

# 数据库发展研究报告

## (2021 年)

中国信息通信研究院云计算与大数据研究所  
2021 年 6 月

---

## 版权声明

---

本报告版权属于中国信息通信研究院，并受法律保护。  
转载、摘编或利用其它方式使用本报告文字或者观点的，应  
注明“来源：中国信息通信研究院”。违反上述声明者，本  
院将追究其相关法律责任。

## 编制说明

本报告的撰写得到了行业内许多专家的支持和帮助，他们分别来自：云和恩墨（北京）信息技术有限公司、成都虚谷伟业科技有限公司、华为技术有限公司、阿里云计算有限公司、北京科蓝软件系统股份有限公司、中国移动通信集团广东有限公司、湖南亚信安慧科技有限公司、星环信息科技（上海）股份有限公司、上海新炬网络信息技术股份有限公司、平凯星辰（北京）科技有限公司、北京奥星贝斯科技有限公司、上海热璞网络科技有限公司、北京快立方科技有限公司、北京滴普科技有限公司、上海爱可生信息技术股份有限公司、北京新数科技有限公司、北京坤中宇信息技术有限公司、北京优炫软件股份有限公司和武汉达梦数据库股份有限公司。在此表示由衷的感谢。

## 前 言

管理和分析数据，是人类进入信息时代后推动社会进步的关键环节。当前，随着数据要素市场化配置上升为国家战略，数据正式成为企业、产业乃至国家的战略性资源。数据库系统作为承载数据存储和计算功能的专用软件，经过半个多世纪的发展演进，已成为最主流的数据处理工具，是各企业数据工作流程的核心。2020 年，全球数据库市场规模达到 671 亿美元，中国数据库市场规模约为 240.9 亿元，占比约 5.2%，市场空间广阔。

当前，新一轮科技革命迅猛发展，数据规模爆炸性增长、数据类型愈发丰富、数据应用快速深化，促使数据库产业再次进入创新周期中的混沌状态。全球范围内创新型数据库产品快速涌现，市场格局剧烈变革，我国数据库产业进入重大发展机遇期。

本研究报告从技术、产业、应用三方面梳理了数据库发展情况，并展望了发展趋势。技术方面，梳理了数据库发展的三个关键阶段，探讨了未来数据库技术演进的主要方向；产业方面，从产业主体、研发模式、产品分布、推广策略等方面，分析了我国数据库产业的主体特点、市场格局、发展态势等；应用方面，以金融、政务、工业、互联网为代表，研究了其当前应用现状、问题以及下一步应用趋势。本报告还讨论了当前我国数据库行业存在的挑战，并给出相关建议。希望本报告的分析可以对业界提供参考，不当之处请多多指正。

# 目 录

一、 数据库是信息系统的核心.....	1
二、 数据库关键技术及发展趋势.....	2
(一) 数据库管理系统典型架构.....	2
(二) 数据库技术发展历程.....	4
(三) 数据库技术发展趋势.....	9
三、 数据库产业发展.....	14
(一) 数据库产业概述.....	15
(二) 数据库产品.....	16
(三) 数据库服务.....	25
(四) 数据库支撑体系.....	28
四、 数据库典型行业应用动态.....	33
(一) 金融行业&电信行业.....	34
(二) 政务行业.....	37
(三) 制造业.....	38
(四) 互联网.....	39
五、 总结与展望.....	41
六、 附录.....	42
(一) 数据库管理系统细分类别.....	42
(二) 数据库企业人才发展计划.....	44
(三) 中国信通院数据库应用创新实验室.....	44
(四) 中国信通院数据库评测体系.....	44
参考文献.....	46

## 图 目 录

图 1 数据库发展历程重要节点.....	2
图 2 数据库管理系统各模块架构图.....	3
图 3 网状模型与层次模型示意图.....	4
图 4 关系模型示意图.....	5
图 5 部分非关系模型示意图.....	7
图 6 集中式与分布式数据库架构示意图.....	8
图 7 数据库产业链全景图.....	15
图 8 中国数据库市场规模及增速.....	15
图 9 我国数据库企业成立时间分布情况.....	17
图 10 我国数据库企业总部分布情况.....	17
图 11 我国数据库企业员工数量分布情况.....	18
图 12 我国数据库企业专利数量分布情况.....	19
图 13 我国数据库产品分布情况.....	20
图 14 关系型数据库中基于主流开源数据库的分布情况.....	21
图 15 2021 年 6 月数据库开源与商业许可证流行度对比.....	24
图 16 数据库服务能力成熟度模型框架.....	26
图 17 全球各国数据库领域发文量及质量.....	28
图 18 2018-2020 年 VLDB、ICDE 和 SIGMOD 论文分布情况.....	29
图 19 全球数据库领域主要研究内容热点图.....	30
图 20 2018-2020 年中国高校及企业学术会议论文贡献情况.....	30
图 21 我国数据库初创企业融资情况.....	33

## 表 目 录

表 1 数据库典型行业应用特点.....	34
表 2 金融和电信行业数据库架构升级典型案例.....	35
表 3 金融行业非关系型数据库替换典型案例.....	36
表 4 金融行业核心系统数据库替换典型案例.....	37
表 5 政务行业非关系型数据库典型案例.....	38
表 6 制造业时序数据库典型案例.....	39
表 7 按数据模型分类的数据库管理系统.....	43
表 8 部分数据库企业人才培养一览表.....	44

## 一、数据库是信息系统的核心

获取与分析信息，是人类推动社会进步的关键过程。如何持续提高信息处理的能力，是人类社会的持续性命题。人类先后利用骨制品、绳结、纸张、算盘等工具，提高信息处理效率。在计算机发明之后，信息可经过其电子化编码后转化为数据。人类开始探索利用计算机算力优势，实现对数据的高效存储与计算，从而大幅提升信息处理效率。人类首先尝试开发出各类特定计算机应用程序完成相关工作，但由于各类特定应用程序复用性差，人类逐渐围绕数据存储和计算功能、抽象出满足组织信息化需求、逻辑关联的数据库。

数据库是支持一个或多个应用领域，按概念结构组织的数据集合，其概念结构描述这些数据的特征及其对应实体间的联系<sup>1</sup>。数据库中的数据按一定的数据模型组织、描述和存储，具有较小冗余度、较高数据独立性和易扩展性，并可为各种用户共享<sup>2</sup>。数据库由于综合成本低、处理能力高，扮演各类信息系统的核心角色。

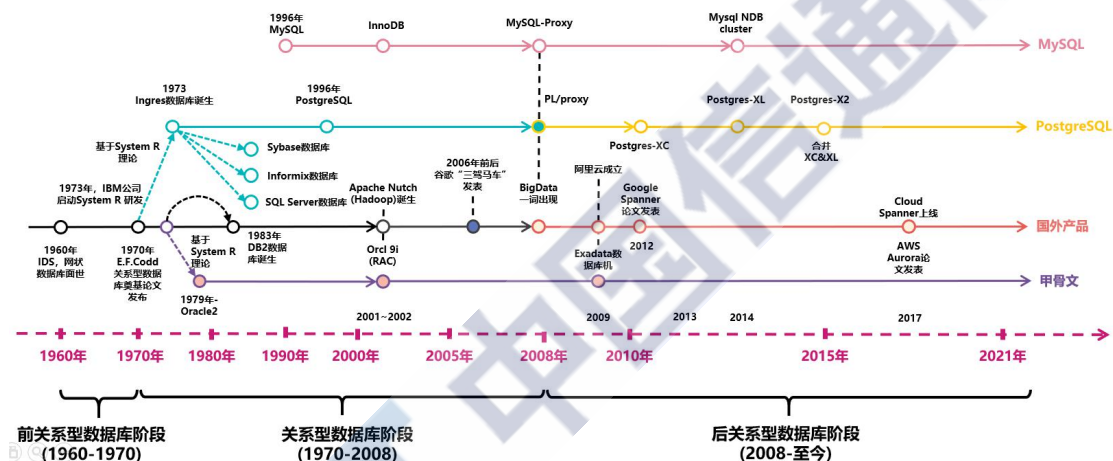
首款企业级数据库产品诞生于上世纪 60 年代，六十余年发展过程中，数据库共经历前关系型、关系型和后关系型三大阶段。前关系型阶段数据库的数据模型主要基于网状模型和层次模型，代表产品为 IDS 和 IMS，该类产品在当时较好地解决了数据集中存储和共享的问题，但在数据抽象程度和独立性上存在明显不足。关系型阶段以 IBM 公司研究员 E.F.Codd 提出关系模型概念，论述范式理论作为开启标志，期间诞生了一批以 DB2、Sybase、Oracle、SQL Server、MySQL、

