



项目 1

认识数据库



项目目标

- 能够描述数据、数据库、数据库管理系统等相关概念；
- 能够根据不同的应用需求正确选择数据库和数据库管理系统。



项目描述

数据库将计算机科学和易于人类理解的数据管理方式完美地衔接在一起，从 20 世纪 50 年代开始，就逐渐影响并改变了世界。改革开放初期，数据库理论正式传入中国。历经 40 多年的技术迭代，国产数据库技术蓬勃发展，见证了中国自主技术的巨大变革。

在本项目中，我们将深入探讨数据、数据库和数据库管理系统的基本概念，以便更好地理解这些概念在计算机科学和现代信息管理中的关键作用，并能够根据不同的应用需求选择合适的数据库和数据库管理系统，从而解决现实世界中的数据管理问题。

任务 1.1 认识数据与数据库



微课：认识数据库

1. 数据

数据（data）是用来描述事物的符号记录，可以有数字、文字、图形、图像、声音、视频、动画等多种表现形式。当这些数据经过数字化处理后，可以存储到计算机中供进一步处理和分析。

数据通常分为结构化数据（structured data）和非结构化数据（unstructured data）两种类型。结构化数据是指具有预定义数据模型和组织方式的数据，通常存储在关系数据库中，以表格形式展现。每个表格都有许多列，用于描述数据的不同属性，而行则代表数据记录。常见的 Excel 数据库及 SQL 数据库中的数据就是典型的结构化数据。结构化数据的



优点在于它们易于进行查询、排序和分析。例如，在表 1.1 中，学号、姓名等信息都可以被视为结构化数据，因为它们可以被组织在表格中，并且每一种信息都有其预定义的数据类型和格式。

表 1.1 结构化二维表举例

学号	姓名	性别	班级
202401	小明	男	计算机 01
202402	小红	女	计算机 02
202403	小兰	女	计算机 02

相反，非结构化数据是指没有预定义数据模型或固定格式的数据，如文本、图像、音频和视频等。这些数据类型不易在传统的关系数据库中存储和分析，但是在日常生活和工作中，非结构化数据占据主导地位。

使用普通的数据分析方法，从这些非结构化数据中挖掘有用的信息可能非常困难，通常需要使用更为复杂的技术。例如，使用文本挖掘技术对文本信息进行分析，利用图像识别技术对图片进行分析，应用自然语言处理技术理解和分析语音和文本数据等。

2. 数据库

数据库（database，DB）是一个系统化的、长期存储在计算机内的、有组织的、可共享的数据集合。它不仅是一个数据的仓库，更是一个能够存储、处理和管理数据的平台。数据库中的数据按照一定的数据模型组织、描述和存储，这些数据模型可以是关系模型、层次模型、网络模型、关系模型、列存储模型、文档模型、键值对模型、图模型、时序模型或者向量模型等。

任务 1.2 认识数据库管理系统和数据库系统

1. 数据库管理系统

数据库管理系统（database management system，DBMS）是一种特殊的系统软件，在用户和数据库之间起到中介的作用，用于科学地组织、存储、管理和检索数据。它提供了一个系统化的方法来创建、更新、管理和检索数据库中的数据，如图 1.1 所示。

DBMS 的主要职责是提供一种方便、有效的方式来存储和检索大量的数据，同时保证数据的安全性、一致性和完整性。它还提供了一种抽象的视图来隐藏物理存储的细节，使得用户可以专注于数据的逻辑结构而不是物理存储。

DBMS 还有助于数据的并发控制，即在多个用户同时访问数据时，确保数据的一致性和完整性。此外，DBMS 还提供了数据的备份和恢复功能，以防止数据丢失或损坏。

2. 数据库系统

数据库系统（database system，DBS）是一个综合性系统，包括数据库、数据库管理系统和与其交互的应用程序及用户，如图 1.2 所示。它是在计算机环境中引入数据库后形成的整体结构，用于有效地管理和处理数据。

