

第 1 章 数据库基本认知

学习目标

- 了解数据库的基本知识。
- 了解结构化查询语言的特点。
- 掌握 MySQL 数据表数据库的安装与配置方法。
- 能够使用多种方式连接启动和运行 MySQL 服务器。

当前人类社会正处于信息时代,数据以惊人的速度增长,如何对其进行有效的管理和利用是人类社会面临的重大课题。数据库技术正是为了适应信息社会的需要而发展起来的一门综合性数据管理技术。为了更好地掌握数据库技术,需要知道数据库是如何存储的、数据库要遵守什么规则、数据库技术都经历了哪些阶段、数据库管理系统会提供哪些功能,以及当前流行的数据库有哪些等。要了解这些内容,首先要认识数据库。

1.1 数据库概述

无论是传统的软件,还是互联网网站,或者是移动端的应用,都要处理数据。数据库技术研究如何有效地管理和存取大量的数据资源。随着计算机技术的不断发展,数据库技术已成为计算机科学的重要分支。今天,数据库技术不仅应用于事务处理,还进一步应用于情报检索、人工智能、专家系统、计算机辅助设计等领域。数据库的建设规模、数据库信息量的规模及使用频率已成为衡量一个企业、一个组织乃至一个国家信息化程度高低的重要标志。

在大数据时代,数据库技术与人们的生活息息相关。下面以小张同学开学第一天的学习生活为例来说明这样一个时代背景。早上起床,小张想知道今天要上哪些课程,所以他登录了学校的“教务管理系统”,在“选课数据库”中查询到他今天的上课信息,包括课程名称、上课时间、上课地点、授课教师等;接着,小张走进食堂买早餐,当他刷餐卡时,学校的“就餐管理系统”根据他的卡号在“餐卡数据库”里读取“卡内金额”,并将“消费金额”等信息写入数据库;课后,小张去图书馆借书,首先他登录“图书管理系统”,通过“图书数据库”查询书籍信息选择要借阅的书籍,然后当他办理借阅手续时,该系统将小张的借阅信息(包括借书证号、姓名、图书编号、借阅日期等)写入数据库;晚上,小张去超市购物,“超市结算系统”根据条码到“商品数据库”中查询物品名称、单价等信息并计算结算金额、找零等数据。由此可见,数据库技术的应用已经深入人们生活的方方面面。研究如何科学地管理数据来为人们提供可共享的、安全的、可靠的数据显得非常重要。

1.1.1 数据与数据处理

1. 数据

数据是人们反映客观世界而记录下来的可以鉴别的物理符号。今天,数据的概念不再仅局限于狭义的数值数据,还包括文字、声音、图形等一切能被计算机接收且处理的符号。

2. 数据处理

数据是重要的资源,人们对收集的大量数据进行加工、整理、转换,可以从中获取有价值的信息。数据处理正是将数据转换成信息的过程,是对各种形式的数据进行收集、存储、加工和传播的一系列活动的总和。

3. 数据管理

数据处理的中心问题是数据管理。数据管理是对数据进行分类、组织、编码、存储、检索与维护的操作。

4. 数据处理技术

数据处理技术研究如何科学地组织和存储数据,如何高效地获取和处理数据。数据库技术是数据处理的核心技术,它是计算机辅助管理数据的方法,是通过研究数据库的结构、存储、设计、管理以及应用的基本理论和实现方法,并利用这些理论来实现对数据库中的数据进行处理、分析和理解的技术。

1.1.2 数据库技术的发展

自从 1946 年第一台电子计算机 ENIAC 诞生以来,计算机的应用范围就迅速扩展。从最初单纯的科学计算到复杂的事务处理再到决策支持甚至人工智能,在这个过程中,计算机所处理的数据呈几何级急剧增长,数据间关系的复杂性也随之增加。

为了人们能方便而充分地利用这些宝贵的信息资源,经过不断的发展,逐渐形成了数据库技术。数据库于 20 世纪 60 年代末产生,几十年来,不断得到迅速的发展,几乎已经渗透到计算机应用的每一个领域。

当我们回顾数据库的发展历史,会由衷地感谢那些本领域的前辈们所做出的重大贡献,而其中的三件大事更被看成数据库技术基本形成的标志。它们是:1968 年 IBM 公司推出的信息管理系统 IMS(Information Management System);1969 年美国 CODASYL 的数据库任务组(DBTG)提出网状数据库模型的数据库规范,并于其后公布了《DBTG 报告》;1970 年 IBM 公司的高级研究员 E.F.Codd 发表论文《大型共享数据库数据的关系模型》,奠定了关系数据库的基础理论。