<sup>1</sup> GB/T 5271.17-2010《信息技术 词汇 第 17 部分：数据库》

<sup>2</sup> 《数据库系统概论（第 5 版）》，王珊、萨师煊。



PostgreSQL 等为代表的广泛应用的关系型数据库,该阶段技术脉络逐步清晰、市场格局趋于稳定。谷歌的三篇论文<sup>3</sup>开启后关系型数据库阶段,该阶段由于数据规模爆炸增长、数据类型不断丰富、数据应用不断深化,技术路线呈现多样化发展。随着各行业数字化转型不断深入,5G、云计算等新兴技术快速发展,传统数据库的应用系统纷纷优化升级。全球市场格局剧烈变革,我国数据库产业进入重大发展机遇期。



来源：中国信息通信研究院、虚谷伟业

图1 数据库发展历程重要节点

## 二、数据库关键技术及发展趋势

数据库管理系统作为能够使用户定义、创建、维护和控制访问数据库的软件系统<sup>4</sup>,其整体架构与技术路线不断深化发展,如今呈现集中式与分布式并存,数据模型不断拓展等技术现状。

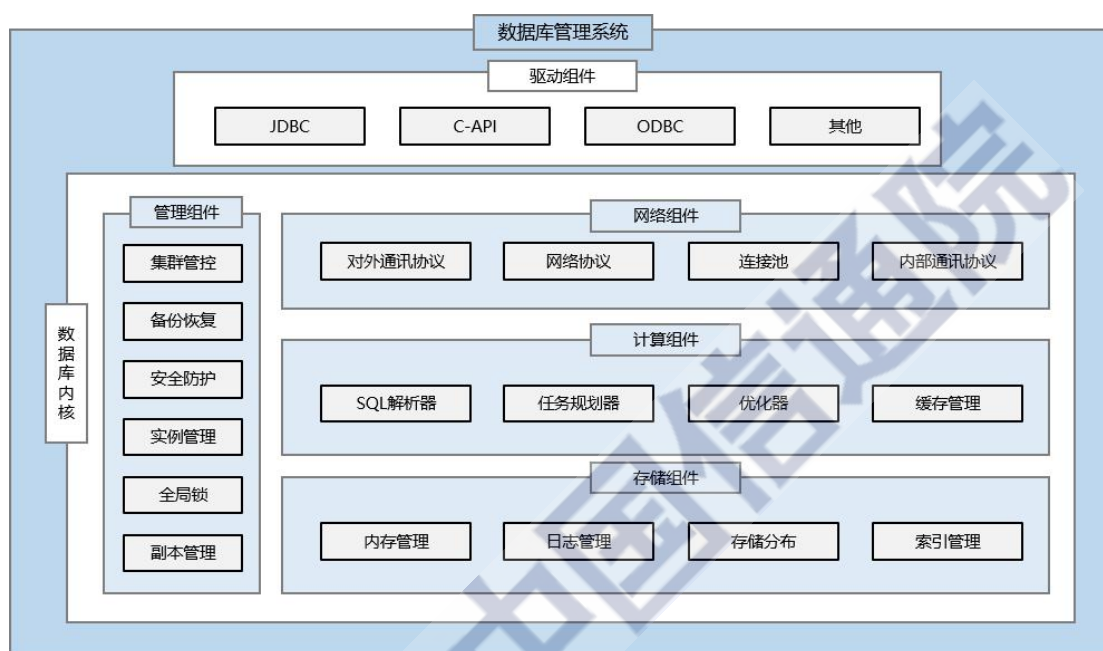
### （一）数据库管理系统典型架构

数据库管理系统由于不同产品实现细节不完全相同,此处仅对部

<sup>3</sup>Google File System、Google Bigtable 和 Google MapReduce

<sup>4</sup>TM Connolly, CE Begg, Database Systems: A Practical Approach to Design, Implementation and Management 4th Edition (England: Addison Wesley, 2005), p.16

分主流数据库产品做进一步抽象处理得出上述架构。数据库大致可以由内核组件集与外部组件集共同组成，其中外部组件集以数据库配套的独立支撑软件为主，例如数据库驱动。内核组件集则一般可以分为管理组件、网络组件、计算组件、存储组件四大模块。



来源：中国信息通信研究院

图2 数据库管理系统各模块架构图

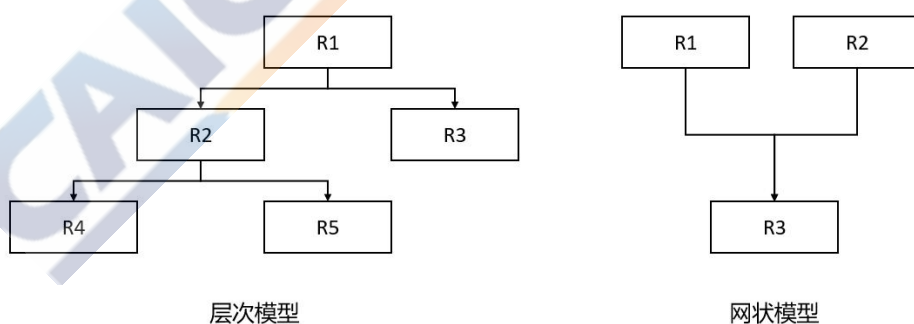
存储组件是负责数据持久化存储的组件，对数据库的日志、索引、堆数据等内容进行管理。在新一代的存算分离体系下，数据库堆数据的存储可能是由外部的分布式存储系统承担；计算组件又可以称为协调组件、服务组件，负责响应数据库访问请求，并将 SQL 语言解析成为数据库对应的内部任务。计算组件在分布式、集群等架构下也承接大部分的计算任务，例如排序、联接等；管理组件用于对数据库全生命状态的管理，例如心跳管理、集群管理等，以及各类中心化任务承接，如死锁仲裁、存储映射管理、元数据管理、事务号管理等；网络组件管理整个数据库管理系统的网络通讯的组件。数据库的网络通

讯有内部和外部之分。内部一般指在集群环境或者分布式环境下的各节点之间的高速数据交换。外部一般指的是各个数据库通过对外访问协议与存在于客户端的驱动进行互联的网络交换；驱动组件是支撑数据库能正常提供服务的配套独立组件，数据库管理系统基于其通用特性，往往可以对不同语言开发的软件提供数据服务。但是由于数据库本身只对外提供网络通讯协议，对协议的封装则由客户端侧的不同驱动组件完成。通常有支持 JAVA 语言的 JDBC 接口、支持 C 语言的 ODBC 接口和 C-API 接口等。

