的数学概念基础之上, 有关系代数作为语言模型, 有关系理论作为理论基础; ②模型概念单一、规范化, 数据结构简单、清晰, 易懂易用; ③存取路径对用户透明, 具有更高的数据独立性和更好的安全性, 也便于维护, 简化了程序员的工作, 提高了开发的效率。主要缺点包括: ①路径透明也导致了查询效率相对较低; ②为了提高性能, 需要进行查询优化, 增加了编程难度。

**4. 面向对象数据库** 面向对象数据库是一种以面向对象数据模型为基础的数据库。在面向对象数据模型中, 全面支持对象、复合对象、封装、类、继承、重载、滞后联编、多态性等面向对象技术中的基本概念。利用类的概念来描述复杂的对象, 利用对象中封装的状态和方法来模拟对象的基本属性和复杂行为, 利用继承性来实现对象结构和方法的重用。面向对象数据库是类的集合, 一组类可以形成一个类层次。一个面向对象数据库可能有多个类层次。在一个类层次中, 一个类继承其所有超类的全部属性、方法和消息。可以说, 面向对象数据库是面向对象技术与数据库技术结合的产物。面向对象数据库的来源主要分为两大类: ①扩充关系数据模型的对象关系数据库; ②在面向对象语言中嵌入数据库功能或者从底层直接开发并实现的面向对象数据库系统。

面向对象数据库的优点主要包括: ①具有表示和构造复杂对象的能力; ②封装和信息隐藏提供的模块化机制; ③封装、继承和类层次提供的软件重用机制; ④通过滞后联编(late binding)等概念得到系统扩充能力。主要缺点包括: ①缺乏通用数据模型; ②缺乏理论基础; ③缺乏友好的用户界面与环境; ④缺乏有力的查询优化技术。

**5. XML数据库** XML数据库是一种以XML数据模型为基础的数据库。在XML数据模型中, 以节点(元素、属性、备注等)和节点间的相互关系为基础来描述和表示现实世界中的实体及其相互之间的联系。同时, 通过对节点(元素、属性、备注等)的标记进行相关操作来实现对数据的操作。XML数据库主要有三种类型: ①在原有数据库系统上扩充对XML数据的处

理功能,使之能适应 XML 数据存储和查询需要的能处理 XML 的数据库 (XML enabled database, XEDB)。<sup>②</sup>以 XML 文档作为基本的逻辑存储单位,以自然的方式处理 XML 数据,并且针对 XML 的数据存储和查询特点设计专门的数据模型和处理方法的纯 XML 数据库 (native XML database, NXD)。<sup>③</sup>综合上述两种 XML 数据库特点的混合 XML 数据库 (hybrid XML database, HXD)。

XML 数据库的优点主要包括:①能够有效管理半结构化数据;②提供了对标签名称的操作,也提供了对路径的操作;③能够清晰表达数据的层次特征,便于对层次化的数据进行操作。主要缺点包括:①数据冗余度高,不仅占用大量的磁盘空间,还会给操作文件带来困难;②数据访问速度不高,检索效率低下;③对 XML 数据的插入、删除和更新等操作的支持不够完善。

**6. NoSQL 数据库** NoSQL 数据库是非关系型的、分布式的、开源的和水平可扩展的新一代数据库系统,放弃了关系模型和 SQL 语言,具有模式自由、支持简易复制、简单的 API、最终的一致性(非 ACID)和大容量数据等特点。NoSQL 数据库采用的数据模型主要包括:①Key-Value 键值对数据模型;②Document 数据模型;③Column 数据模型。典型的 NoSQL 数据库主要包括:①Google Bigtable,是一个稀疏的、分布式的、持久化存储的多维排序 Map。不支持类似关联这样高级的 SQL 操作,取而代之的是多级映射的数据结构,并支持大规模数据处理、高容错性和自我管理特性,提供 PB 级的存储能力,使用结构化的文件来存储数据,整个集群每秒可处理数百万的读写操作。②Apache Cassandra,是一套开源大规模分布式 Key-Value 存储系统,是一个混合型的非关系的数据库。由一堆数据库节点共同构成一个分布式网络服务,对 Cassandra 的一个写操作,会被复制到其他节点上去,对 Cassandra 的读操作,也会被路由到某个节点上面去读取。③Apache Hbase,是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统,通过 Google Bigtable 的开源实现。④MongoDB,是一个介于关系数据库和 NoSQL 数据库之间的产品,是 NoSQL 数据库当中功能最丰富、最像关系数据库的产品。⑤CouchDB,是用 Erlang 开发的面向文档的数据库系统。CouchDB 不是一个传统的关系数据库,而是面向文档的数据库,其数据存储方式有点类似 lucene 的 index 文件格式,CouchDB 最大的意义在于它是一个面向 Web 应用的新一代存储系统,事实上,CouchDB 的广告语就是:下一代的 Web 应用存储系统。此外,还有 Neo4j、Redis、MemcacheDB 和 Project Voldemort 等。