数据库技术的发展大致经过三个阶段:人工管理阶段、文件系统阶段和数据库系统阶段。

1. 人工管理阶段

人工管理阶段(20 世纪 50 年代以前)计算机应用的前景是：计算机主要用于科学计算，从硬件上看，外存只有磁带、卡片、纸带，没有磁盘等直接存取的存储设备；从软件上看，没有操作系统，没有管理数据的软件，数据处理的方式是批处理。

这一阶段的数据管理具有以下几个特点。

(1) 数据不保存。因为计算机主要应用于科学计算，一般不需要将数据长期保存。只是在计算某一课题时将数据输入，用完就撤走，不仅对用户数据这样处理，有时对系统软件也是如此。

(2) 没有专用的软件对数据进行管理。程序员不仅要规定数据的逻辑结构，而且要在程序中设计物理结构。

(3) 只有程序(Program)的概念，没有文件(File)的概念。数据的组织方式必须由程序员自行设计。

(4) 一组数据对应一个程序，数据是面向应用的。即使两个应用程序涉及某些相同的数据，也必须各自定义，无法互相利用、互相参照。所以程序和程序之间有大量重复的数据。

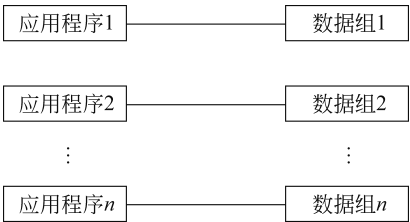


图 1-1 人工管理阶段数据管理的特点

这个时期程序与数据的关系如图 1-1 所示。

2. 文件系统阶段

文件系统阶段(20 世纪 50 年代后期到 60 年代中期)计算机不仅用于科学计算，还大量用于管理。外存储器有了磁盘、磁鼓等直接存取的存储设备。在软件方面，操作系统中已经有了专门的管理数据软件，一般称为文件系统(有时称为“信息处理模块”)。从处理方式角度讲，不仅有了文件批处理，而且能够联机实时处理。

这一阶段的数据管理形成了以下几个特点。

(1) 计算机大量用于数据处理。数据需要长期保留在外存上反复处理，即经常需要对文件进行查询、修改、插入和删除等操作。

(2) 用软件进行数据管理。程序和数据之间有软件提供存取方法进行转换，有共同的数据查询、修改的管理模块。

(3) 文件组织已多样化。有索引文件、链接文件和直接存取文件等。文件之间是独立的，联系要通过程序去构造。

(4) 数据不再属于某个特定的程序，可以重复使用。但程序仍然基于特定的物理结构和存取方法，因此数据结构与程序之间的依赖关系并未根本改变。

上述特点比人工管理阶段有了很大的改进，但随着数据量的急剧增加，数据管理规模的扩大，文件系统显露出以下三个缺点。

- (1) 数据冗余度(Redundancy)大。
- (2) 数据不一致性。
- (3) 数据和程序缺乏独立性。

文件系统中的文件，是为某一特定应用服务的，因此，要想对现有的数据再增加一些新

的应用是很困难的。一旦数据的逻辑结构改变,就必须修改应用程序,修改文件结构的定义。而应用程序的改变,如应用程序所使用的高级语言的变化等,也将影响文件的数据结构的改变。这个时期的数据和程序缺乏独立性,其相互关系如图 1-2 所示。

3. 数据库系统阶段

数据库系统阶段(20 世纪 60 年代后期开始)计算机应用越来越广泛,数据量急剧增长,而且数据的共享要求越来越高。这时,有大容量的磁盘,联机实时处理需求更多了,并开始提出和考虑分布处理。另外,软件价格开始上升,硬件价格下降,使编制和维护系统软件及应用程序所需的成本相对增加。在这种情况下,为了解决多用户、多应用共享数据的需求,使数据为尽可能多的应用服务,出现了数据库这样的数据管理技术。这一阶段的数据管理呈现以下特点。

(1) 有复杂的数据模型(结构)。数据模型描述数据本身的特点和数据之间的联系,这种联系通过存取路径实现。

(2) 有较高的数据独立性。数据的物理结构与逻辑结构间差别可以很大。用户以简单的逻辑结构操作数据而无须考虑数据的物理结构。数据库的结构分成用户的逻辑结构(外模式)、逻辑结构、物理结构(内模式)三级。