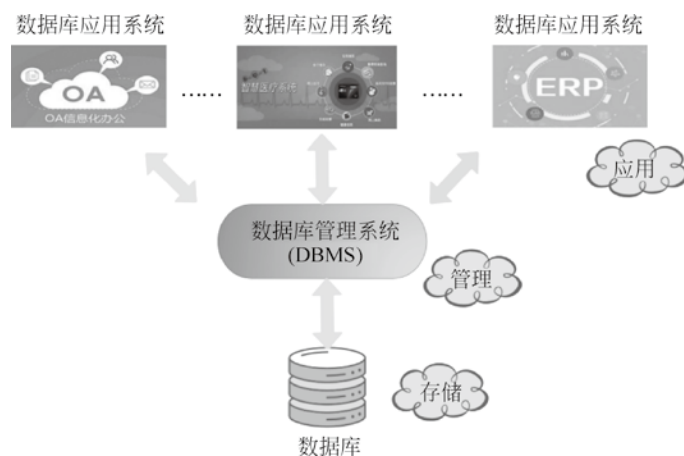


图 1.1 数据库管理系统作用示意图

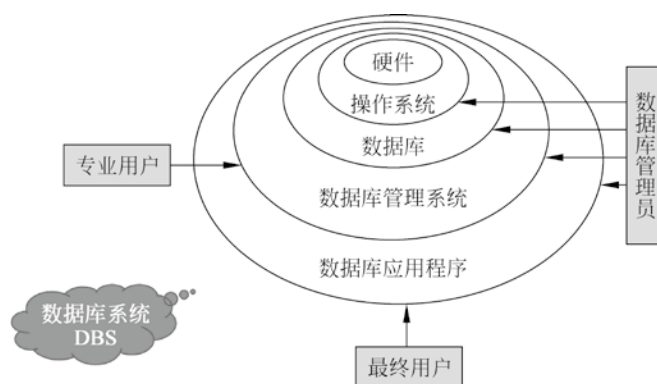


图 1.2 数据库系统构成示意图

任务 1.3 认识常见的数据库管理系统

随着数据库技术的演进，目前主要有两类数据库管理系统，即关系型数据库管理系统和非关系型数据库管理系统（见图 1.3）。关系型数据库管理系统主要用于处理结构化数据，而非关系型数据库管理系统则更适合处理非结构化数据。

1. 关系型数据库管理系统

关系型数据库是一种采用关系模型来组织数据的数据库。数据以二维表格的形式存储，每个表格由行和列组成。数据库则由这些表格及其间的关系构成。常见的关系型数据库管理系统包括 MySQL、SQL Server 和 DB2 等。

1) MySQL

MySQL 由瑞典 MySQL AB 公司开发，属于 Oracle（甲骨文）公司旗下产品。其体积小、速度快、总体拥有成本低，尤其是开放源码这一特点，适合于一般中小型系统，在 Web 应用开发领域，MySQL 占有举足轻重的地位。

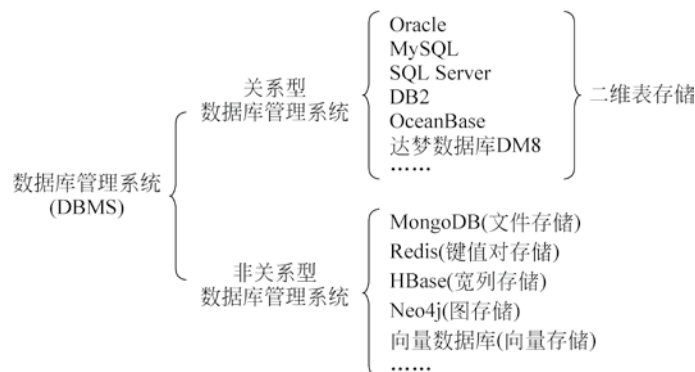


图 1.3 数据库管理系统分类

目前使用 MySQL 的著名公司和产品有 Facebook、Github、YouTube、Twitter、阿里巴巴、美团外卖、腾讯、微信等。

2) SQL Server

SQL Server 由 Microsoft 公司推出，具有使用方便、可伸缩性好、与 Windows 操作系统和微软相关软件集成程度高等优点。早期版本的 SQL Server 只能运行在 Windows 平台上，SQL Server 2017 以后的版本也能支持 UNIX 平台。

3) DB2

DB2 由 IBM 公司开发，运行环境主要是 UNIX 和 Windows Server。DB2 提供了高层次的数据利用性、完整性、安全性、可恢复性，采用了数据分级技术，主要应用于大型分布式应用系统，具有较好的可伸缩性。

2. 非关系型数据库管理系统

关系型数据库虽然强大，但在大数据时代，面对海量和复杂的数据模型，其处理能力有限。非关系型（NoSQL）数据库因其易扩展性、海量数据处理能力、高性能和灵活的数据模型，已在数据库领域得到了广泛应用。

