

中国软件根技术发展白皮书

(数据库册)



中国软件行业协会

目 录

前言	1
一、数据库概述	3
（一）数据库相关概念	3
（二）数据库分类	5
（三）数据库架构标准	6
（四）数据库产业链	8
二、全球数据库竞争格局和发展态势	9
（一）全球数据库行业竞争格局	9
（二）主流数据库厂商产品分析	10
1. Oracle	10
2. IBM	11
3. Microsoft	12
4. Amazon	12
5. Google	13
（三）发展现状	14
1. 数据库是数字经济发展的基础技术和设施	14
2. 美国数据库厂商长期垄断全球数据库市场	14
3. 数据库技术路径分化带来市场分裂与竞合	15
4. “开源 + 云”已经成为数据库革新的关键力量	16
（四）全球数据库发展新态势	16
1. 数据库和硬件协同成为新兴竞争力量	16
2. 数据库领域的投资热情正在逐年升温	16
3. 中国数据库迎来技术积累爆发增长期	17
4. 技术收敛是国产数据库突围有效路径	18

三、国产数据库发展状况	19
（一）发展现状	19
（二）市场规模	20
（三）国产数据库主要厂商产品分析	21
1. 达梦数据库	22
2. 南大通用	22
3. 神舟通用	23
4. 人大金仓	23
5. 华为公司	24
6. 阿里云	24
7. 腾讯云	25
8. 平凯星辰	25
9. 云和恩墨	26
10. 海量数据	27
（四）国产数据库在重点行业应用分析	27
（五）国产数据库厂商积极布局生态建设	28
四、主流数据库技术发展分析	30
（一）关系型数据库	30
1. 技术特点	30
2. 主流产品	31
3. 发展趋势	31
（二）非关系型数据库	31
1. 技术特点	31
2. 主流产品	32
3. 发展趋势	33
五、国产数据库发展面临的形势、问题及机遇	36
（一）面临的主要形势	36

(二) 存在的主要问题	38
1. 产品种类繁多, 竞争能力不足	38
2. 技术基础薄弱, 专业人才匮乏	38
3. 安全性能存忧, 用户使用存虑	38
4. 产品标准各异, 缺乏行业标准	39
5. 产业资源分散, 生态建设滞后	40
(三) 发展机遇	40
1. 利好政策出台释放巨大红利	40
2. 数字化激发数据库市场潜力	41
3. 数据库国产化替代空间巨大	41
4. 行业技术实力积累整体增强	41
六、国产数据库发展基本判断	43
(一) 国产数据库技术路线收敛势在必行	43
(二) 国产数据库厂商应重视用户的需求	43
(三) 存算分离 + 多读多写成为发展新趋势	43
(四) 开源是国产数据库发展的有效路径	44
七、建立以 openGauss 为代表的国产数据库体系	45
(一) 内核创新研究, 打造坚实数据底座	45
(二) 联合产业资源, 共建资源池化架构	46
(三) 满足行业诉求, 打造关基行业标杆	47
(四) 打造开源社区, 共建技术开发生态	48
(五) 走进高校课堂, 培养专业技术人才	49
八、发展对策建议	50
(一) 引导数据库技术路线收敛提升竞争能力	50
(二) 强化存算分离 + 共享存储标准体系建设	50
(三) 出台专项政策支持首台套和规模化应用	50
(四) 强化数据库技术创新集聚全产业链资源	51
(五) 集中支持开源社区加大产业链生态建设	51

前 言

党的二十大报告明确提出，坚持创新在我国现代化建设全局中的核心地位，加快实现高水平科技自立自强，加快建设科技强国，并对完善科技创新体系、加快实施创新驱动发展战略等作出了系列重要部署。

当前，在错综复杂的国际环境和新技术快速迭代的背景下，基础软件根技术已经成为国家战略科技竞争的制高点。要从根本上摆脱对单一技术供应链的依赖，必须坚持原始技术创新，坚定不移地摸索出一条独立自主的软件根技术产业体系发展之路。数据库是业界公认的最复杂、跨技术领域最多的基础软件，随着数字化的深入推进，数据库在技术架构、数据管理、数据安全创新，以及数据科学基础研究、生态建设、“产学研用”深入合作等方面都面临着巨大的挑战。

现阶段，我国数据库行业市场前景广阔，整体呈现出稳步上升的发展势头，经过多年的技术沉淀和市场洗礼，正在经历从“引进”到“自研”，由“可用”向“好用”进阶的关键节点，这为国产数据库发展提供了良好的成长土壤。国产数据库要实现突破式发展，需要与行业发展变化需求相协同，也需要产业界形成合力持续推动原始创新，还需要构建起完善健康的产业生态体系。