(3) 提供方便的用户接口。数据库系统为用户提供了方便的用户接口,用户可使用查询语言或简单的终端命令操作数据库,也可以用程序方式操作数据库。

(4) 提供数据控制功能。数据库管理系统提供以下四方面的数据控制功能。

① 数据完整性。保证数据库始终包含正确的数据,用户可以设计一些完整性规则以确保数据值的正确性。

② 数据安全性。保证数据的安全和机密,防止数据丢失或被窃取。

③ 数据库的并发控制。避免并发程序之间的相互干扰,防止数据库被破坏,杜绝提供给用户不正确的数据。

④ 数据库的恢复。在数据库被破坏或数据不可靠时,系统有能力把数据库恢复到最近某时刻的正确状态。

这一阶段程序和数据的关系如图 1-3 所示。

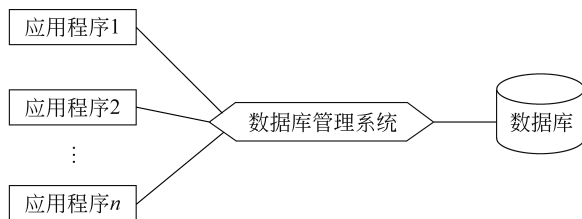


图 1-3 程序和数据的关系

综上所述,可以说数据库是个通用的综合性的数据集合,它可以供各种用户共享,且具有最小的冗余度、较高的数据与程序的独立性。

由于多种程序并发地使用数据库,为了能有效、及时地处理数据,并提供安全性和完整性,必须有一个软件系统在建立、运用和维护时对数据库进行统一控制,这种软件系统称为数据库管理系统(DBMS)。

1.1.3 数据库系统

数据库系统(Database System)是采用数据库技术构建的复杂计算机系统。它不是单纯的数据库或数据库管理系统,而是一种综合了计算机硬件、软件、数据集合和数据库管理员,遵循数据库规则,向用户和应用程序提供信息服务的集成系统。因此,数据库、软件系统、硬件系统、数据库管理员被称为数据库系统的四要素。

数据库系统的四个要素构成有机的整体,它们之间互相紧密配合和依靠,为各类用户提供信息服务。

1. 数据库

数据库是按一定结构组织的、各种应用相关的所有数据的集合。它包含了数据库管理系统处理的全部数据。其内容主要分为两个部分:一是物理数据库,记载了所有数据;二是数据字典,描述了不同数据之间的关系和数据组织的结构。

2. 软件系统

软件系统包括数据库管理系统(DBMS)、操作系统(Operating System)、应用程序开发工具及各种应用程序。数据库管理系统是整个数据库系统的核心,所有对数据库的操作,如查询、增加、删除、新建、更新等都要通过 DBMS 的分析,由 DBMS 调用操作系统的相关部分来执行。也就是说,数据库管理系统是位于用户与操作系统之间的一层数据管理软件。操作系统创建并维持了 DBMS 的运行环境。

1) 数据库管理系统

DBMS 的主要功能包括以下几个方面。

(1) 数据定义功能。提供数据定义语言(Data Definition Language,DDL),用户通过它可以方便地定义数据。

(2) 数据操纵功能。提供数据操纵语言(Data Manipulation Language,DML)实现对数据库的基本操作。

(3) 数据运行管理。这是 DBMS 运行时的核心部分,包括并发控制、安全检查、完整性约束条件的检查和执行、数据库的内部维护(如索引、数据字典的自动维护)等。所有的数据库操作都要在这些控制程序的统一管理下进行。

(4) 数据库的建立和维护功能。它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,数据库的重组功能和性能监测、分析功能等。这些功能通常由一些实用程序完成。

2) 软件系统的工作过程

图 1-4 详细描述了应用程序通过数据管理系统和操作系统访问(读取)数据库的过程。

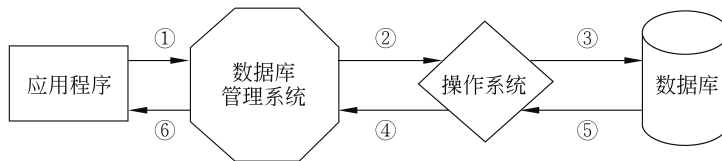


图 1-4 访问数据库

① 应用程序向数据库管理系统发出读取数据的请求,此请求在用户程序中是一条读取记录的数据操作(DML)语句。

② 数据库管理系统接到应用程序的请求,判断此操作是否在用户权限范围内,如果是,则将 DML 语句转换成数据库内部记录的格式,确定要读取记录在存储器上的物理地址,然后向操作系统发送读取记录的命令及相关的地址信息。