## （二）数据库技术发展历程

从 20 世纪 60 年代至今，数据库的发展历程大致分为以下三个阶段：

### 1. 前关系型阶段（1960–1970）：网状层次数据库初尝探索



来源：中国信息通信研究院

图 3 网状模型与层次模型示意图

1963 年，通用电气公司的 Charles Bachman 等人开发出世界上第一个数据库管理系统（以下简称 DBMS）也是第一个网状 DBMS——

集成数据存储（Integrated Data Store, IDS）。网状 DBMS 的诞生对当时的信息系统产生了广泛而深远的影响，解决了层次结构无法建模更复杂的数据关系的建模问题<sup>5</sup>。

同时期为解决“阿波罗登月”计划处理庞大数据量的需求，北美航空公司(NAA)开发出 GUAM(Generalized Update Access Method)软件。其设计思想是将多个小组件构成较大组件，最终组成完整产品。这是一种倒置树的结构，也被称之为层次结构<sup>6</sup>。随后 IBM 加入 NAA，将 GUAM 发展成为 IMS(Information Management System)系统并发布于 1968 年，成为最早商品化的层次 DBMS。

## 2.关系型阶段（1970-2008）：关系型数据库大规模应用



The diagram illustrates a relational model using a table. A callout box labeled '属性' (Attribute) points to the column headers. Another callout box labeled '元组' (Tuple) points to a row of data. The table contains five columns: '学号' (Student ID), '姓名' (Name), '年龄' (Age), '性别' (Gender), and '专业' (Major). It contains four rows of data representing individual students.

学号	姓名	年龄	性别	专业
001	张三	20	男	计算机
002	李四	25	男	软件工程
003	王五	23	男	自动化
004	赵六	22	女	电子工程

来源：中国信息通信研究院

图 4 关系模型示意图

第一阶段的 DBMS 解决了数据的独立存储、统一管理和统一访问的问题，实现了数据和程序的分离，但缺少被广泛接受的理论基础，同时也不方便使用，即便是对记录进行简单访问，依然需要编写复杂程序，所以数据库仍需完善理论从而规模化应用落地。第二阶段开启

<sup>5</sup>TM Connolly, CE Begg, Database Systems: A Practical Approach to Design, Implementation and Management 4th Edition (England: Addison Wesley, 2005), p. 24.

<sup>6</sup>出处同上

的标志性事件为 1970 年，IBM 实验室的 Edgar Frank Codd 发表了一篇题为《大型共享数据库数据的关系模型》论文，提出基于集合论和谓词逻辑的关系模型，为关系型数据库技术奠定了理论基础。这篇论文弥补了之前方法的不足，促使 IBM 的 San José 实验室启动验证关系型数据库管理系统的原型项目 System R，数据库发展正式进入第二阶段。

1974 年，Ingres 原型诞生，为后续大量基于其源码开发的 PostgreSQL、Sybase、Informix 和 Tandem 等著名产品打下坚实基础。1977 年，Oracle 前身 SDL 成立。1978 年，SDL 发布 Oracle 第一个版本。

20 世纪 80 年代，关系型数据库进入商业化时代。1980 年，关系型数据库公司 RTI（现名 Actian）成立并销售 Ingres，同年，Informix 公司成立。1983 年，IBM 发布 Database2 (DB2) for MVS，标志 DB2 正式诞生。1984 年，Sybase 公司成立。1985 年，Informix 发布第一款产品。1986 年，美国国家标准局 (ANSI) 数据库委员会批准 SQL 作为数据库语言的美国标准并公布标准 SQL 文本。1987 年，国际标准化组织 (ISO) 也做出了同样决定，对 SQL 进行标准化规范并不断更新，使得 SQL 成为关系型数据库的主流语言。此后相当长的一段时间内，不论是微机、小型机还是大型机，不论是哪种数据库系统，都采用 SQL 作为数据存取语言，各个公司纷纷推出各自支持 SQL 的软件或接口<sup>7</sup>。同年 5 月，Sybase 发布首款产品。

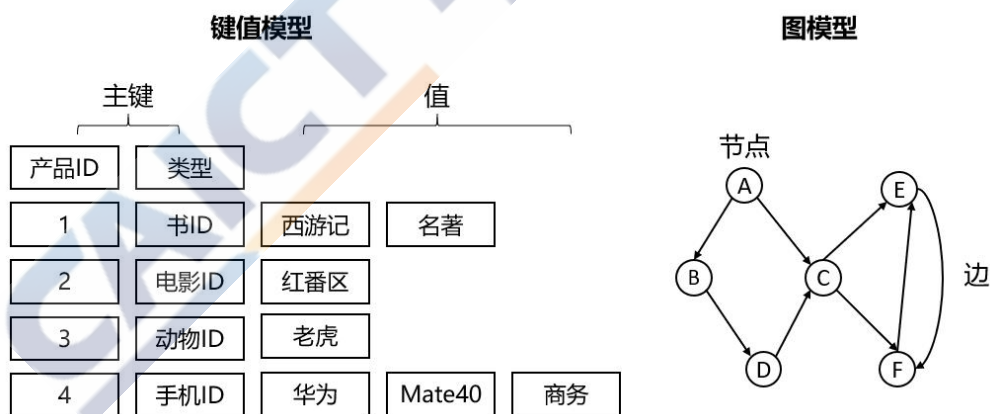
<sup>7</sup> 吴鹤龄. 关系数据库的标准语言——SQL[J]. 计算机研究与发展, 1989(06):7.



20 世纪 90 年代，Access、PostgreSQL 和 MySQL 相继发布。至此，关系型数据库理论得到了充分的完善、扩展和应用，在后关系型阶段，关系型数据库仍在发展演进，从未中止。

### 3.后关系型阶段（2008-至今）：模型拓展与架构解耦并存

进入 21 世纪，随着信息技术及互联网不断进步，数据量呈现爆发式增长，各行业领域对数据库技术提出了更多需求，数据模型不断丰富、技术架构逐渐解耦，一部分数据库走向分布式、多模处理、存算分离的方向演进。谷歌在 2003 至 2004 年公布了关于 GFS、MapReduce 和 BigTable 三篇技术论文，为分布式数据库奠定基础，Mike Stonebraker 提出“one size does not fit all”并依照此理念推出多种数据模型、存储介质的数据库，数据库发展正式进入第三阶段。