NoSQL 数据库主要包括以下特点。

（1）易扩展：由于数据间无须建立关系，NoSQL 数据库易于扩展。

（2）大数据量，高性能：NoSQL 数据库具有高读写性能，尤其在处理海量数据时，得益于其无关系性和简单的数据库结构。

（3）灵活的数据模型：NoSQL 数据库不需要预定义字段，可随时存储自定义数据格式，这在关系型数据库中是困难的。

（4）高可用：NoSQL 数据库可在不影响性能的情况下实现高可用架构。

常见的非关系型数据库管理系统包括 MongoDB、Redis、HBase、Neo4j 和向量数据库等。

1) MongoDB

MongoDB 是一个由 C++ 语言编写的、基于分布式文件存储的开源数据库管理系统，为 Web 应用提供可扩展的高性能数据存储解决方案。MongoDB 使用 BSON（Binary JSON）来存储数据、进行网络交换，BSON 是一种类 JSON 的二进制存储格式，属于 schema-less 的存储形式，支持内嵌的文档对象和数组对象，比 JSON 数据类型更丰富。



MongoDB 是一个介于关系型数据库管理系统和非关系型数据库管理系统之间的产品，是功能最丰富，最像关系型数据库管理系统的非关系型数据库管理系统。它的特点是高性能、易部署、易使用、存储数据非常方便。

2) Redis

Redis 是一个开源的、高性能的键值数据库管理系统，使用 ANSI C 语言编写。作为一个基于内存运行且支持数据持久化功能（如 RDB 和 AOF）的系统，Redis 不仅能进行高效读写操作，还能保证数据安全。它具备主从复制机制，支持多种数据类型，如 string、list、set、zset、hash 等，还有众多由社区贡献和维护，针对 Java、C/C++、C#、PHP、JavaScript、Python 等编程语言的客户端 API 库。

3) HBase

HBase（Hadoop database）是一个构建在 Hadoop 文件系统（hadoop distributed file system, HDFS）之上的分布式、可伸缩的宽列存储数据库，它为 HDFS 提供了高并发随机写入和实时查询功能。其设计目标是面向大规模数据存储，支持快速随机读写操作，适用于海量半结构化或非结构化数据场景。

4) Neo4j

Neo4j 是一个高性能的 NoSQL 图形数据库管理系统。它将结构化数据存储在网络上而不是传统的表格中。图数据模型是一种用于表示实体（节点）及其之间的关系（边）的数据结构，在处理复杂关系数据时表现出色，尤其适用于社交网络、推荐系统和知识图谱等应用场景。

Neo4j 是一个嵌入式的、基于磁盘的、具备完全的事务特性的 Java 持久化引擎。Neo4j 因其嵌入式、高性能、轻量级等优势，越来越受到关注。

5) 向量数据库

向量数据库是一类专门用于高效存储和检索高维向量数据的数据库管理系统，这种类型的数据库在 AI 领域中具有重要的应用。例如，向量数据库可以用于存储和快速检索图像、文本等数据的特征向量，以支持相似性搜索和推荐系统。

常见的向量数据库有以下三种。

(1) Faiss：由 Facebook AI Research 开发，支持高效的相似性搜索和聚类。

(2) Annoy：由 Spotify 开发，适用于内存有限的情况下进行近似最近邻搜索。

(3) Milvus：Milvus 是一款由 Zilliz 开发的开源向量数据库，专为高效存储和搜索大规模高维向量数据而设计。它适用于处理文本、图像、视频和音频等非结构化数据的特征空间位置表示，并通过嵌入技术将这些数据转换为向量，以便计算机理解 and 处理。

3. 国产数据库管理系统

目前，流行的数据库管理系统主要由国外公司开发。然而，将重要信息和核心数据存储在外国公司开发的数据库中可能存在安全隐患，基础软件作为国家信息安全的重要组成部分，“国产化”具有重要意义。

经过多年的发展，国产数据库管理系统已经取得了显著的进步，不仅有传统的“三剑客”：达梦、人大金仓和南大通用，还有互联网巨头、网络通信巨头和创业公司的新进入者，形成了百花齐放、群雄逐鹿的局面。国产数据库管理系统已经取得了明显的突破：快



速发展、进入国际视野、获得权威机构认可、产品达到金融级性能和可用性，逐渐被大中型企业采用。

目前，国产数据库管理系统正处于发展的黄金时期。创新的商业模式和丰富的应用场景正在推动数据库进入“升级换代”的新时代。同时，业务和成本的挑战也在加快数据库的“升级换代”。

典型的国产数据库管理系统包括以下几种。

1) TiDB

TiDB 是一款由 PingCAP 公司开发的分布式数据库管理系统，旨在提供水平扩展能力的同时，兼具在线事务处理（on-line transaction processing, OLTP）和在线分析处理（on-line analytical processing, OLAP）能力，实现 HTAP（Hybrid Transactional and Analytical Processing）混合事务分析处理。其核心架构由计算引擎层、分布式协调层、以及存储引擎层三大模块组成，每个层次包含不同的组件，协同工作以满足高性能、高并发和复杂查询的需求。