③ 操作系统执行该命令,打开数据库文件,按照上一步提供的地址信息,读取相应的记录。

④ 二进制记录信息已经从数据库中读出,并发送到操作系统的系统缓冲区,供 DBMS 调用。

⑤ DBMS 从系统缓冲区中调用二进制的系统信息,并将取得的信息转换成应用程序所要求的格式。

⑥ 应用程序接收从 DBMS 中取得的所需数据,继续运行下一步操作。

需要注意的是,一种数据库一般只支持一种或两种操作系统,不过,近年来,人们也越来越认识到跨平台作业的重要性,许多大型数据库都同时支持多种操作系统。

3. 硬件系统

硬件系统是指支持数据库系统运行的全部硬件,一般由中央处理器、内存、外存等硬件设备组成。不同的数据库对硬件系统的要求有所不同,普通的桌面数据库一般只能运行在个人计算机上,而一些大型数据库如 Oracle、Sybase 等,则对硬件系统有较高的要求。另外,如果是联网的数据库系统,则还需要购买配套的网络设备。

4. 数据库管理员

数据库管理员(Database Administrator, DBA),是专门负责数据库系统设计、运行和维护的专职人员。他们在数据库系统的规划、设计、运行阶段都担负着重要的任务。在数据库规划设计阶段,DBA 创建数据模式,并根据此数据模式决定数据库的内容和结构;在运行维护阶段,DBA 对不同的用户授予不同的权限,并监督用户对数据库的使用;在管理方面,DBA 运用数据库管理系统提供的实用程序进行数据库的装配、维护、日志、恢复、统计分析等工作,运用数据字典了解系统的运行情况,并将系统的相关变化记录到数据字典中。

数据库管理员的工作十分繁重而复杂,尤其是大型数据库的 DBA 往往是由几个人组成的小组协同工作。数据库管理员的职责又十分重要,直接关系到数据库系统的顺利运作。所以,DBA 必须由专业知识较深、经验较丰富的专业人士来担任。

1.1.4 数据库基本体系结构

人们为数据库设计了一个严谨的体系结构,数据库领域公认的标准结构是三级模式结构,它包括外模式、概念模式、内模式,有效地组织、管理数据,提高了数据库的逻辑独立性和物理独立性。

1. 外模式

外模式又称子模式或用户模式,对应于用户级数据库。它是某个或某几个用户所看到的数据库的数据视图,是与某一应用有关的数据的逻辑表示。外模式是从模式导出的一个子集,包含模式中允许特定用户使用的那部分数据。用户可以通过外模式描述语言来描述、定义对应于用户的数据记录(外模式),也可以利用数据操纵语言(Data Manipulation Language,DML)对这些数据记录进行操作。外模式反映了数据库系统的用户观。

2. 概念模式

概念模式又称模式或逻辑模式,对应于概念级数据库。它是由数据库设计者综合所有用户的数据,按照统一的观点构造的全局逻辑结构,是对数据库中全部数据的逻辑结构和特征的总体描述,是所有用户的公共数据视图(全局视图)。它是由数据库管理系统提供的数据库模式描述语言(Data Description Language,DDL)来描述、定义的。概念模式反映了数据库系统的整体观。

3. 内模式

内模式又称存储模式,对应于物理级数据库。它是数据库中全体数据的内部表示或底层描述,是数据库最低一级的逻辑描述,描述了数据在存储介质上的存储方式和物理结构,对应着实际存储在外存储介质上的数据库。内模式是由内模式描述语言来描述、定义的。内模式反映了数据库系统的存储观。

总体而言,概念模式描述数据的全局逻辑结构,外模式涉及的是数据的局部逻辑结构,即用户可以直接接触到的数据的逻辑结构,而内模式更多的是由数据库系统内部实现的。

为了实现三个抽象层次的联系和转换,数据库系统在这三级模式中提供了映像机制,即外模式/概念模式映像和概念模式/内模式映像。其中,外模式/概念模式映像定义某个外模式和概念模式之间的对应关系,使得当概念模式改变时,通过外模式/概念模式映像的相应改变来保证模式不变。另外,概念模式/内模式映像定义数据的逻辑结构和存储结构之间的对应关系,使得当数据库的存储结构改变时,通过概念模式/内模式映像的相应改变来维持模式不变。

