

阶段 1 正确设计数据库并维护

能正确设计数据库并确保数据库及数据记录的正确性,是实施数据库应用项目的前提和基础。

本阶段通过若干任务,介绍数据库的设计、数据库完整性的控制、记录的正确维护等相关的知识和技能。

阶段任务

- 任务 1 数据库技术概述
- 任务 2 MySQL 简介
- 任务 3 学生成绩数据库的设计
- 任务 4 学生成绩数据库的创建
- 任务 5 学生成绩数据库中表的创建
- 任务 6 学生成绩数据库的完整性控制
- 任务 7 学生成绩数据库中记录的维护

阶段知识目标

- (1) 了解数据库相关知识、技术的发展。
- (2) 了解 MySQL 数据库管理系统的内涵和特点。
- (3) 理解关系模型,包括关系的基本概念——关系、关系模式、元组、属性、键等;关系的基本性质——元组的唯一性、顺序无关性,属性的唯一性、同属性的同质性等;关系的完整性——实体完整性、参照完整性等。
- (4) 理解数据库的设计过程,包括需求分析、概念设计(E-R 图)、逻辑设计、物理设计等。
- (5) 理解数据库的字符集和存储引擎的作用。
- (6) 理解 MySQL 的常用数据类型、常用系统函数。
- (7) 理解列级约束,包括主键约束和外键约束、非空约束、唯一性约束、默认值约束等的作用。
- (8) 理解数据库实体完整性、参照完整性的概念和作用。
- (9) 理解 INSERT、DELETE、UPDATE 语句的适用情境和语法要素。
- (10) 理解各类运算符的作用和条件表达式的设计。

阶段技能目标

- (1) 能安装、配置、启动、登录 MySQL 数据库服务器。
- (2) 掌握数据库设计过程中各主要阶段的实施技能,得到各阶段的目标文档,包括简

单的需求分析文档,概念设计和 E-R 图,逻辑设计中规范的(符合范式)的关系,物理设计和库、表、字段的具体命名和类型、键的确定。

(3) 能用图形界面和 SQL 语句创建、修改和查看数据库内容,确定其字符集和存储引擎。

(4) 能用图形界面和 SQL 语句创建、修改和查看数据表,确定字段名及其类型,并设置其列级约束,包括主键约束和外键约束、唯一性约束、非空约束和默认值约束。

(5) 能正确地设置主键,确保实体完整性;正确地设置外键,确保参照完整性。

(6) 能正确地应用运算符、设计合理的条件表达式。

(7) 掌握记录添加的技术模型,分析其应用情境;设计 INSERT 语句,并调试和观察效果。

(8) 掌握记录删除的技术模型,分析其应用情境;设计 DELETE 语句,并调试和观察效果。

(9) 掌握记录修改的技术模型,分析其应用情境;设计 UPDATE 语句,并调试和观察效果。

任务 1 数据库技术概述



小雅和小佳都是 2024 年入学的软件专业的学生。在第二学期时,学校开设了“数据库技术”这门课。开学时,学生们上的是网课,很多环节都需要自学,然后咨询任课老师杨老师。在学习前,她们想先知道的是:这门课学些什么?在今后的工作中有什么作用?杨老师觉得,这些问题提得很好,非常有必要先让学生们对该课程提前认识一下。

杨老师说:“在当代社会,大家都会用手机 App 进行购物、订票、查询分数等,在这些 App 的后台,所有的购物信息、订阅信息、分数信息等,其实都被存储在系统所带的数据库里。”

购物时,各类信息通过操作界面被录入进数据库,之后就可以进行各类查询了,如查询本月消费、折扣等。考试分数,也是由考试院录入数据库后,学生再进行查询。订票也是这样。事实上,如今社会几乎所有事务都是有软件操作界面的,所有信息都可以被录入、保存在数据库中,然后进行查询、统计、大数据分析和推送。这些推送,比你更懂你。

数据库课程主要学习软件系统中数据库的设计、完整性控制、记录录入、查询、统计,以及安全管理和权限控制等。在后续的各类程序开发课程以及软件公司的项目中,所有的软件系统都带有数据库。程序员需要在操作界面收集用户信息,写添加语句将这些信息送入数据库,并根据用户的查询需求,写查询语句查得结果集并反馈给客户,等等。所以,数据库课程所学的知识技能,将伴随软件及相关专业的学生度过今后的课程学习过程和工作过程。