2) OceanBase

OceanBase 是阿里巴巴和蚂蚁集团自研的原生分布式关系数据库管理系统，能在普通硬件上实现金融级高可用，具备卓越的水平扩展能力。OceanBase 是全球首家通过 TPC-C 标准测试的分布式数据库管理系统，单集群规模超过 1500 个节点，具有云原生、强一致性，以及高度兼容 Oracle、MySQL 特性，广泛应用于金融行业。

3) 达梦（DM8）

DM8 是达梦公司开发的新一代自主研制数据库管理系统，集成了分布式技术、弹性计算及云计算等先进技术的优势。该系统对灵活性、易用性、可靠性和高安全性等方面进行了显著改进，能够支撑超大规模并发事务处理，并适用于同时进行事务处理和分析型业务操作。此外，DM8 能动态分配计算资源，实现精细化资源利用，并降低成本投入。

◆ 项目任务单 ◆

在本项目中，我们深入探讨了数据库的基本知识，包括数据、数据库、数据库管理系统、数据库系统等核心概念，以及数据库技术的发展历史。我们还大致了解了全球流行的数据库管理系统和国产数据库管理系统。为了检验读者的理解和掌握程度，请完成以下任务。

1. 用自己的语言解释数据、数据库、数据库管理系统和数据库系统这些基本概念，尽可能地把这些概念讲清楚。

2. 列出并简述几种全球流行的数据库管理系统，包括其特点和主要的应用场景。



3. 选择一个流行的非关系型 (NoSQL) 数据库, 如 MongoDB、Redis 或 Cassandra 等, 研究并了解其基本概念、特点和应用场景。

◆ 拓展任务 ◆

根据提供的微课视频, 自主学习并掌握“中国数据库发展简史”模块的内容。请利用墨天轮数据社区 (一个可以方便获取国产数据库最新动态的平台), 选择一款流行的国产数据库管理系统, 如华为的 GaussDB、阿里巴巴的 PolarDB、达梦的 DM 等。研究并编写一份报告, 介绍该数据库管理系统的发展历程, 包括主要的产品及其特色。请尽可能详细地描述产品的主要技术特点、应用场景, 以及它们在数据库领域的创新之处。



微课: 中国数据库发展简史



安装与配置 MySQL 开发环境



项目目标

- 能够独立完成 MySQL 8.0 的下载、安装与配置；
- 能够使用不同方法启动、停止 MySQL 服务；
- 能够使用第三方工具访问 MySQL 数据库。



项目描述

MySQL 是全球流行度排名第二、开源数据库排名第一的数据库管理系统。其 Logo 是一只名为 Sakila 的小海豚，灵感来源于创始人 David Axmark 在澳大利亚潜水时遇到的一只海豚——“瑞士军刀鱼”。海豚象征着速度、力量和精确性，这些也是 MySQL 的核心特性。无论是处理的数据量、查询速度，还是系统稳定性，MySQL 都表现出色。

在本项目中，我们将深入了解 MySQL 数据库管理系统，学习如何搭建 MySQL 开发环境，包括下载、安装、配置、连接测试，以及如何使用图形化管理工具 Navicat 访问 MySQL，为后续的学习做好准备。

任务 2.1 认识 MySQL



微课：认识 MySQL

MySQL 于 1995 年发布，并迅速崭露头角，成为开源数据库的领头羊。后期相继被 Sun Microsystems、Oracle 公司收购。由于担心 MySQL 可能被转为闭源，MySQL 的创始人创建了分支项目 MariaDB。MySQL 采用了 GPL（GNU general public license）协议，这意味着任何人都可以通过修改源代码来开发自己的 MySQL 系统，而无须支付额外费用。

MySQL 能够处理大型数据库，支持数千万条记录的数据库，并能处理含有 5000 万条记录的数据仓库。在 32 位系统中，表文件最大可支持 4GB，而在 64 位系统中，最大的表



文件可支持 8TB。由于 MySQL 使用标准的 SQL 数据语言，它可以在多个系统上运行，并支持多种编程语言，包括但不限于 C、C++、Python、Java、Perl、PHP 和 Ruby。

任务 2.2 了解 MySQL 发展史上的重大事件

MySQL 的历史可以说是互联网发展史的一个缩影。从社交、电商到金融领域，互联网业务的发展，都对数据库的需求提出了更高的挑战。这些挑战包括高并发、高性能、高可用、轻资源、易维护和易扩展等需求，无不推动了 MySQL 的发展。表 2.1 展示了 MySQL 发展历史上的重大事件。