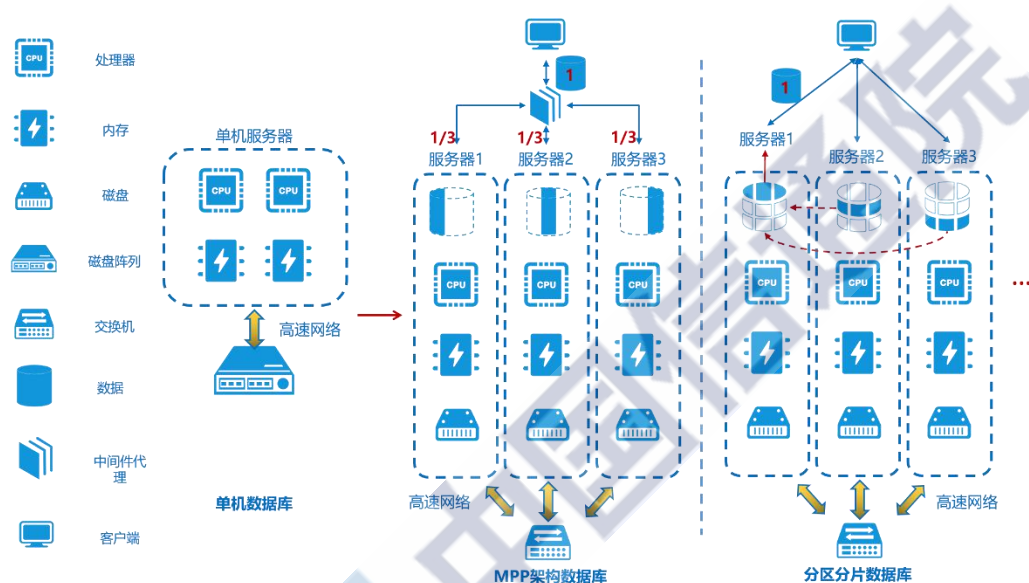


来源：中国信息通信研究院

图 5 部分非关系模型示意图

**数据模型不断拓展。**步入互联网 Web 2.0 和移动互联网时代，许多互联网应用表现出高并发读写、海量数据处理、数据结构不统一等特点，关系型数据库并不能很好地支持这些场景。另一方面，非关系

型数据库有着高并发读写、数据高可用性、海量数据存储和实时分析等特点，能较好地支持这些应用的需求。因此，一些非关系型数据库也开始兴起。为了解决大规模数据集合和多种数据类型带来的挑战，NoSQL 数据库应运而生，其访问速度快，适宜处理互联网时代容量大、多样性高、流动性强的数据。



来源：中国信息通信研究院

图6 集中式与分布式数据库架构示意图

**分布式架构逐渐成熟。**由于传统基于集中式数据库在应对海量数据及复杂分析处理时，存在数据库的横向扩展能力受限、数据存储和计算能力受限、不能满足业务瞬时高峰的性能等根本性的架构问题。利用分布式计算和内存计算等新技术设计的分布式数据库能够解决上述遇到的性能不足等问题。分布式数据库的数据分散在网络上多个互联的节点上，数据量、写入读取的负载均衡分散到多个单机中，集群中某个节点故障整个集群仍然能继续工作，数据通过分片、复制、分区等方式实现分布存储。每个数据节点的数据会存在一个或者多个

副本，提供数据冗余。当某个数据节点出现故障时，可以从其副本节点获取数据，避免数据的丢失，进而提升了整个分布式集群的可靠性。为保障分布式事务在跨节点处理时事务的原子性和一致性，一般使用分布式协议处理。常用两阶段提交、三阶段提交协议保障事务的原子性；使用 Paxos、Raft 等协议同步数据库的事务日志从而保障事务的一致性。分布式数据库技术架构大致可分为如下三类：

以 Apache Cassandra 、Apache HBase 为代表的分布式存储为基础的数据库，底层存储基于分布式文件系统具备了分片或者分区存储的能力，扩大了普通存储设备的存储系统的上限。

以 Greenplum 为代表的 Shared-Nothing 架构，通过多节点协同工作扩大分布式存储能力的同时，相应的还通过 MPP 架构可以支持多级并行计算处理，增强查询和分析能力。

以 Kylin 为代表的多维数据库产品，以及以 OpenTSDB 为代表的时序数据库，使用其它分布式数据库作为后台存储，通过构建相应的数据模型和索引技术，扩展成为新的数据库。

其他还包括分库分表等中间件解决方案，严格来说不属数据库系统，但是提供类分布式数据库解决方案，适用于合适的业务场景对分布式数据库的需求。

### （三）数据库技术发展趋势

大数据时代，数据量不断爆炸式增长，数据存储结构也越来越灵活多样，日益变革的新兴业务需求催生数据库及应用系统的存在形式愈发丰富，这些变化均对数据库的各类能力不断提出挑战，推动数据



库技术的不断演进，总结起来体现为三个方向：1) 多模数据库实现一库多用、利用统一框架支撑混合负载处理、运用 AI 实现管理自治，**提升易用性、降低使用成本**（下文趋势一、二、三）；2) 充分利用新兴硬件、与云基础设施深度结合，**增强功能、提升性能**（下文趋势四、五）；3) 利用隐私计算技术助力安全能力提升、区块链数据库辅助数据存证溯源，**提升数据可信与安全**（下文趋势六、七）。

### 1.趋势一：多模数据库实现一库多用

后关系型阶段，数据结构越来越灵活多样，如表格类型的关系数据、半结构化的用户画像数据以及非结构化的图片和视频数据等。面对这些多种结构的数据，应用程序对不同数据提出了不同存储要求，数据的多样性成为数据库平台面临的一大挑战，数据库因此需要适应多类型数据管理的需求。多模数据库支持灵活的数据存储类型，将各种类型的数据进行集中存储、查询和处理，可以同时满足应用程序对于结构化、半结构化和非结构化数据的统一管理需求。目前行业以 Azure Cosmos DB、ArangoDB、SequoiaDB 和 Lindorm 等多模数据库为典型代表。未来在云化架构下，多类型数据管理是一种新趋势，也是简化运维、节省开发成本的一个新选择。

### 2.趋势二：统一框架支撑分析与事务混合处理

业务系统的数据处理分为联机事务处理（OLTP）与联机分析处理（OLAP）两类。企业通常维护不同数据库以便支持两类不同的任务，管理和维护成本高。因此，能够统一支持 OLTP 和 OLAP 的数据库成为众多企业的需求。产业界当先正基于创新的计算存储框架研发

HTAP 数据库，其能够基于统一套引擎同时支撑业务系统运行和分析决策场景，避免在传统架构中，在线与离线数据库之间大量的数据交互。目前 HTAP 大致有两种实现方式，第一种是主备库物理隔离，主库运行 OLTP 负载，备库运行 OLAP 负载，主备之间通过重做日志进行数据同步。第二种是采用一体化设计，通过同一套引擎实现混合负载，区分 OLTP 与 OLAP 请求所在资源组，对资源组进行逻辑隔离，例如 Oracle 多租户隔离机制。HTAP 典型产品有 Oracle、SQL Server、Greenplum、TiDB、OceanBase 和 PolarDB 等。需要注意的是，HTAP 的价值在于更加简单通用，对于绝大部分中等规模的客户，数据量不会特别大，只需要一套系统即可，但对于超大型互联网企业，HTAP 数据库的分析性能可能不如专用 OLAP 数据库或大数据平台。

### 3.趋势三：运用 AI 实现管理自治

面对大规模数据和不同的应用场景，传统数据库组件存在业务类型不敏感、查询优化能力弱等问题。目前有研究通过将传统数据库组件用机器学习算法替代，来实现更高的查询和存储效率，自动化处理各种任务，例如自动管理计算与存储资源、自动防范恶意访问与攻击、主动实现数据库智能调优。机器学习算法可以分析大量数据记录，标记异常值和异常模式，帮助企业提高安全性，防范入侵者破坏，还可以在系统运行时自动、连续、无人工干预地执行修补、调优、备份和升级操作，尽可能减少人为错误或恶意行为，确保数据库高效运行、安全无失。2019 年 6 月，Oracle 推出云上自治数据库 Autonomous Database；2020 年 4 月，阿里云发布“自动驾驶”级数据库平台 DAS；

2021 年 3 月，华为发布了融入 AI 框架的 openGauss2.0 版本。其均采用上述思想降低数据库集群的运维管理成本，保障数据库持续稳定、高效运行。未来 80%以上的日常运维工作有望借助 AI 完成。