相关知识

在短短的几十年中,很多人特别是中年以上人士,经历了从农业社会转向工业社会,转向以台式计算机为工具的信息社会,再转向以移动终端为典型代表的万物互联社会,转变是如此迅速。农业社会的要素是土地和人,那时的人必须掌握基本的农业生产技能。工业社会的要素是技术、设备、场地和人,人必须掌握必要的操作机器的技能。在信息社会人们开始用信息技术优化各类生产和活动,因此必须掌握相关的信息技术。在万物互联社会中,大量的互联设备和现代传媒产生了海量的信息,如何不迷失在汪洋大海般的信息中,反而是当下的人,特别是年轻人必须拥有的意识。

数据库技术可以帮助人们收集数据、保存数据、筛选数据、查询数据、统计和汇总,挖掘出隐秘然而相关联的信息,从而辅助决策等。因此,数据库技术可以在信息的汪洋大海里起到挖掘有用信息、从而引航的作用,这在当今社会是一种不可或缺的技能。

1.1 数据库相关概念

1. 信息

计算机把人类带入信息社会。信息(information)就是对各种事物的属性、状态及相互间的关联的描述。这种描述可以采用任何人可以理解的形式,如文字、数字、各类图像、各类视频等,都属于信息。

2. 数据

数据(data)是用来记录信息的、具有一定格式的符号的集合。这种符号必须是计算机系统能识别、存储和处理的,如数字、字符、符合格式规范的图像等。

3. 信息技术

将现实世界中需要进行管理的系统(如工资管理系统、订票系统、库存管理系统、销售系统等)中涉及的信息世界抽象为数据世界,通过对数据的管理,实现对现实世界的管理,这种技术称为信息技术(information technic, IT)。信息技术在当今世界是不可或缺的。如学习、购物、出行订票、各类交易等,都必须通过信息技术利用软件系统来实现运营和管理。

4. 数据库

为了实现软件系统的功能,需要将与之相关的各类数据(包括所有的原始数据、运行数据等)组织在专门的集合中,实现数据的保存,并通过对数据的查询、统计、挖掘和分析,实现信息化管理,这个专门的集合就是数据库(database, DB)。

数据库是为了实现软件的功能,将所需要的所有数据、按一定格式组织而成的集合。



关系的概念与性质

5. 数据库管理系统

数据库是由数据库管理系统来实施管理的。数据库管理系统(database management system,DBMS)是计算机系统软件的一种,它的职能是有效地组织和存储数据、获取和管理数据,接收和完成用户提出的访问数据的各种请求。MySQL、Access、SQL Server、Oracle 等是目前常用的数据库管理系统。其通用的功能如下。

1) 数据定义

DBMS 提供了数据定义语言(data definition language,DDL),用户通过它可以方便地对数据库中的数据对象进行定义、生成、修改、删除等操作。

2) 数据操纵

DBMS 还提供了数据操纵语言(data manipulation language,DML),用户可以使用 DML 实现对数据库记录的基本操作,如查询、增加、删除和修改等。

3) 数据库的创建和维护

数据库的创建和维护包括数据库初始数据的输入、完整性控制、存储、转换功能,数据库的备份、恢复功能,数据库的组织功能和性能监视、分析功能等。

4) 数据库的运行管理

数据库在创建、运用和维护时,由 DBMS 对使用它的所有用户及其权限进行统一管理和控制,以保证数据库的安全性和多用户对数据的并发使用等操作。

6. 数据库系统

数据库系统(database system,DBS)是指在计算机系统中引入数据库后的系统,由计算机硬件、操作系统、数据库管理系统(及其开发工具)、应用程序、数据库用户构成。数据库系统的示意图如图 1.1 所示。

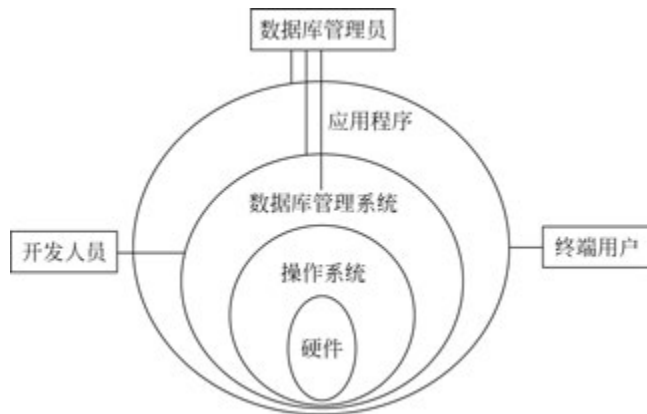


图 1.1 数据库系统的示意图

可见,数据、数据库、数据库管理系统、数据库系统,这一系列的概念,其范围逐步扩大,从单一的符号、到数据集合、到管理数据集合的系统软件,最后到包含软硬件、各类使用人员的系统。在学习数据库之前,了解这一系列概念有助于大家对数据库形成全面和