我们倡导国产数据库技术收敛路线，鼓励产业上下游生态链资源联合起来共同打造具有国际竞争力的国产数据库技术、产品和解决方案，推动我国数据库产业高质量发展。



一、数据库概述

数字化正在改变世界万物，而数字世界的底层，则是庞大的不断产生、汇集、运算的数据。实现数字化的前提是数据化，而数据的可存、可用、可管、可流动、可分析，都离不开数据库。

数据库作为信息系统的核心，在计算机中扮演着承上启下的作用，向下实现调动计算、网络、存储等基础资源，向上完成支撑各种应用软件需求，是计算功能得以实现的基础设施底座。数据库的发展不仅影响着计算机技术的发展进程，也推动着经济社会和数字中国的持续进步。

（一）数据库相关概念

数据（Data）：数据库中存储的基本对象，是描述实物的符号记录。描述实物的符号可以是数字，也可以是文字、图形、图像、声音、语言等，数据有多重表现形式，它们都可以经过数字化后存入计算机。

数据库（DB）：长期存储在计算机内的、有组织的、可共享的、统一管理的大量数据的集合。数据库中的数据按一定的数据模型组织、描述、存储，其特点是冗余度小，具有较高的数据独立性、共享性和易扩展性。

专栏 1：数据模型是数据库系统的核心和基础
数据模型是对现实世界数据特征的抽象。通俗地讲数据模型就是对现实世界的模拟映射。

→数据模型分类：

概念模型，也称信息模型。按用户的观点来对数据和信息建模，用于数据库设计。

逻辑模型：包括网状模型、层次模型、关系模型、面向对象数据模型、对象关系模型、半结构化数据模型等。

物理模型：对数据最底层的抽象，描述数据在系统内部的表示方法和存取方法、或者在磁盘或磁带上的存储方式和存取方法。

→数据模型的组成要素：

严格定义的一组概念的集合。精确描述了系统的静态特性、动态特性和完整性约束条件。

数据库管理系统（DBMS）：一种用于管理、维护和操纵数据库的大型软件系统。它通过提供统一的管理和控制，保证数据库的安全性和完整性，满足多个应用程序和用户在同一时刻或不同时刻建立、修改和查询数据库。DBMS 主要由管理模块、控制模块和各种数据库操作语言组成。

专栏 2：数据库管理系统的主要特性

独立性：通过 DBMS，不同应用程序和用户可以在同一数据库下操作，各数据之间互相独立，互不干扰。

共享性：数据库系统可以进行多用户共享，多个用户或应用程序能够同时访问同一数据库，进行数据的查询和修改。

集中性：通过 DBMS，可以将不同的数据集中到同一数据库中进行管理和维护，提高了数据的管理效率和数据间的联系性。

安全性：通过 DBMS，可以实现对数据的安全保护，数据访问控制和数据备份等功能。

一致性：通过 DBMS，不论是在单用户还是多用户的情况下，数据的一致性能得到保障。

（二）数据库分类

历经 70 年的不断发展，数据库已经衍生出多种技术路径和方向。为了更清楚地掌握数据库之间的特性差异，数据库产品可根据不同标准进行分类，常见的分类依据包括数据模型、架构模型、业务负载特征、部署方式、适配硬件架构等，其中数据结构类型是最常用的分类标准。

- 根据数据模型分类：关系型数据库、非关系型数据库。
- 根据架构模型分类：集中式数据库、分布式数据库。
- 根据业务负载特性分类：事务型数据库（OLTP）、分析型数据库（OLAP）、混合型数据库（HTAP）。

现阶段，最常见的数据库模型是关系型数据库和非关系型数据库，国内超过 90% 的数据库市场选择使用关系型数据库，故本文将以数据结构为重点对数据库进行研究。

关系型数据库（RDB）：一种采用关系模型来组织数据的数据库，而关系模型是由二维表及其联系组成的数据组织。在关系型数据库中，对数据的操作几乎全部建立在一个或多个关系表格上，通过对这些关联表的表格进行分类、合并、连接或选取等运算来实现数据的管理。

专栏 3：关系型数据库的优缺点
→主要优点： 使用表结构，格式一致，易于维护；使用通用的 SQL 语言操作，使用方便，可用于复杂查询；数据存储在磁盘中，安全性高。 →主要缺点： 读写性能比较差，不能满足海量数据的高效率读写；不节省空间，

建立在关系模型上，要遵循某些规则，比如数据中某字段值即使为空仍要分配空间；固定的表结构，灵活度较低。

非关系型数据库（NOSQL）：一种数据结构化存储方法的集合，可以是文档或者键值对等。作为关系型数据库的一个补充，能在特定场景和特点问题下发挥高效率和高性能。非关系型数据库可以分为 **key-value** 型（针对高性能并发读写场景）、文档型（针对海量数据访问场景）、列式数据库、图形数据库四种类型。