数据库三级模式如图 1-5 所示。

上述的概念虽然比较复杂,但是读者通过了解数据库的基本结构,可以在规划设计数据库时对整个过程有一个更全面的认识。

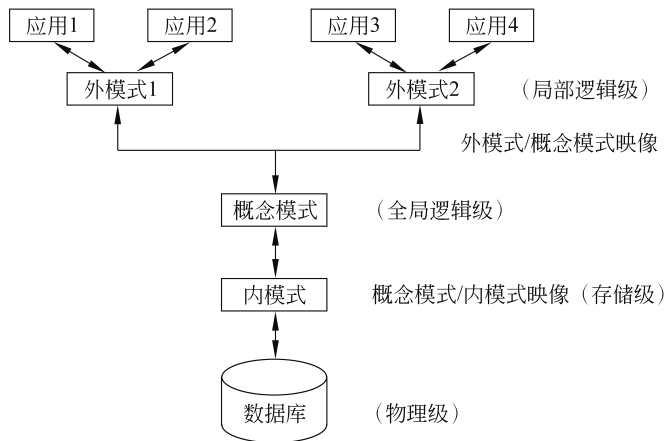


图 1-5 数据库三级模式

1.1.5 结构化查询语言

SQL(Structured Query Language)叫作结构化查询语言。SQL 是一种标准的关系数据库查询语言,它用于对关系数据库中的数据进行存储、查询、更新等操作。正因为 SQL 的标准化,所以绝大多数数据库系统都支持 SQL,而且在不同的数据库管理系统中,SQL 的差别很小。只要掌握了标准 SQL,就可以应用到任何其他关系数据库管理系统上。

SQL 是一种交互式的计算机操作语言,也是一种数据库编程语言,它不仅能够在交互式环境下提供对数据库的各种访问操作,而且作为一种分布式数据库语言用于数据库应用的开发。

SQL 由三部分组成,包括数据库生成、维护和安全性问题的所有内容。

(1) 数据定义语言(Data Definition Language,DDL): DDL 提供完整定义数据库必需的所有内容,包括数据库生成后的结构修改、删除功能。DDL 是 SQL 中用来生成、修改、删除数据库基本要素的部分。

(2) 数据操作语言(Data Manipulation Language,DML): DML 是 SQL 中运算数据库的部分,它是对数据库中的数据进行输入、修改及提取的有力工具。

(3) 数据控制语言(Data Control Language,DCL): DCL 提供的防护措施是数据库安全性所必需的。SQL 通过限制可以改变数据库的操作来保护它,包括事件、特权等。

SQL 共有 9 个核心动词,表 1-1 列出了这 9 个动词和它们分别所属的功能。

表 1-1 SQL 核心动词和它们分别所属的功能

功 能	动 词
数据定义	CREATE、DROP、ALTER
数据操作	SELECT、UPDATE、DELETE、INSERT
数据控制	GRANT、REVOKE

1.1.6 常见的数据库产品

随着数据库技术的发展,数据库产品越来越多,如 Oracle、SQL Server、DB2、MongoDB、MySQL 等。

1. Oracle 数据库

Oracle 数据库管理系统是由甲骨文(Oracle)公司开发的,在数据库领域一直处于领先地位。目前,Oracle 数据库覆盖了大、中、小型计算机等几十种计算机型,成为世界上使用最广泛的关系数据管理系统(由二维表及其之间的关系组成的一种数据库)之一。

Oracle 数据库管理系统采用标准的 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,而且它可以在 VMS、DOS、UNIX、Windows 等操作系统下工作。不仅如此,Oracle 数据库管理系统还具有良好的兼容性、可移植性和可连接性。

2. SQL Server 数据库

SQL Server 是由微软公司开发的一种关系数据库管理系统,它已广泛应用于电子商务、银行、保险、电力等行业。

SQL Server 提供了对 XML 和 Internet 标准的支持,具有强大的、灵活的、基于 Web 的应用程序管理功能。而且界面友好、易于操作,深受广大用户的喜爱,但它只能在 Windows 平台上运行,并对操作系统的稳定性要求较高,因此很难应对日益增长的用户数量。

3. DB2 数据库

DB2 数据库是由 IBM 公司研制的一种关系数据库管理系统,主要应用于 OS/2、Windows 等平台,具有较好的可伸缩性,可支持从大型计算机到单用户环境。

DB2 支持标准的 SQL,并且提供了高层次的数据利用性、完整性、安全性和可恢复性,以及从小规模到大规模应用程序的执行能力,适用于海量数据的存储,但相对于其他数据库管理系统而言,DB2 的操作比较复杂。