系统的认识。

1.2 数据库的使用人员

在学习数据库之前,如果对自己未来在数据库系统中担任的角色有清晰的认识,也能在学习时更有针对性。

1. 开发人员

数据库系统处于开发阶段时,其用户是开发人员。开发人员有以下两类。

(1) 系统分析员:实现从信息世界向数据世界的抽象,通过需求分析、概念设计、逻辑设计和物理设计,实现数据库的设计,这是数据库应用的前提和基础。

(2) 程序员:在开发程序时,应用数据库,编写 SQL 语句和其他代码,实现数据的录入和查询,以及其他预设的功能。

这两类用户对数据库的结构和数据都应该有全面、深刻的认识和理解,当然要想成为薪资很高的系统分析员,必须先耐心、勤恳地做好程序员工作。

2. 数据库管理员和终端用户

数据库系统交付使用后,也有两类用户。

(1) 数据库管理员(database administrator,DBA):管理数据库、管理使用数据库的用户及其权限,进行备份恢复、日志管理等,确保数据库的安全。这类用户对数据库的结构、数据、特别是各类安全相关的问题,必须有不容出错的保障能力。

(2) 使用软件的终端用户:如使用网络购物软件,其登录记录、购物车、购物记录、维权记录等,都会被及时存储在软件的数据库中。当用户需要修改、查询、统计时,就利用这些记录为其提供所需的服务。因此,终端用户其实是不断在与数据库打交道,当然这种接触是通过软件间接实现的,不可能允许其直接与数据库接触。

在软件系统交付使用后,开发人员将完成使命并退出,终端用户只能通过软件间接地在数据库中增加记录、查询等,只有 DBA 能直接操作数据库。在现代社会,数据库是各类机构最重要和关键的资产之一。

所以,大家需要考虑清楚:自己在学习时,属于什么角色?以后开发软件系统时,属于什么角色?在日常使用各类系统时,又属于什么角色?

1.3 SQL 概述

SQL(structured query language,结构化查询语言)是 ISO 组织认可的国际通用的、关系型数据库查询语言,其中的语句(或称命令)可以用于定义和操作数据库等。

1. SQL 的特点

SQL 主要有以下特点。

(1) 一体化。SQL 集数据定义、数据操作和数据控制功能于一体,可以满足数据库相关的全部工作要求。

(2) 使用方式多样。SQL 语句既可以直接用命令方式交互式执行;也可以嵌入 C、C++、Java、Python 等几乎所有的主流编程语言中使用。

(3) 非过程化。SQL 语句只需要告诉计算机“做什么”,不需要告知它“怎么做”;只需要提供操作需求,不需要提供操作步骤。

(4) 语言简洁,语法简单,好学好用。SQL 的核心功能只需 6 个动词:创建(CREATE)、修改(ALTER)、查询(SELECT)、增加(INSERT)、删除(DELETE)和更新(UPDATE),语法接近英语口语。

2. SQL 的主要语句

SQL 语句主要分为以下几类。

1) 数据定义语句

数据定义语句用于定义和修改数据库、表等数据库对象,常用语句有 CREATE、ALTER、DROP 等,如表 1.1 所示。

2) 数据操纵语句

数据操纵语句用于在数据库的表中增加、更新、删除、查询、更新记录等。常用的语句有 SELECT、INSERT、DELETE、UPDATE 等,如表 1.2 所示。

表 1.1 主要的数据库定义语句

语 句	功 能 描 述
CREATE	创建库、表、约束等各类对象
ALTER	修改库、表、约束等各类对象
DROP	删除库、表、约束等各类对象

表 1.2 主要的数据库操纵语句

语 句	功 能 描 述
SELECT	在表、视图中查询、统计
INSERT	在表、视图中增加记录
DELETE	在表、视图中删除记录
UPDATE	在表、视图中更新数据