NoSQL 数据库的主要优点包括:①系统的动态可扩展性好;②操作逻辑简单,具有快速的读写能力;③开源软件,使用成本低廉。主要缺点包括:①不提供 SQL,增加了系统的开发和维护成本;②提供的功能有限,不支持事务等功能;③处于起步阶段,产品成熟度不高。

### 3.1.5 数据库系统与数据库管理系统

数据库系统(database system, DBS)是引入了数据库的计算机系统,是计算机软硬件和数据资源共同组成的系统,目的是为了实有组织地、动态地存储大量关联数据,以方便多用户对数据进行访问。数据库系统一般由数据库、数据库管理系统、应用程序和用户几个部分组成。按照数据库服务器的部署方式,可以将数据库系统分为单用户结构、主从式结构、分布式结构、客户/服务器、浏览器/应用服务器/数据库服务器等几个体系结构。

1) 单用户数据库系统:应用程序、数据库管理系统、数据都安装在一台计算机上,为一个用户独占,不同机器之间不能共享数据,是最简单的数据库系统。

2) 主从式数据库系统:应用程序、数据库管理系统和数据都集中存放在一个主机上,所有处理任务都由该主机来完成,多个用户通过该主机的不同终端并发访问数据库,共享数据资源,多

用于基于大型机的系统。

3) 分布式数据库系统: 数据库中的数据逻辑上是一个整体, 物理上却分布于网络中的不同计算节点上, 每个节点都可以独立处理本地数据库中的数据, 执行局部应用; 也可以同时存取和处理多个异地数据库中的数据, 执行全局应用。

4) 客户机/服务器数据库系统: 网络中的一台或多台计算机专门用于执行数据库管理系统功能, 称为数据库服务器。用户使用客户端向数据库服务器发送请求, 数据库服务器处理后将结果返回。该结构能显著减少网络中的数据传输量, 从而提高系统性能。

5) 浏览器/应用服务器/数据库服务器: 用户通过浏览器访问后台的应用服务器和数据库服务器, 而应用服务器和数据库服务器紧密联系, 可以使数据库应用系统获得更高的性能, 在部署上也可具有更大的灵活性和可扩充性。

对数据库系统的基本要求: ①能够保证数据的独立性。数据和程序相互独立有利于加快软件开发速度, 节省开发费用。②冗余数据少, 数据共享程度高。③系统的用户接口简单, 用户容易掌握, 使用方便。④能够确保系统运行可靠, 出现故障时能迅速排除; 能够保护数据不受非授权者访问或破坏; 能够防止错误数据的产生, 一旦产生也能及时发现。⑤有重新组织数据的能力, 能改变数据的存储结构或数据存储位置, 以适应用户操作特性的变化, 改善由于频繁插入、删除操作造成的数据组织零乱和时空性能变坏的状况。⑥具有可修改性和可扩充性。⑦能够充分描述数据间的内在联系。

数据库系统一般由4个部分组成:

1) 数据库 (database, DB): 指长期存储在计算机内的、有组织、可共享的数据的集合。数据库中的数据按一定的数学模型组织、描述和存储, 具有较小的冗余、较高的数据独立性和易扩展性, 并可为各种用户共享。