#### 4.趋势四：充分利用新兴硬件

最近十几年，新兴硬件在经历学术研究、工程化和产品化阶段发展，对数据库系统设计提供了广阔思路。期间最主要的硬件技术进步是多处理器（SMP）、多核（MultiCore）、大内存（Big Memory）和固态硬盘（SSD），多处理器和多核为并行处理提供可能，SSD 大幅提升了数据库系统的 IOPS 和降低延迟，大内存促进了内存数据库引擎的发展。根据第三方机构 Wikibon 预测，2026 年 SSD 单 TB 成本将低于机械硬盘，达到 15 美元/TB；非易失性内存（NVM）具有容量大、低延迟、字节寻址、持久化等特性，能够应用于传统数据库存储引擎各个部分，如索引、事物并发控制、日志、垃圾回收等方面；GPU 适用于特定数据库操作加速，如扫描、谓词过滤、大量数据的排序、大表关联、聚集等操作，互联网公司在 FPGA 加速进行了很多探索，例如微软利用 FPGA 加速网卡处理，百度用 FPGA 加速查询处理等。随着新型硬件成本逐渐降低，充分利用新兴硬件资源提升数据库性能、降低成本，是未来数据库发展的重要方向之一。

#### 5.趋势五：与云基础设施深度结合

Gartner 预测到 2022 年 75%的数据库将托管在云端。云计算技术的不断发展催生出将数据库部署在云上的需求，通过云服务形式提供数据库功能的云数据库应运而生。云与数据库的融合，减少了数据库

参数的重复配置，具有快速部署、高扩展性、高可用性、可迁移性、易运维性和资源隔离等特点，具体有两种形态，一种是基于云资源部署的传统数据库；另一种是基于容器化、微服务、Serverless 等理念设计的存算分离架构的云原生数据库。云原生数据库能够随时随地从多前端访问，提供云服务的计算节点，并且能够灵活及时调动资源进行扩缩容，助力企业降本增效。以亚马逊 AWS、阿里云、Snowflake 等为代表的企业，开创了云原生数据库时代。未来，数据库将深度结合云原生与分布式特点，帮助用户实现最大限度资源池化、弹性变配、超高并发等能力，更加便捷、低成本实现云上数字化转型与升级。

## 6.趋势六：隐私计算技术助力安全能力提升

随着数据上云趋势显著，云数据库面临的风险相较于传统数据库更加多样化、复杂化。如何解决第三方可信问题是云数据库面临的首要安全挑战。当前云数据库数据安全隐私保护是针对数据所处阶段来制定保护措施，如在数据传输阶段使用安全传输协议 SSL/TLS，在数据持久化存储阶段使用透明存储加密，在返回结果阶段使用数据脱敏策略等。这些传统技术手段可以解决单点风险，但不成体系，且对处于运行或者运维状态下的数据则缺少有效的保护。近年来以同态加密等密码学为代表的软件解决方案和以可信执行环境（TEE）为代表的硬件方案为数据库安全设计提供许多新思路。密码学方案的核心思路是整个运算过程都是在密文状态，通过基于数学理论的算法来直接对密文数据进行检索与计算。硬件方案的核心思路是将存放于普通环境（REE）的加密数据传递给 TEE 侧，并在 TEE 侧完成数据解密和



计算任务。基于隐私计算技术的数据库产品有 CryptDB、ZeroDB、openGauss 等。未来，此类数据库将围绕算法安全性和性能损耗等问题，逐步突破，进而提供覆盖数据全生命周期的安全保护机制。

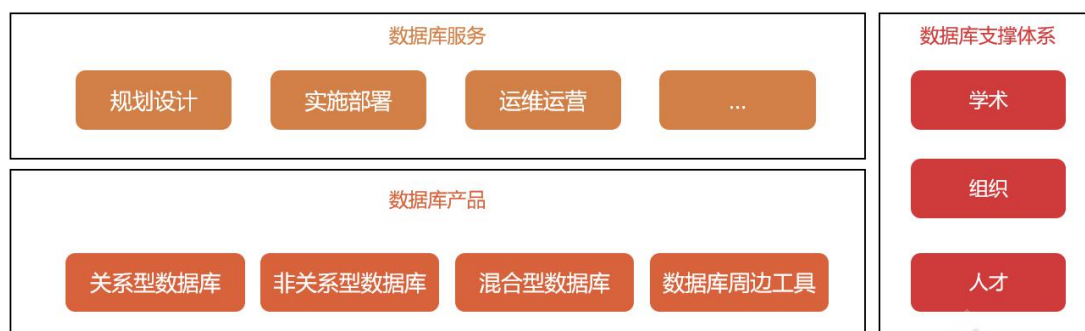
### 7.趋势七：区块链数据库辅助数据存证溯源

数据库管理员或黑客对数据库历史记录的修改是一个经常爆发的问题。区块链具有去中心化、信息不可篡改等特征，区块链数据库能够长期留存有效记录，数据库的所有历史操作均不可更改并能追溯，适用于金融机构、公安等行业的应用场景。区块链数据库典型产品有 BlockchainDB、BigchainDB 和 ChainSQL 等。区块链数据库由于要容忍节点拜占庭行为而不得不采用代价更高的 PBFT、PoW 等共识算法成为落地应用的一大挑战，此外，由于没有统一的协调者，如何保证区块链网络分片时分布式系统的安全性，高并发下的并行控制如何保证 ACID 也都是设计者不可忽视的问题。未来，提升区块链数据库性能将成为学术界与工业界共同探索的命题。

## 三、数据库产业发展

全球数据库产业生态成熟壮大，在发展过程中，逐渐细分出数据库产品、数据库服务和数据库支撑体系三个细分产业。

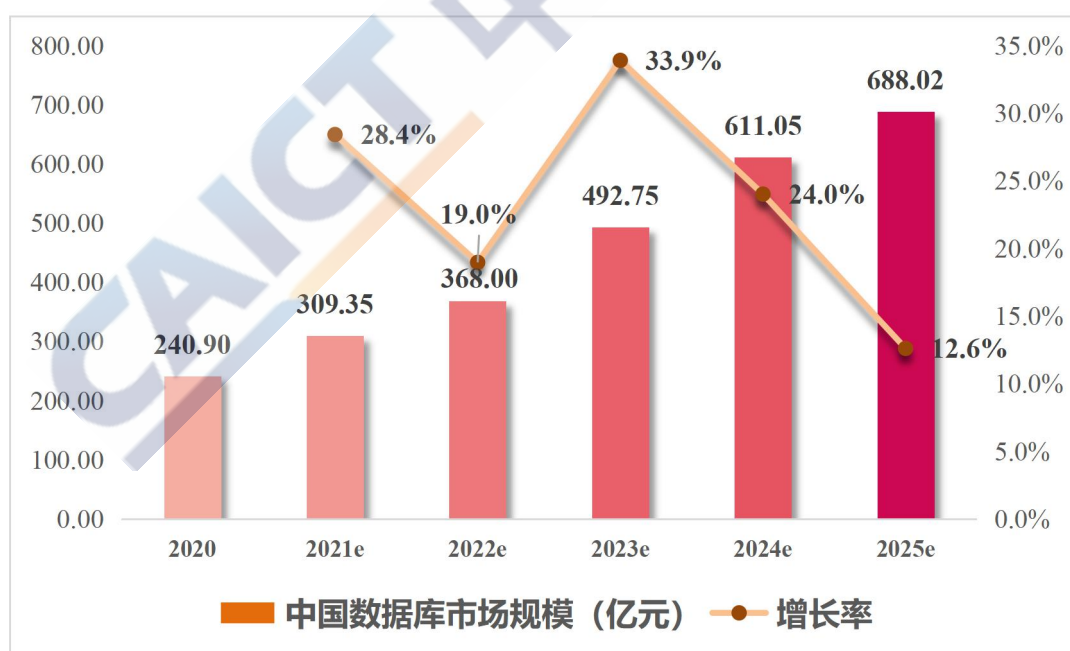
## （一）数据库产业概述



来源：中国信息通信研究院

图7 数据库产业链全景图

**数据库产品**主要由关系型数据库、非关系型数据库、混合型数据库及数据库周边工具构成。**数据库服务**是指围绕数据库的咨询规划、实施部署和运维运营等环节，为数据库系统的正常、高效、持续、安全使用提供信息技术服务工作。**数据库支撑体系**由从事数据库学术研究、人才培养、开源社区、评测认证等工作的相关主体共同构成。



