

数据库的概念与用途

数据库的概念

什么是数据库呢?当人们从不同的角度来描述这一概念时就有不同的定义(当然是描述性的)。例如,称数据库是一个“记录保存系统”(该定义强调了数据库是若干记录的集合)。又如称数据库是“人们为解决特定的任务,以一定的组织方式存储在一起的相关的数据的集合”(该定义侧重于数据的组织)。更有甚者称数据库是“一个数据仓库”。当然,这种说法虽然形象,但并不严谨。严格地说,数据库是“按照数据结构来组织、存储和管理数据的仓库”。在经济管理的日常工作中,常常需要把某些相关的数据放进这样“仓库”,并根据管理的需要进行相应的处理。例如,企业或事业单位的人事部门常常要把本单位职工的基本情况(职工号、姓名、年龄、性别、籍贯、工资、简历等)存放在表 20.6.3 中,这张表就可以看成是一个数据库。有了这个“数据仓库”我们就可以根据需要随时查询某职工的基本情况,也可以查询工资在某个范围内的职工人数等等。这些工作如果都能在计算机上自动进行,那我们的人事管理就可以达到极高的水平。此外,在财务管理、仓库管理、生产管理中也需建立众多的这种“数据库”,使其可以利用计算机实现财务、仓库、生产的自动化管理。

J. Martin 给数据库下了一个比较完整的定义:数据库是存储在一起的相关数据的集合,这些数据是结构化的,无有害的或不必要的冗余,并为多种应用服务;数据的存储独立于使用它的程序;对数据库插入新数据,修改和检索原有数据均能按一种公用的和可控制的方式进行。当某个系统中存在结构上完全分开的若干个数据库时,则该系统包含一个“数据库集合”。

数据库的优点

使用数据库可以带来许多好处:如减少了数据的冗余度,从而大大地节省了数据的存储空间;实现数据资源的充分共享等等。此外,数据库技术还为用户提供了非常简便的使用手段使用户易于编写有关数据库应用程序。特别是近年来推出的微型计算机关系数据库管理系统 dBASE,操作直观,使用灵活,编程方便,环境适应广泛(一般的十六位机,如 IBM/PC/XT,国产长城 0520 等均可运行种软件),数据处理能力极强。数据库在我国正得到愈来愈广泛的应用,必将成为经济管理的有力工具。

数据库是通过数据库管理系统(DBMS-DATA BASE MANAGEMENT SYSTEM)软件来实现数据的存储、管理与使用的 dBASE 就是一种数据库管理系统软件。

数据库结构与数据库种类

数据库通常分为层次式数据库、网络式数据库和关系式数据库三种。而不同的数据库是按不同的数据结构来联系和组织的。

1. 数据结构模型

(1) 数据结构

所谓数据结构是指数据的组织形式或数据之间的联系。如果用 D 表示数据,用 R 表示数据对象之间存在的关系集合,则将 $DS = (D, R)$ 称为数据结构。例如,设有一个电话号码簿,它记录了 n 个人的名字和相应的电话号码。为了方便地查找某人的电话号码,将人名和号码按字典顺序排列,并在名字的后面跟随着对应的电话号码。这样,若要查找某人的电话号码(假定他的名字的第一个字母是 Y),那么只须查找以 Y 开头的那些名字就可以了。该例中,数据的集合 D 就是人名和电话号码,它们之间的联系 R 就是按字典顺序的排列,其相应的数据结构就是 $DS = (D, R)$, 即一个数组。

(2) 数据结构种类

数据结构又分为数据的逻辑结构和数据的物理结构。数据的逻辑结构是从逻辑的角度(即数据间的联系和组织方式)来观察数据,分析数据,与数据的存储位置无关。数据的物理结构是指数据在计算机中存放的结构,即数据的逻辑结构在计算机中的实现形式,所以物理结构也被称为存储结构。本节只研究数据的逻辑结构,并将反映和实现数据联系的方法称为数据模型。

目前,比较流行的数据模型有三种,即按图论理论建立的层次结构模型和网状结构模型

以及按关系理论建立的关系结构模型。

2. 层次、网状和关系数据库系统

(1) 层次结构模型

层次结构模型实质上是一种有根结点的定向有序树(在数学中“树”被定义为一个无回的连通图)。这个组织结构图像一棵树,校部就是树根(称为根结点),各系、专业、教师、学生等为枝点(称为结点),树根与枝点之间的联系称为边,树根与边之比为 1:N,即树根只有一个,树枝有 N 个。

(2) 网状结构模型

例如某医院医生、病房和病人之间的联系。即每个医生负责治疗三个病人,每个病房可住一到四个病人。如果将医生看成是一个数据集合,病人和病房分别是另外两个数据集合,那么医生、病人和病房的比例关系就是 M:N:P(即 M 个医生, N 个病人, P 间病房)。这种数据结构就是网状数据结构,它的一般结构模型。记录 $R_i (i=1, 2, 8)$ 满足以下条件:

①可以有一个以上的结点无双亲(如 R_1, R_2, R_3)。

②至少有一个结点有多于一个以上的双亲。在“医生、病人、病房”例中,“医生集合有若干个结点(M 个医生结点)无“双亲”,而“病房”集合有 P 个结点(即病房),并有一个以上的“双亲”(即病人)。

