

第 1 章主要介绍数据库系统的基本概念、发展历史和现状。内容涵盖数据库系统的定义与构成，数据库领域的核心问题（DBMS 实现、数据库应用设计和数据库存取），数据库技术的两大创意（声明式接口和事务），数据库系统四个发展阶段（起步、关系数据库主导、NoSQL 探索、云数据库时代），以及关系数据库和主流数据库管理系统的演变与现状分析，并以阿里巴巴数据库架构演变和谷歌四篇论文为例进行说明。

请带着如下问题阅读教材第 1 章：

1. 什么是数据库、数据库管理系统 (DBMS) 和数据库系统 (DBS)? 它们之间的关系是什么?
2. 数据库系统需要解决哪些核心问题?
3. 数据库系统经历了哪四个主要发展阶段? 每个阶段的主要特点和关键技术是什么?
4. 什么是 NoSQL 和 NewSQL 数据库? 它们与传统的关系数据库 (RDBMS) 有什么区别?
5. 多模数据库 (Multi-model Database) 是什么? 它有哪些优势?
6. 什么是 HTAP (混合事务与分析处理)? 它与传统的 OLTP 和 OLAP 有什么不同?
7. 向量数据库是什么? 它与其他类型的数据库有什么区别?
8. 云原生数据库的架构特点及其对未来数据库发展的影响?
9. 谷歌公司的四篇著名论文 (GFS, MapReduce, BigTable, Spanner) 对现代数据库技术产生了哪些影响?
10. 以阿里巴巴集团的数据库架构演变为例, 探讨大型企业在数据库技术选择和架构演进过程中的关键因素。
11. 大数据时代数据库系统的发展趋势, 并结合具体的应用场景加以说明。

阅读过程中可能遇到的专有名词

- **数据库 (Database)**: 描述或模拟真实世界某些方面内在关联的数据集合。
- **数据库管理系统 (DBMS)**: 专门对数据库进行组织、存储和检索的软件系统。
- **数据库系统 (DBS)**: 由数据库和数据库管理系统组成的系统。
- **关系数据库 (Relational Database)**: 基于关系模型组织数据的数据库, 使用表格存储数据, 并使用 SQL 进行查询。
- **SQL (Structured Query Language)**: 用于管理和查询关系数据库的标准语言。
- **NoSQL (Not Only SQL)**: 非关系型数据库, 通常用于处理大规模、非结构化或半结构化数据。
- **NewSQL**: 结合关系数据库的 ACID 事务特性和 NoSQL 数据库的横向扩展能力的数据库系统。
- **声明式接口 (Declarative Interfaces)**: 用户只需描述想要的结果, 而无需指定如何获得结果的接口, 如 SQL。
- **事务 (Transaction)**: 将一系列数据库操作视为一个不可分割的整体, 要么全部成功提交, 要么全部失败回滚。
- **ACID (Atomicity, Consistency, Isolation, Durability)**: 关系数据库事务的四个关键特性, 保证数据可靠性。
- **CAP 理论 (Consistency, Availability, Partition Tolerance)**: 分布式系统中三个关键特性的权衡理论。
- **BASE (Basically Available, Soft state, Eventually consistent)**: NoSQL 数据库通常遵循的一致性模型, 与 ACID 模型相对。

- **OLTP (Online Transaction Processing)**: 联机事务处理, 主要处理日常事务操作。
- **OLAP (Online Analytical Processing)**: 联机分析处理, 主要用于数据分析和报表。
- **HTAP (Hybrid Transactional/Analytical Processing)**: 混合事务/分析处理, 同时支持 OLTP 和 OLAP 负载。
- **云原生数据库 (Cloud-Native Database)**: 为云计算环境而设计的数据库, 具有存算分离、资源池化、弹性伸缩、高可用性和容器化部署等特点。
- **多模数据库 (Multi-model Database)**: 支持多种数据模型 (如关系型、文档型、图型) 的数据库。
- **向量数据库(Vector Database)**: 专门存储、索引和管理向量数据的数据库, 通常用于 AI 应用。
- **共享一切架构 (Shared-Everything Architecture)**: 所有计算节点共享同一存储层的数据库架构。
- **共享存储架构 (Shared-Storage Architecture)**: 所有计算节点共享同一存储层, 但各自拥有独立计算资源的数据库架构。
- **SSTable (Sorted String Table)**: 一种用于在磁盘上持久化存储排序键值对的数据结构。
- **LSM 树(Log-Structured Merge Tree)**: 一种用于实现键值存储系统的磁盘数据结构。
- **元数据 (Metadata)**: 描述数据的数据, 例如数据库的模式和结构。
- **分布式数据库 (Distributed Database)**: 将数据存储多个计算机上的数据库系统。
- **分片 (Sharding)**: 将数据分割并存储在不同的数据库实例或节点上的技术。
- **数据湖 (Data Lake)**: 存储各种类型原始数据的存储库。
- **数据仓库(Data Warehouse)**: 为商业智能和数据分析而设计的中央数据存储。
- **弹性(Elasticity)**: 根据业务需求自动增加或减少计算资源的能力。
- **RDMA(Remote Direct Memory Access)**: 一种允许计算机直接访问另一台计算机内存的技术。
- **NVM(Non-Volatile Memory)**: 非易失性内存, 即使断电也能保留数据的存储器。
- **IOE**: 指 IBM 的小型机服务器, Oracle 数据库和 EMC 存储。
- **自驱动数据库(Self-Driving Database)**: 集成人工智能技术, 可以自主执行数据库管理任务的数据库系统。