3) 数据控制语句

数据控制语句主要用于安全,为各类用户指定其操作权限。常用的语句有 GRANT、REVOKE、DENY 等,如表 1.3 所示。

表 1.3 主要的数据库控制语句

语 句	功 能 描 述
GRANT	授予某用户对某对象的操作权限
REVOKE	撤销某用户对某对象的操作权限
DENY	拒绝某用户对某对象的操作权限

学习数据库技术时,面对具体的问题,首先分析问题、定性,然后决定使用哪些 SQL 命令来解决问题,是必须掌握的核心技能。

1.4 数据管理技术的发展

在了解了与数据库有关的重要概念如数据库(DB)、数据库管理系统(DBMS)、数据库系统(DBS)、各类使用人员和 SQL 之后,有必要了解一下计算机技术出现后,数据管理技术的发展历程,从而对数据库的作用有更清晰的认识。

计算机技术出现后,数据管理就成为一个无法避免和必须解决的重要环节。从那个时候到目前为止,主要的数据管理技术有人工管理、文件系统、数据库系统、分布式数据库系统、数据仓库、数据挖掘、云计算和大数据等。

1. 人工管理

20 世纪 50 年代前期,计算机主要用于科学计算,数据管理是通过手动方式进行的。数据由计算机或处理它的程序自行携带,数据和应用程序一一对应,处理速度是大幅提高了,但数据与程序交织的形式,带来了数据的不规范、重复、随程序改动而更改、无法独立处理等缺陷。数据是不独立的,只能嵌在每个具体的程序里,如图 1.2 所示。

因此,这一时期数据管理的特点是:数据的独立性差、数据不能被长期保存、数据的冗余度大等。

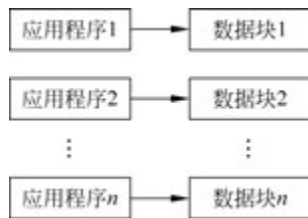


图 1.2 人工管理

2. 文件系统

20 世纪 50 年代后期至 60 年代中后期,磁盘成为计算机的主要外存储器。在软件方面,出现了高级语言和操作系统。因此,数据得以以文件的形式进行组织,并能长期保留在外存储器上。用户可以对数据文件进行查询、修改、增加和删除等操作,数据的独立性在一定程度上提高了。但数据文件没有统一的格式和操作规范,还是紧密地捆绑在具体的程序上。

此时,程序与数据有了一定的独立性,程序和数据分开存储,然而依旧存在数据的冗余度较大及数据的不一致性等缺点,如图 1.3 所示。

3. 数据库系统

20 世纪 60 年代后期,计算机的硬件和软件都有了进一步的发展,信息量的急剧膨胀带来了数据量的急剧增长。为了解决日益增长的数据量带来的数据管理上的沉重负担,数据库技术逐渐发展和成熟起来。

数据库技术使数据有了统一的格式,能对各种数据进行规范、统一、独立的管理,以实现数据的收集、共享,保证数据的完整和安全,提高了数据管理效率。

到了此阶段,数据的管理能够消除重复和冗余,能够比较方便地保持一致性。无论是使用效率还是独立性、正确性、稳定性,都得到了较好的保障,如图 1.4 所示。

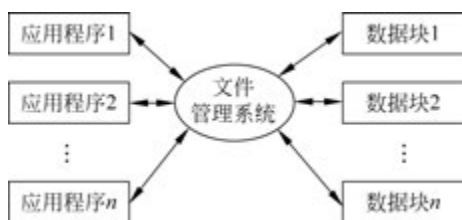


图 1.3 文件系统

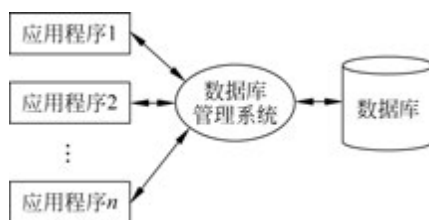


图 1.4 数据库系统

4. 分布式数据库系统

随着地理上分散的用户对数据共享的要求日益增强,以及计算机网络技术的发展,在传统的集中式数据库系统基础上发展出了分布式数据库系统。

分布式数据库系统(distributed database system, DDBS)并不是简单地把集中式数据库安装在不同场地,用网络连接起来实现,这仍然是分散的集中式数据库系统。DDBS 具有自己的性质和特征。集中式数据库系统中的许多概念,如数据独立性,数据共享和减少冗余,并发控制和事务恢复等,在分布式数据库系统中有了不同的、更加丰富的内涵。