2) 硬件: 构成计算机系统的各种物理设备, 包括存储所需的外部设备。硬件的配置应满足整个数据库系统的需要。

3) 软件: 包括操作系统、数据库管理系统及应用程序。数据库管理系统 (database management system, DBMS) 是数据库系统的核心软件, 在操作系统的支持下工作, 它是解决科学地组织和存储数据以及高效获取和维护数据问题的系统软件。其主要功能包括: 数据定义功能、数据操纵功能、数据库的运行管理和数据库的建立与维护。

4) 人员: 主要有4类。第一类为系统分析员和数据库设计人员: 系统分析员负责应用系统的需求分析和规范说明, 他们和用户及数据库管理员一起确定系统的硬件配置, 并参与数据库系统的概要设计。数据库设计人员负责数据库中数据的确定、数据库各级模式的设计。第二类为应用程序员, 负责编写使用数据库的应用程序。这些应用程序可对数据进行检索、建立、删除或修改。第三类为最终用户, 他们利用系统的接口或查询语言访问数据库。第四类用户是数据库管理员 (data base administrator, DBA), 负责数据库的总体信息控制。DBA的具体职责包括: 具体数据库中的信息内容和结构, 决定数据库的存储结构和存取策略, 定义数据库的安全性要求和完整性约束条件, 监控数据库的使用和运行, 负责数据库的性能改进、数据库的重组和重构, 以提高系统的性能。

数据库管理系统是定义、建立、管理、使用和维护数据库的大型软件, 不仅实现了数据库的统一管理和控制, 还保证了数据库的安全性和完整性。同时, 它还提供数据定义语言 (data definition language, DDL) 与数据操作语言 (data manipulation language, DML), 以使用户定义数据库的模式结构与权限约束, 以及实现数据的追加和删除等功能。概括来说, 数据库管理系统的功

能主要包括:

1) 数据定义: DBMS 提供数据定义语言, 用于建立、修改数据库的结构, 用户可以自由定义数据库的模式结构、完整性约束和保密限制等。

2) 数据操作: DBMS 提供数据操作语言, 用户可通过操作语言实现对数据的插入、删除、更新和查询等操作。

3) 数据库运行管理: DBMS 提供运行控制和管理功能, 保证数据库系统的正常运行。主要功能包括: 多用户环境下的并发控制、安全性检查和存取限制控制、完整性检查和执行、运行日志的组织管理、事务的管理和自动恢复等。

4) 数据组织、存储与管理: DBMS 提供分类组织、存储和管理各种数据的功能。决定以何种文件结构和存取方式组织数据, 以及如何实现数据之间的联系, 进而提高存储空间的利用率和数据的存取效率。

5) 数据库的保护: DBMS 提供数据库的恢复、数据库的并发控制、数据库的完整性控制、数据库安全性控制功能, 实现对数据库中数据的保护。此外, 还提供系统缓冲区管理及数据存储的某些自适应调节机制等功能。

6) 数据库的维护: DBMS 提供数据的载入、转换、转储, 数据库的重组和重构, 以及系统性能监控等功能。

7) 通信: DBMS 具有与操作系统的在线处理、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。基于网络的数据库系统, 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。

根据处理对象的不同, 数据库管理系统的层次结构由高级到低级依次为应用层、语言翻译处理层、数据存取层、数据存储层、操作系统。应用层是 DBMS 与终端用户和应用程序的界面层, 处理的对象是各种各样的数据库应用。语言翻译处理层是对数据库语言的各类语句进行语法分析、视图转换、授权检查、完整性检查等。数据存取层处理的对象是单个元组, 它将上层的集合操作转换为单记录操作。数据存储层处理的对象是数据页和系统缓冲区。操作系统是 DBMS 的基础。操作系统提供的存取原语和基本的存取方法, 通常作为与 DBMS 存储层的接口。

数据库管理系统的技术特点如下:

1) 采用复杂的数据模型表示数据结构, 数据冗余小, 易扩充, 实现了数据共享。

2) 具有较高的数据和程序独立性, 数据库的独立性有物理独立性和逻辑独立性。

3) 数据库系统为用户提供了方便的用户接口。

4) 数据库系统提供 4 个方面的数据控制功能, 分别是并发控制、恢复、完整性和安全性。数据库中各个应用程序所使用的数据由数据库系统统一规定, 按照一定的数据模型组织和建立, 由系统统一管理和集中控制。