专栏 4：非关系型数据库的优缺点

→主要优点：

存储数据的格式可以是 **key-value** 形式、文档形式、图片形式等，使用灵活，应用场景广泛；速度快，效率高，NoSQL 可以使用硬盘或者随机存储器作为载体，而关系型数据库只能使用硬盘；海量数据的维护和处理非常轻松；具有扩展简单、高并发、高稳定性、成本低廉的优势；可以实现数据的分布式处理。

→主要缺点：

暂时不提供 SQL 支持，学习和使用成本较高；没有事务处理，没有保证数据的完整性和安全性。适合处理海量数据，但是不一定安全；功能没有关系型数据库完善。

（三）数据库架构标准

产业数字化加速产生了诸如图、流、时序和地图空间等多种数据类型的海量数据，计算机体系架构正从以 **CPU** 为中心的架构向多样性算力协同的对等架构演进。在多模数据和多样性算力双轮驱动下，数据库架构需要与时俱进，有效利用多样

性算力，进行资源的集约化管理和调度，实现多模数据的高效处理和数据价值挖掘。

数据库组件化架构标准以 SQL 标准作为应用接入的标准界面，并进一步往下延伸，分层定义分发、执行、存储引擎的交互接口，实现数据库架构的可组装可演进，打造面向未来的数据库组件化技术生态。

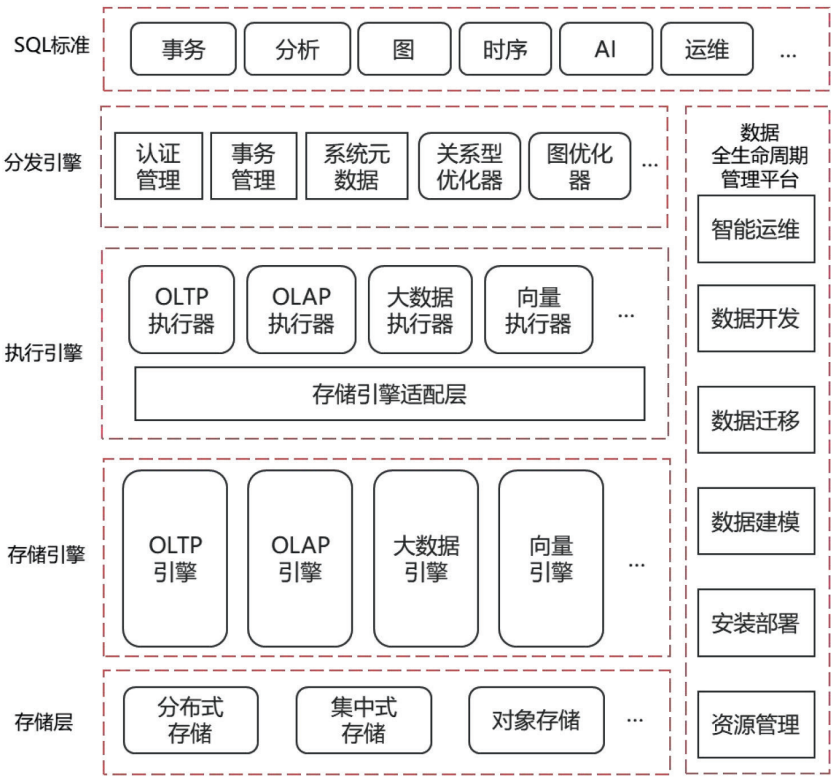


图 1 数据库组件化架构标准

图 1 所示的数据库组件化架构标准中，SQL 标准定义了全场景 SQL 标准；分发引擎实现 SQL 语句到执行计划的生成；执行引擎实现高效数据处理；存储引擎实现对多模数据的管理；存储层实现数据的透明化企业级存储能力；数据全生命周期管

理平台实现智能运维、集群管理等公共能力。

组件化架构标准在每一层都提供了扩展协同接口，以支持多样性的 SQL 接口、多模优化器、多模执行器、多模存储引擎。同时在部署形态上，每一层既可以单机部署，也可以集群部署，从而实现每一层基于业务负载的资源弹性伸缩。

（四）数据库产业链

数据库与芯片、操作系统并列为全球技术三大件，也是企业 IT 系统必不可少的核心技术。从图 2 所示的数据库上下游产业链图谱中可以看到，产业链上游主要是以 CPU 厂商、服务器厂商、网络设备厂商、存储厂商组成的网络和硬件厂商；中游主要是以数据库、操作系统和中间件等组成的基础软件厂商；下游主要是以各行业的集成商、开发商组成的应用开发厂商。

从产业链发展视角来看，数据库处在产业链中游枢纽位置，向上承接各类硬件、网络服务资源，向下提供数据存储、管理与分析服务，因此，数据库的地位和作用突显重要。



图 2 数据库上下游产业链图谱