按照网状数据结构建立的数据库系统称为网状数据库系统,其典型代表是 DBTG(Data Base Task Group)。用数学方法可将网状数据结构转化为层次数据结构。

(3) 关系结构模型

关系式数据结构把一些复杂的数据结构归结为简单的二元关系(即二维表格形式)。例如某单位的职工关系就是一个二元关系。这个四行六列的表格的每一列称为一个字段(即属性),字段名相当于标题栏中的标题(属性名称);表的每一行是包含了六个属性(工号、姓名、年龄、性别、职务、工资)的一个六元组,即一个人的记录。这个表格清晰地反映出该单位职工的基本情况。

通常一个 m 行、n 列的二维表格的结构。

表中每一行表示一个记录值,每一列表示一个属性(即字段或数据项)。该表一共有 m 个记录。每个记录包含 n 个属性。

作为一个关系的二维表,必须满足以下条件:

- (1)表中每一列必须是基本数据项(即不可再分解)。
- (2)表中每一列必须具有相同的数据类型(例如字符型或数值型)。
- (3)表中每一列的名字必须是唯一的。
- (4)表中不应有内容完全相同的行。
- (5)行的顺序与列的顺序不影响表格中所表示的信息的含义。

由关系数据结构组成的数据库系统被称为关系数据库系统。

在关系数据库中,对数据的操作几乎全部建立在一个或多个关系表格上,通过对这些关系表格的分类、合并、连接或选取等运算来实现数据的管理。dBASEII 就是这类数据库管理系统的典型代表。对于一个实际的应用问题(如人事管理问题),有时需要多个关系才能实现。用 dBASEII 建立起来的一个关系称为一个数据库(或称数据库文件),而把对应多个关系建立起来的多个数据库称为数据库系统。dBASEII 的另一个重要功能是通过建立命令文件来实现对数据库的使用和管理,对于一个数据库系统相应的命令序列文件,称为该数据库的应用系统。因此,可以概括地说,一个关系称为一个数据库,若干个数据库可以构成一个数据库系统。数据库系统可以派生出各种不同类型的辅助文件和建立它的应用系统。

数据库的要求与特性

为了使各种类型的数据库系统能够充分发挥它们的优越性,必须对数据库管理系统的使用提出一些明确的要求。

1. 建立数据库文件的要求

(1)尽量减少数据的重复,使数据具有最小的冗余度。计算机早期应用中的文件管理系统,由于数据文件是用户各自建立的,几个用户即使有许多相同的数据也得放在各自的文件中,因而造成存储的数据大量重复,浪费存储空间。数据库技术正是为了克服这一缺点而出现的,所以在组织数据的存储时应避免出现冗余。

(2) 提高数据的利用率，使众多用户都能共享数据资源。

(3) 注意保持数据的完整性。这对某些需要历史数据来进行预测、决策的部门(如统计局、银行等)特别重要。

(4) 注意同一数据描述方法的一致性，使数据操作不致发生混乱。如一个人的学历在人事档案中是大学毕业，而在科技档案中却是大学程度，这样就容易造成混乱。

(5) 对于某些需要保密的数据，必须增设保密措施。

(6) 数据的查找率高，根据需要数据应能被及时维护。

2. 数据库文件的特征

无论使用哪一种数据库管理系统，由它们所建立的数据库文件都可以看成是具有相同性质的记录的集合，因而这些数据库文件都有相同的特性：

(1) 文件的记录格式相同，长度相等。

(2) 不同的行是不同的记录，因而具有不同的内容。

(3) 不同的列表示不同的字段名，同一列中的数据的性质(属性)相同。

(4) 每一行各列的内容是不能分割的，但行的顺序和列的顺序不影响文件内容的表达。

3. 文件的分类

对文件引用最多的是主文件和事物文件。其他的文件分类还包括表文件、备份文件、档案的输出文件等。下面将讲述这些文件。

(1) 主文件。主文件是某特定应用领域的永久性的数据资源。主文件包含那些被定期存取以提供信息和经常更新以反映最新状态的记录。典型的主文件有库存文件、职工主文件和收帐主文件等。

(2) 事务文件。事务文件包含着作为一个信息系统的数据库活动(事务)的那些记录。这些事务被分批以构成事务文件。例如，从每周工资卡上录制下来的数分批存放在一个事务文件上，然后对照工资清单文件进行处理以便打印出工资支票和工资记录簿。

(3) 表文件。表文件是一些表格。之所以单独建立表文件而不把表设计在程序中是为了便于修改。例如，一个公用事业公司的税率表或国内税务局的税率就可以存储在表中文件。

(4) 备用文件。备用文件是现有生产性文件的一个复制品。一旦生产性文件受到破坏，利用备用文件就可以重新建立生产性文件。

(5) 档案文件。档案文件不是提供当前处理使用的，而是保存起来作为历史参照的。例如，国内税务局(IRS)可能要求检查某个人最近 15 年的历史。实际上，档案文件恰恰是在给定时间内工作的一个“快照”。

(6) 输出文件。输出文件包含将要打印在打印机上的、显在屏幕上的或者绘制在绘图仪上的那些信息的数值映象。输出文件可以是“假脱机的”(存储在辅存设备上)，当输出设备可用时才进行实际的输出。