5) 增加了系统的灵活性。

### 3.1.6 数据库应用开发流程

数据库应用系统的开发是一项软件工程。为了实现一个好的数据库应用系统, 需要认真细致地做好调研、设计、实施和运行维护等各个阶段的工作。概括来说, 数据库应用系统的开发流程可分为系统规划、需求分析、概念设计、逻辑设计、物理设计、系统编码及调试和系统运行及维护 7 个阶段, 各阶段间相互联接, 而且常常需要回溯修正。

1. 系统规划阶段 系统规划的任务是分析应用系统开发的必要性和可行性。在收集整理相关



资料的基础上,分析应用系统与周边系统的关系,对应用系统的规模、地位和作用进行全面的分析和论证;明确应用系统的基本功能,划分数据库支持的范围,分析数据来源、采集方式和采集范围,研究数据结构的特点,估算数据量大小,确立数据处理的基本要求和业务的规范标准;规划人力资源调配,对系统开发维护人员的业务水平提出要求,对用户和操作员的素质做出评估;拟定设备配置方案,论证计算机、网络和其他设备在时间、空间两方面的处理能力,既满足应用需求,又留有余量,选择合适的 OS、DBMS 和其他辅助软件,设备配置方案要在使用要求、系统性能、购置成本和维护代价各方面综合权衡;对系统的开发、运行、维护的成本做出估算;预测系统效益的期望值;拟订开发进度计划,还要对现行工作模式如何向新系统过渡做出具体安排;最后,写出详尽的可行性分析报告和数据库应用系统规划书,以此作为系统后续开发工作的总纲。

**2. 需求分析阶段** 需求分析的任务是分析系统用户的实际应用需求,了解用户业务状况,根据用户的实际需求给出恰当设计。需求分析阶段的重点是调查、收集与分析用户在数据管理中的信息、处理、安全性与完整性要求。并用数据字典和数据流图描述出用户的数据需求和数据处理过程的要求。需求分析的过程大致可分成三步:①需求信息的收集:系统需求信息的收集工作一般以机构设置和业务活动为主线,从高层、中层到低层逐步展开。②需求信息的分析整理:对收集到的信息进行分析、整理工作。用数据流图(data flow diagram, DFD)详细地描述业务流程及业务中的数据联系;用数据字典(data dictionary, DD)详细地描述系统中的全部数据。③需求信息的评审:开发过程中的每一个阶段都要经过评审,确认任务是否全部完成,以便避免或纠正工作中出现的错误和疏漏。评审过程中,需要聘请外部专家参与评审过程,从而保证评审的质量和客观性。评审可能导致开发过程回溯,甚至会反复多次。但是,一定要达到全部预期目标才能让需求分析阶段的工作暂告一个段落。最后,将分析的成果进行总结形成系统的需求说明书,并附以详尽的数据流图和数据字典。

**3. 概念设计阶段** 概念模型设计的任务是在需求分析结果的基础上,对用户的需求进行综合、归纳与抽象,形成一个数据库的概念模型,并给出描述。概念模型不依赖于任何具体的计算机系统,反映了系统信息需求的概念结构。例如,在采用关系型数据库时建立实体-关系图(entity-relationship diagram, ERD),即 E-R 图,或者在采用面向对象的数据库时建立对象的类图(class diagram, CD)。

下面,以关系模型为例介绍概念模型设计的过程:①设计局部概念模型:主要任务包括确定局部概念模型的范围、定义实体、定义联系、确定属性以及逐一画出所有的局部 ER 图,并附以相应的说明文件。②设计全局概念模型:主要任务包括确定公共实体类型、合并局部 ER 图、消除不一致因素、优化全局 ER 图以及画出全局 ER 图,并附以相应的说明文件。③概念模型的评审:主要包括用户评审和开发人员评审两个部分。