表 2.1 MySQL 发展历史上的重大事件

年份	重 大 事 件
1995 年	MySQL 首次发布。由 Michael Widenius 和 David Axmark 创立，最初是为了提高表和查询的速度而开发的
2000 年	MySQL AB 公司推出 MySQL 的第一个开源版本（GPL 许可），这标志着 MySQL 成为开源社区的一部分
2008 年	Sun Microsystems 以约 10 亿美元的价格收购 MySQL AB。这次收购将 MySQL 带入了一个更大的技术生态系统中
2009 年	Oracle 公司收购 Sun Microsystems，从而成为 MySQL 的所有者。这次收购在开源社区中引起了广泛的关注和担忧，用户担心 MySQL 的开发可能会受到影响
2010 年	为了缓解社区的担忧，Oracle 公司公布了对 MySQL 的五年承诺，承诺将继续开发 MySQL 并保持其作为开源项目
2013 年	MySQL 5.6 发布，带来了许多性能改进，包括更好的复制特性、优化的查询执行计划及改进的分区支持
2015 年	MySQL 5.7 发布，进一步增强了性能，引入了 JSON 支持，提高了安全性和可伸缩性
2018 年	MySQL 8.0 发布，这是一个重要的版本，带来了诸如窗口函数、公共表达式递归、角色管理、锁定等待超时设置等众多新特性和改进
2020 年至今	MySQL 继续发展，引入更多的特性和改进，包括更好的云集成、数据安全性和隐私功能，以及性能优化

MySQL 从 5.7 版本直接跳跃到了 8.0 版本，是一个重要的里程碑。MySQL 8.0 版本在功能上做了显著的改进和增强，其中包括对 MySQL Optimizer 优化器的改进。这些改进不仅提升了运行速度，也为用户带来了更好的性能体验。

Oracle 公司目前拥有了排名前两位的关系型数据库 Oracle 和 MySQL，那么它们各适合什么样的用户呢？

Oracle 数据库更适合大型跨国企业，因为这些企业对费用不敏感，但对性能和安全性有更高的要求。由于 MySQL 体积小、速度快、总体拥有成本低，且能处理上千万条记录的大型数据库，再加上其开源的特点，很多互联网公司和中小型网站选择 MySQL 作为其数据库。

众多厂商选择 MySQL 的原因主要有以下几点。

- (1) 开源，使用成本低。
- (2) 性能卓越，服务稳定。
- (3) 软件体积小，使用简单，并且易于维护。
- (4) 历史悠久，社区用户非常活跃，遇到问题有大量资源可以获取帮助。
- (5) 许多互联网公司在用，经过了时间的检验。

任务 2.3 搭建 MySQL 开发环境

2.3.1 MySQL 的版本

MySQL 主要有以下四个版本。

(1) MySQL Community Server 社区版本：一个开源且免费的版本，任何人都可以自由下载和使用。但是，这个版本不提供官方的技术支持，更适用于大多数普通用户或者对 MySQL 有一定了解的开发者。

(2) MySQL Enterprise Edition 企业版本：一个收费版本，不能在线下载，但可以试用 30 天。它提供了更多的功能和更完备的技术支持，更适合于对数据库的功能和可靠性有较高要求的企业客户。

(3) MySQL Cluster 集群版：一个开源且免费的版本，主要用于架设集群服务器，可以将几个 MySQL Server 封装成一个 Server。这个版本需要在社区版或企业版的基础上使用。

(4) MySQL Cluster CGE 高级集群版：一个收费版本，提供了更高级的集群功能。

本书使用的版本是 2021 年 1 月 18 日发布的 MySQL Community Server 社区版，版本号为 8.0.28。

2.3.2 MySQL 的安装与配置

MySQL 可以在多个平台上运行，包括但不限于 Windows、Linux 和 macOS。每个平台上的安装和配置步骤可能有所不同，本书将重点介绍在 Windows 平台上的安装和配置方法。

1. MySQL 的下载

要下载 MySQL，首先需要访问 MySQL 的官方网站并找到适合操作系统和版本号的安装文件，如图 2.1 所示。

下载完成后，得到安装文件 mysql-installer-community-8.0.28.msi。

2. MySQL 的安装和配置

(1) 双击安装文件，在 Choosing a Setup Type 界面，选择 Server only 选项（这里只安装服务器部分），如图 2.2 所示。然后，单击 Next 按钮进入下一步。

(2) 进入 Installation 界面，单击 Execute 按钮开始安装过程，如图 2.3 和图 2.4 所示。



微课：MySQL 的安装与配置



微课：下载 MySQL 安装包