来源：中国信息通信研究院，2021 年 6 月

图8 中国数据库市场规模及增速

据中国信通院测算<sup>8</sup>，2020 年全球数据库市场规模为 671 亿美元，其中中国数据库市场规模为 35 亿美元（约合 240.9 亿元人民币），占全球 5.2%<sup>9</sup>。

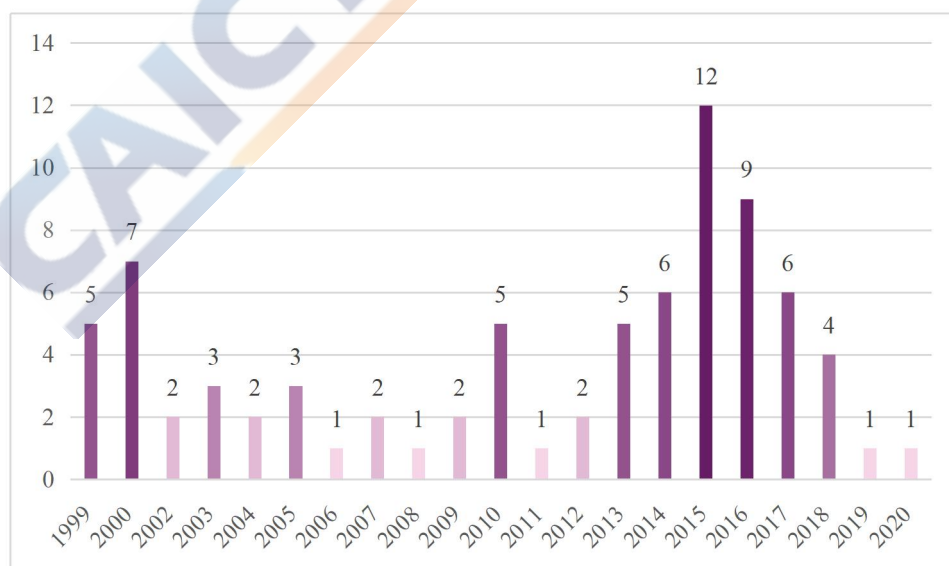
预计到 2025 年，全球数据库市场规模将达到 798 亿美元。中国的 IT 总支出将占全球 12.3%。我们预计，中国数据库市场在全球的占比将在 2025 年接近中国 IT 总支出在全球的占比，中国数据库市场总规模将达到 688 亿元，市场年复合增长率（CAGR）为 23.4%。

## （二）数据库产品

本节主要聚焦国内数据库产品提供商、产品分布及市场竞争三方面，并逐一展开分析。

### 1. 企业主体大部分仍处于发展初期阶段

据中国信通院统计分析，截止 2021 年 5 月底，我国数据库产品提供商共计 80 家。



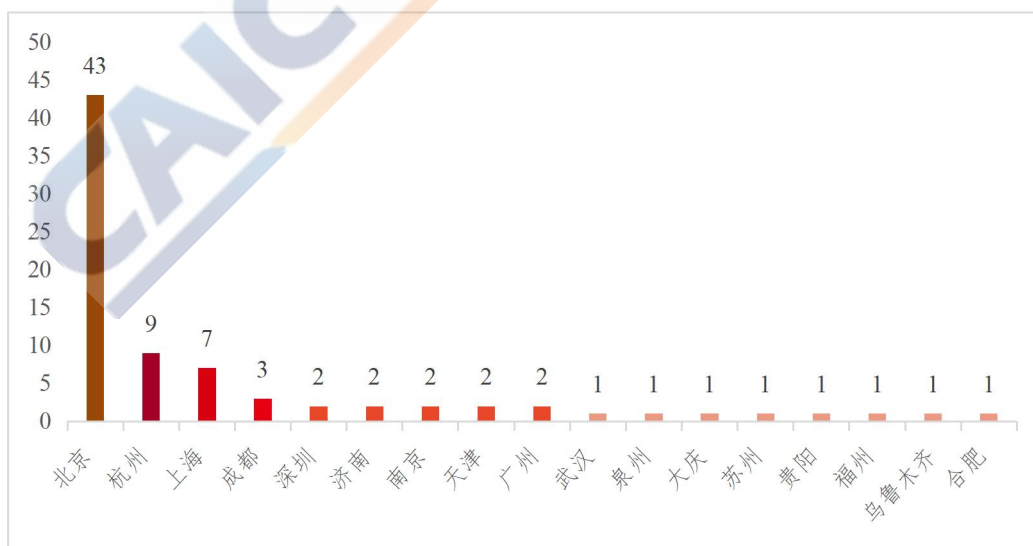
来源：中国信息通信研究院，2021 年 6 月

<sup>8</sup>市场规模为 2021 年统计，主要依据企业财报、人员访谈、评测评估、历史公开数据等得出。

<sup>9</sup>根据国家统计局显示，2020 年全年人民币平均汇率为 1 美元兑 6.8974 元人民币。

图 9 我国数据库企业成立时间分布情况

**成立时间呈现两个热周期。**从企业成立时间看<sup>10</sup>，我国数据库企业成立时间主要集中在 1999-2000 年和 2013-2017 年两个时间段，数量分别是 12 个和 38 个，依次占比 15%和 48%。上世纪 90 年代，以 Oracle、DB2 等为代表的国际商业数据库进入大陆市场，先后在电信、金融、政务等重要行业拿下大单，应用于各类核心系统和周边系统。国内也由于顶层设计加码，掀起一股国产数据库的浪潮。2012 年，大数据成为国家级发展战略。在此背景下，我国涌现出一大批以大数据和数据库为主营业务的初创公司。2015 年，平凯星辰、星瑞格、华胜信泰、上海丛云、恒辉信达等企业成立。2016 至 2018 年，图数据库和时序数据库关注度不断提升，以费马科技、创邻科技、欧若数网、蜀天梦图等为代表的初创图数据库企业相继成立，以浙江智臾、涛思数据等为代表的时序数据库企业不断涌现，政策利好与资本关注为我国数据库产业不断注入新活力，国产数据库产业迎来第二轮浪潮。