分布式数据库系统包括分布式数据库管理系统(DDBMS)和分布式数据库(DDB)。在分布式数据库系统中,一个应用程序可以对数据库进行“透明”的操作。此时,数据库中的数据分别在不同的局部数据库中存储、由不同的 DBMS 进行管理、在不同的机器上运行、由不同的操作系统支持、被不同的通信网络连接在一起。而应用程序的感觉却仍然像是在操作一个传统的数据库一样,这就是计算机领域内所谓的“透明”。分布式数据库系统如图 1.5 所示。

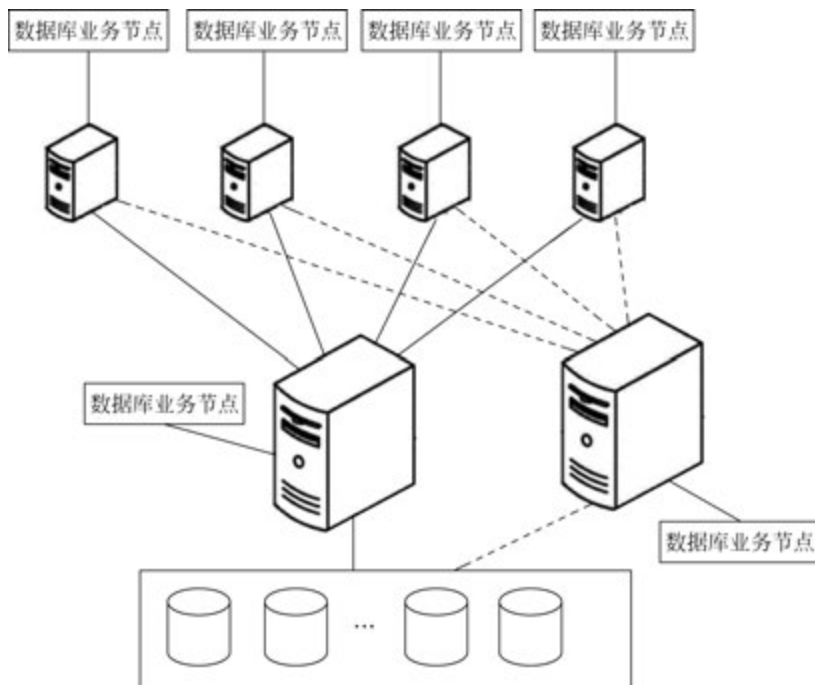


图 1.5 分布式数据库系统

5. 数据仓库

数据库技术经过几十年的发展和广泛应用,在社会各行各业积累了大量数据,使数据不断膨胀。从数据海洋中提取、检索出有用的信息——那种能够支持决策的信息,为企业的管理决策提供精准的数据分析和支持,成为数据库技术的发展趋势。因此,数据仓库技术,包括数据挖掘技术已成为数据库技术发展的前沿。

比尔·恩门(Bill Inmon)在《建立数据仓库》一书中对数据仓库的定义是:数据仓库是面向主题的、集成的、随时间变化的、非易失性数据的集合,用于支持管理层的决策。

数据仓库的特点有:面向主题性,数据集成性,数据的多样性、多源性,数据库的非易失性,数据的集合性和支持决策,具有现代化收集、搜索、分析和统计手段的、很大的数据仓库,如图 1.6 所示。