4. MongoDB 数据库

MongoDB 是一个介于关系数据库和非关系数据库之间的产品,是非关系数据库中功能最丰富、最像关系数据库的。它支持的数据结构非常松散,是类似 JSON 的 bson 格式,因此可以存储比较复杂的数据类型。

Mongo 数据库管理系统最大的特点是它支持的查询语言非常强大,其语法有点类似于面向对象的查询语言,可以实现类似关系数据库单表查询的绝大部分功能,而且支持对数据建立索引。不仅如此,它还是一个开源数据库,并且具有高性能、易部署、易使用、存储数据非常方便等特点。对于大数据量、高并发、弱事务的互联网应用,MongoDB 完全可以满足 Web 2.0 和移动互联网的数据存储需求。

5. MySQL 数据库

MySQL 数据库管理系统是由瑞典的 MySQL AB 公司开发的,但是几经辗转,现在成为 Oracle 公司的产品。它是以“客户/服务器”模式实现的,是一个多用户、多线程的小型数据库服务器。MySQL 是开源数据的,任何人都可以获得该数据库的源代码并修正 MySQL 的缺陷。

MySQL 具有跨平台的特性,它不仅可以在 Windows 平台上使用,还可以在 UNIX、Linux 和 macOS 等平台上使用。相对其他数据库而言,MySQL 的使用更加方便、快捷,而且 MySQL 是免费的,运营成本低,因此,越来越多的公司使用 MySQL。

如图 1-6 所示是数据库流行度排行榜 DB-Engines Ranking。从图中可以看到,在关系数据库中,Oracle、MySQL、Microsoft SQL Server 排前三位,流行度远远超过其他数据库。在非关系数据库中,比较流行的有 MongoDB、Elasticsearch、Redis 等。

391 systems in ranking, April 2022									
Rank			DBMS	Database Model	Score				
Apr 2022	Mar 2022	Apr 2021			Apr 2022	Mar 2022	Apr 2021		
1.	1.	1.	Oracle 📈	Relational, Multi-model 📈	1254.82	+3.50	-20.10		
2.	2.	2.	MySQL 📈	Relational, Multi-model 📈	1204.16	+5.93	-16.53		
3.	3.	3.	Microsoft SQL Server 📈	Relational, Multi-model 📈	938.46	+4.67	-69.51		
4.	4.	4.	PostgreSQL 📈 🗯	Relational, Multi-model 📈	614.46	-2.47	+60.94		
5.	5.	5.	MongoDB 📈	Document, Multi-model 📈	483.38	-2.28	+13.41		
6.	6.	↑ 7.	Redis 📈	Key-value, Multi-model 📈	177.61	+0.85	+21.72		
7.	↑ 8.	↑ 8.	Elasticsearch 📈	Search engine, Multi-model 📈	160.83	+0.89	+8.66		
8.	↓ 7.	↓ 6.	IBM Db2	Relational, Multi-model 📈	160.46	-1.69	+2.68		
9.	9.	↑ 10.	Microsoft Access	Relational	142.78	+7.36	+26.06		
10.	10.	↓ 9.	SQLite 📈	Relational	132.80	+0.62	+7.74		

图 1-6 数据库流行度排名

1.2 MySQL 的安装与配置

大型商业数据库虽然功能强大,但价格也非常昂贵,因此,许多中小型企业将目光转向开源数据库。开源数据库具有速度快、易用性好、支持 SQL、对网络的支持性好、可移植性好、费用低等特点,完全能够满足中小企业的需求。尤其在知识产权越来越受重视的今天,开源数据库更加成为企业应用数据库的首选。本书将以 MySQL 关系数据库系统为平台讲述数据库的设计与实现。

下面首先回顾 MySQL 发展史。

作为最受欢迎的开源关系数据库管理系统,MySQL 最早来自 MySQL AB 公司的 ISAMS 与 mSQL 项目(主要用于数据仓库场景),1996 年 MySQL 1.0 诞生,当时只支持 SQL 特性,还没有事务支持。MySQL 3.11.1 是第一个对外提供服务的版本,MySQL 主从复制功能也是从这个时候加入的。

2000 年前后,设计人员尝试将 InnoDB 引擎加入 MySQL 中。

2003 年 12 月,MySQL 5.0 提供了视图、存储过程等功能。