**4. 逻辑设计阶段** 逻辑设计的任务是将概念模型转换为计算机中 DBMS 所支持的数据模型,如关系模型或对象模型等。例如,当采用的是关系型数据库时,便要在逻辑设计阶段将以 E-R 图描述的概念模型转换成关系数据模型,也就是要将实体、实体的属性和实体之间的关联转化成关系模式。定义出符合标准化命名要求的数据库表和字段,再根据用户的处理要求和安全性考虑,建立起必要的视图(view)。除此之外,还应该预先考虑到系统的可扩展性,分析出哪些字段将来可能会发生变更。当然,为了进一步提高数据库应用的性能,还要进行数据结构的优化设计工作。

逻辑设计的过程如下:①设计模式与子模式:根据概念模式、用户需求、约束条件和 DBMS 特性,经过初始关系模式建立、规范化处理、模式评价和修正模式的多次迭代和修正,确定最终

模式和子模式。②编写应用程序设计指南：根据设计好的模式、子模式和实际应用需求，规划应用系统架构，描述应用系统的数据存取和处理功能，设计系统的逻辑接口。③编写物理设计指南：根据设计好的模式、子模式和实际应用需求，归纳物理设计阶段所需的一些重要数据。例如，数据库容量、应用处理频率、操作顺序、响应速度等。

**5. 物理设计阶段** 物理设计的任务是在逻辑数据模型的基础上，为应用系统选取一个最合适的物理结构，包括物理数据库结构、存储记录格式、存储记录位置分配及访问方法等。

物理设计过程如下所述：①存储记录结构设计：分析数据的存储要求和实际应用需求，设计数据存取记录格式。②存储空间分配：将存储记录整体合理地分配在物理存储区域。存储空间分配有两个原则：存取频度高的数据尽量安排在快速、随机设备上，存取频度低的数据则安排在速度较慢的设备上；相互依赖性强的数据尽量存储在同一台设备上，且尽量安排在邻近的存储空间上。③访问方法设计：一个访问方法包括存储结构和检索机构两部分。存储结构限定了访问存储记录时可以使用访问路径；检索机构定义了每个应用实际使用的访问路径。④物理设计性能评价：在物理设计过程中，要对应用系统的整体性能进行评价。性能评价包括时间、空间、效率、开销等各个方面。

**6. 系统编码及调试阶段** 系统编码及调试的任务是建立数据库并编制系统应用程序代码，先进行系统的初步调试，再进行系统的联合调试。系统通过联合调试后，进行数据库应用系统的试运行。

系统调试的过程如下：①建立数据库结构：根据逻辑设计和物理设计结果，建立数据库结构。②调试准备：数据库结构建立后，装入实验数据，使数据库进入调试运行阶段。③调试运行：试运行应用程序，对相关功能进行调试与测试。④系统运行准备：装入实际初始数据，为系统运行提供数据，制定数据库的重新组织方案、故障恢复规范和安全规范。

### 7. 系统运行及维护阶段

数据库应用系统经过系统调试和成功试运行后，即可开始投入正式运行。在数据库系统运行过程中，需要继续对它进行评价、调整与修改。主要工作包括：数据库性能监视、分析和改进，数据库备份和恢复，数据库安全性和完整性控制以及数据库结构的重新构造等工作，使得应用系统能为用户更好地提供服务。

数据库正式投入运行后，运行维护阶段的主要工作是：①维护数据库的安全性与完整性。按照制定的安全规范和故障恢复规范，在系统出现安全问题时，及时调整授权和更改密码；及时发现系统运行时出现的错误，迅速修改，确保系统正常运行。把数据库的备份和转储作为日常的工作，一旦发生故障，立即使用数据库的最新备份予以恢复。②监察系统的性能。运用DBMS提供的性能监察与分析工具，时刻监控系统运行情况。当数据库的存储空间或响应时间等性能下降时，立即进行分析研究，找出原因，及时采取措施改进。③扩充系统的功能。在维持原有系统功能和性能的基础上，适应环境和需求的变化，采纳用户的合理意见，对原有系统进行扩充，增加新的功能。

## 3.2 信息系统

信息、物质和能源是人类社会发展的三大资源，人类在开发利用物质和能源上已经取得了巨大的成功。随着以计算机技术、通信技术、网络技术为代表的现代信息技术的飞速发展，人们越来越重视信息技术对传统产业的改造以及对信息资源的开发和利用，信息系统在管理应用中的应用水平日趋提高。