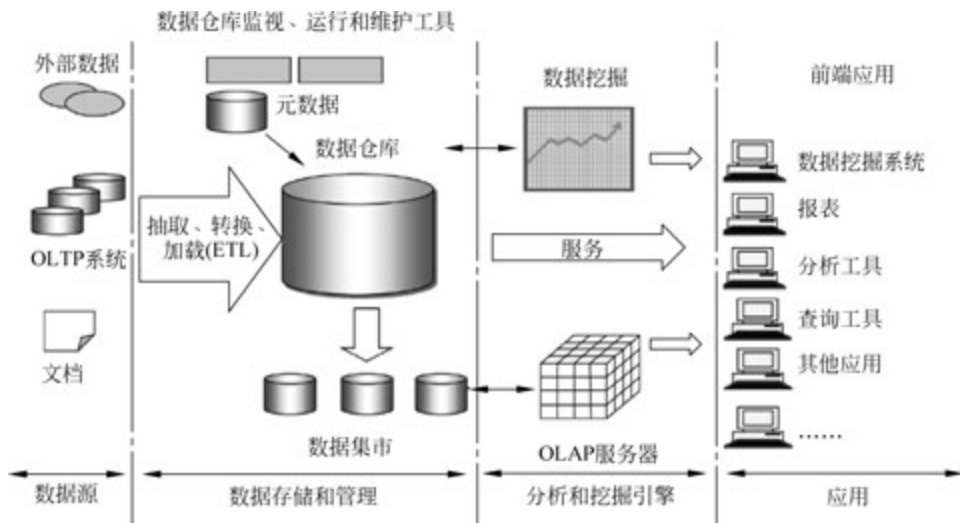


图 1.6 数据仓库

6. 数据挖掘

数据挖掘(data mining)又称为数据库中的知识发现(knowledge-discovery in databases, KDD),是指从大量数据中通过算法,搜索隐藏于其中的、有用的关联信息的过程。例如,著名的啤酒和尿布关联现象的发掘(指买尿布的男人们同时,往往会买啤酒,从而指导超市的商品上架策略)。一般必须通过各类高级算法,如统计、在线分析处理、机器学习、专家系统和模式识别等来实现这种智能挖掘。

在数据挖掘中进行数据分析常用的算法主要有分类、聚类、回归分析、人工智能神经网络等,它们分别根据不同的理论对数据进行归类和挖掘。相信大家都有体会,当你在网站上浏览或购买过某些商品后,它会自动向你推送同类或类似的产品,这就是数据挖掘的

一种实际应用。

数据挖掘的过程如图 1.7 所示。

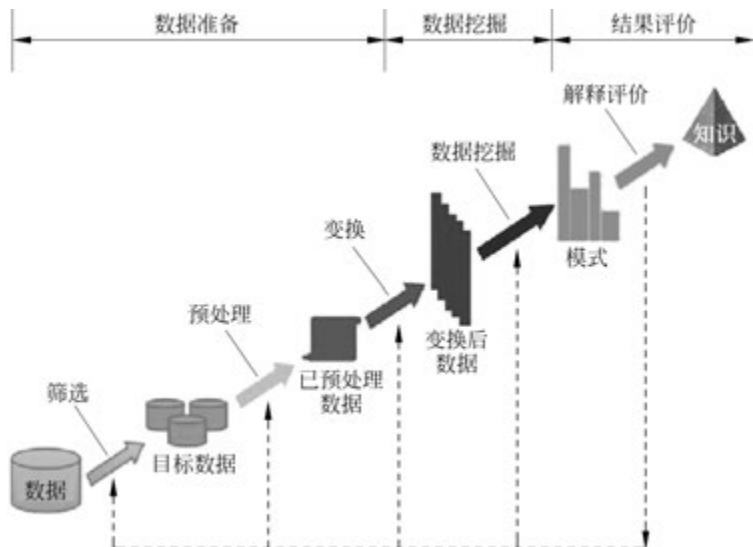


图 1.7 数据挖掘的过程

7. 云计算和大数据

云计算(cloud computing)通过使计算分布在大量的分布式计算机上,或远程服务器中而非本地计算机上,使得企业数据中心的运行与互联网更相似。云计算通常是指通过互联网来提供动态、易扩展并且是虚拟化的计算资源。这是一种基于互联网的使用和交付模式,可以根据需求访问网络上的计算机和存储系统。这样计算不再局限于在本地计算机上进行,而是由网络上的资源共同负担来实现。

大数据(bigdata)是指基于云计算的处理模式而获得的海量的、高增长率和多样化的信息资产。IBM 提出的大数据的 5V 特点是 Volume(大量)、Velocity(高速)、Variety(多样)、Value(低价值密度)、Veracity(真实性)。

对于云计算与大数据的关系,在定义层面,云计算注重资源分配,是硬件资源的虚拟化;而大数据是海量数据的高效处理。在技术层面,大数据与云计算的关系就像一枚硬币的正反面。大数据必然无法用单台计算机进行处理,必须采用分布式架构。大数据的内涵在于对海量数据进行分布式数据挖掘,所以必须依托云计算的分布式处理、云存储和虚拟化技术,如图 1.8 所示。或者说,就作用而言,云计算像硬件,大数据像软件。

分布式数据库、数据仓库、数据挖掘、云计算、大数据,都是目前炙手可热的前沿技术。这一方面说明人们对数据海洋中有效信息进行挖掘的迫切需求,另一方面也凸显了数据库技术的基石地位。