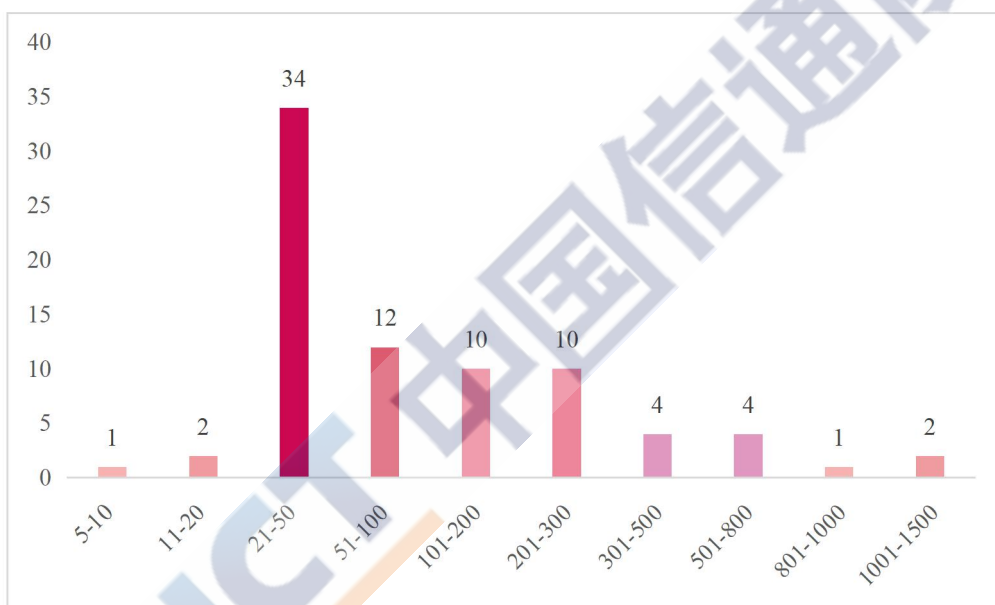
来源：中国信息通信研究院，2021 年 6 月

图 10 我国数据库企业总部分布情况

<sup>10</sup>成立时间以开展数据库业务为成立标志。



**地域分布以一线城市为主。**总部分布情况代表企业所在城市对数据库产业的重视与发展程度。从企业总部的数量看，由于人才规模聚集效应，企业总部通常设在超一线城市，数量最多的前五名是北京、杭州、上海、成都和深圳，分别是 43、9、7、3、2 个，占企业总数约为 54%、11%、9%、4%和 3%，除此之外，济南、南京、天津、武汉、广州、贵阳、福州、合肥和乌鲁木齐等直辖市和省会城市平均孵化出 1-2 个数据库企业。

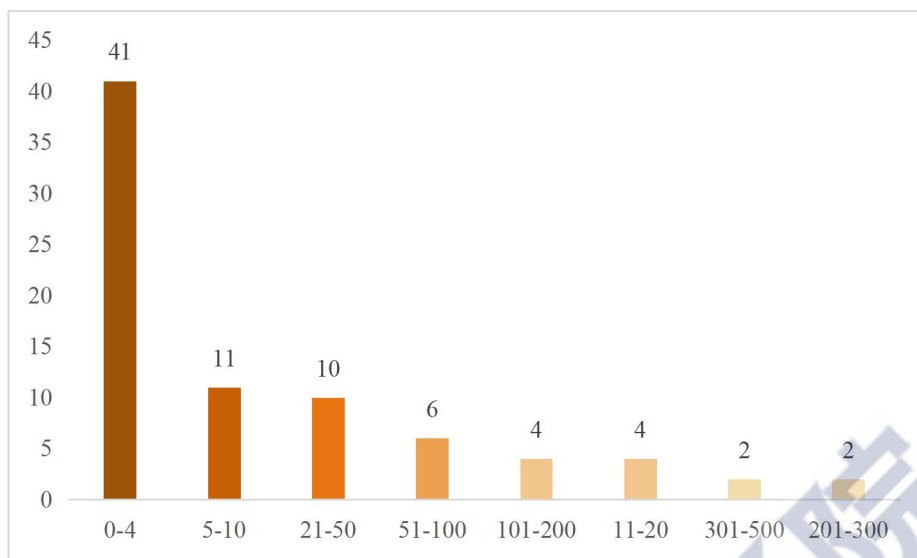


来源：中国信息通信研究院，2021 年 6 月

图 11 我国数据库企业员工数量分布情况

**员工数量普遍在百人以下<sup>11</sup>。**我国数据库企业人员平均人数约为 184 人，最高为 1200 人左右规模，最低为 10 人左右规模。其中 21-50 人左右规模企业占比最高，数量 34 个，比例达到 43%，人数在 51-100 人左右规模次之，数量为 12 个，占比 15%，101-200 人和 201-300 人规模并列第三，均为 10 个，分别占比 13%，由此可见，我国数据库虽然数量众多，但平均从业人员数量较少，仍在快速发展阶段。

<sup>11</sup>统计口径为数据库团队的研发、测试、技术支持等技术人员数量。



来源：中国信息通信研究院，2021 年 6 月

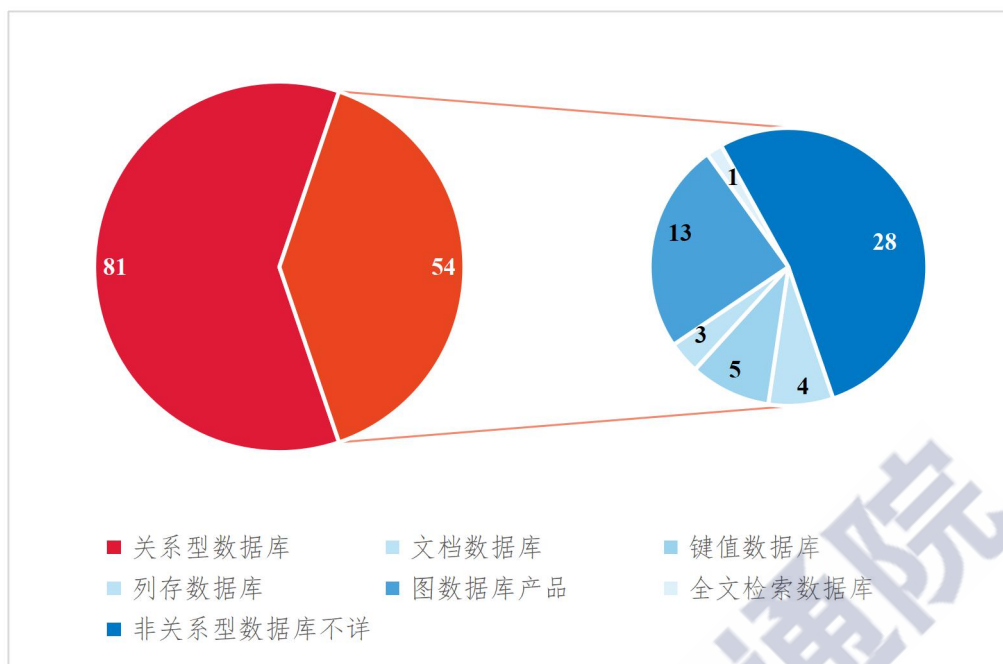
图 12 我国数据库企业专利数量分布情况

平均专利数量不足五十个<sup>12</sup>。我国数据库企业针对数据库领域的平均专利数量（含国内外专利）为 38 个，最高为 500 个左右规模，数量为 0 的企业个数是 19 个，占比 24%。拥有专利数 0-4 个的企业占比最高为 51%，专利数 5-10 个的企业次之，占比 14%，专利数 21-50 个的企业数量排名第三，占比 12%。从企业专利数量上看，Oracle 以 1.4 万个全球领先，SAP 居次席，国内数据库的全部企业技术专利累计千余，仍有较大发展空间。

## 2. 产品类型仍以关系型为主，非关系型产品正在快速发展

我国数据库产品数量分布呈现以关系型为主，非关系型及混合型数据库为辅的局面。数据库产品根据研发方式不同，分为完全自研和基于开源二次研发两类。

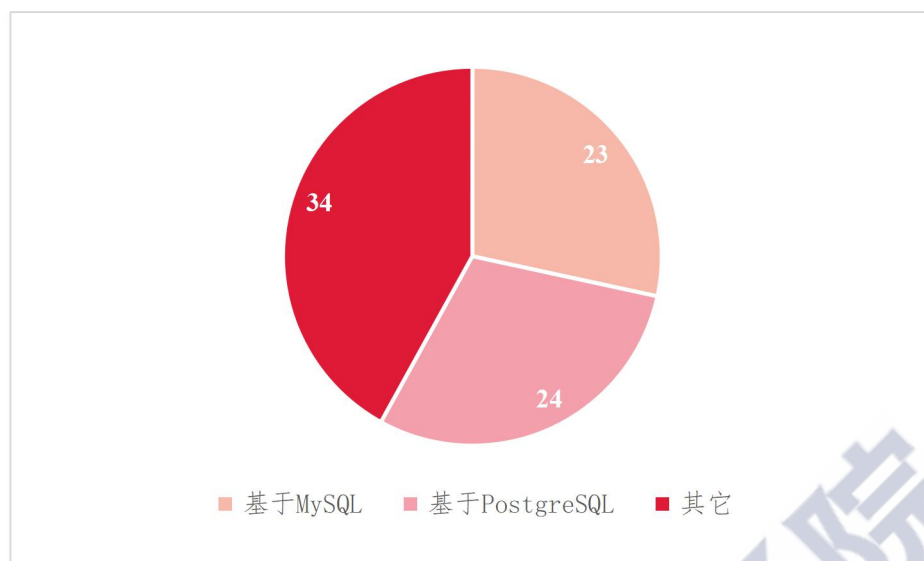
<sup>12</sup>统计口径包含国内及国际专利数量。



来源：中国信息通信研究院，2021 年 6 月

图 13 我国数据库产品分布情况

我国关系型数据库产品多数基于 MySQL 和 PostgreSQL 二次开发而来。据中国信通院统计分析，截止 2021 年 6 月，我国数据库产品共有 135 款。其中关系型数据库 81 个，非关系型数据库有 54 个，占比分别为 60%和 40%。按二级细分类别，以云服务为主要服务交付方式的关系型云数据库 19 个；非关系型数据库中，键值型数据库 5 个，列存数据库 3 个，文档数据库 4 个，图数据库 13 个，全文检索数据库 1 个，在非关系数据库中依次占比 9.26%、5.56%、7.41%、24.07%和 1.85%，由于一些不可控因素，其余非关系型数据库的数据模型暂不可知。



来源：中国信息通信研究院，2021 年 6 月

图 14 关系型数据库中基于主流开源数据库的分布情况

关系型数据库中基于开源数据库 MySQL 和 PostgreSQL 进行二次开发的个数分别为 23 和 24 个，依次占关系型数据库比例为 28.40% 和 29.63%，总计占 58.03%。

我国非关系型数据库产品发展势头良好，逐渐受到国际认可。非关系型数据库中以基于开源数据库如 Redis、InfluxDB、CouchDB 等产品进行二次开发为主。时序数据库因其存储处理海量时序数据的特性，常应用于工业控制、物联网、车联网等领域。据中国信通院统计分析，我国时序数据库从 2000 年后迅速发展，产品数量已达 15 款，1 款为开源数据库，其余均为商业数据库。4 个由云厂商提供，还有来自石化、电力、钢铁等传统工业实时数据库企业的产品 9 个。同时根据 DB-Engines 官网显示<sup>13</sup>，2021 年 5 月的时序数据库的流行度排名中，我国上榜的数据库产品已有两个，分别是浙江智臾和阿里云 TSDB，依次位列第 11 和第 21 名。由于图数据库能够支撑社交网络、

<sup>13</sup><https://db-engines.com/en/ranking/time+series+dbms>

金融反欺诈等互联网与金融场景的关联分析业务，所以行业关注热度自 2016 年以来逐渐升温。我国图数据库产品数量为 13 款，自研程度较高，自研产品占总数比例为 69.23%。从供应商类型看，初创公司、云厂商、高校纷纷入局，其比例分别为 7:5:1。根据 DB-Engines 官网显示<sup>14</sup>，2021 年 5 月的图数据库的流行度排名中，我国上榜的数据库产品有 3 个，分别是欧若数网 Nebula Graph、华为云 GraphBase 和百度智能云开源产品 HugeGraph，依次位列第 15、第 28 和第 30 名。

### 3. 市场份额正逐渐倾向云上，线下市场迎来激烈竞争

**线上市场呈现快速增长。**随着云计算技术不断成熟，云上数据库市场快速增长。根据 Gartner 公司 2019 年发布的市场分析，2017 年至 2018 年的全球数据库产业总营业额的 18.4% 增长中，云数据库管理产品的营业额占比 68%，而 Microsoft 与亚马逊的 AWS 占到总增长的 75.5%。同时根据 2021 年 Gartner 对数据库产品提供商的排名情况看，Microsoft 凭借云数据库的后发优势，抢占了 Oracle 已经持续占据十年的榜首位置，前十名中已经有四家以云服务为主要供应方式的企业，分别为 Microsoft、Amazon、Google 和阿里云。据中国信通院统计分析，2020 年，中国公有云数据库市场规模为 107.68 亿元，未来 5 年，公有云数据库市场年复合增长率将达到 36.1%，预计到 2025 年，中国公有云数据库市场总规模将达到 503.31 亿元。Gartner 预测到 2023 年，全球数据库市场中 75% 的数据库将完成到云平台的迁徙，仅有 5% 的数据保持在原本的本地模式当中。**线上市场格局巨头涌现。**

<sup>14</sup><https://db-engines.com/en/ranking/graph+dbms>



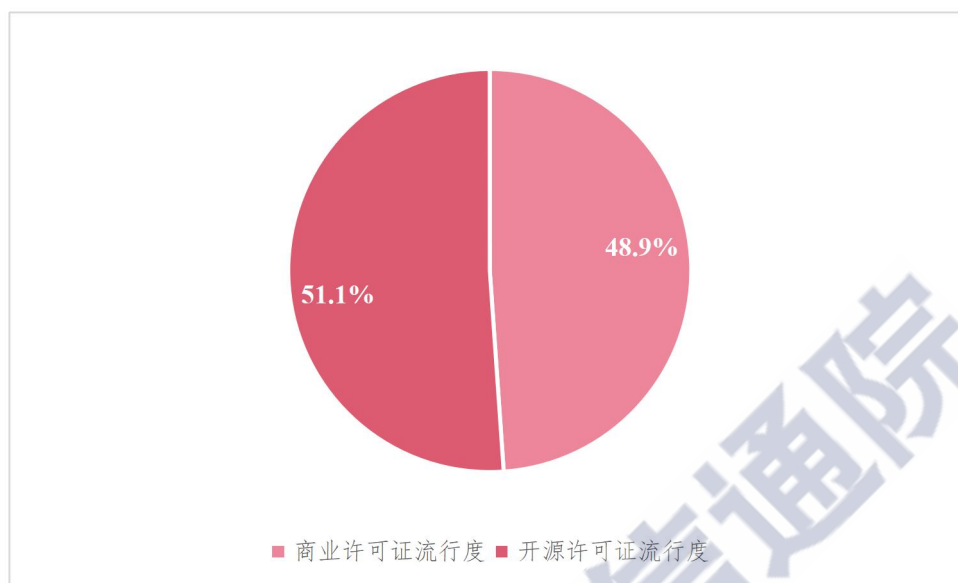
中国信通院调研显示，阿里云、华为云和腾讯云作为我国头部云服务商，其在云计算基础设施、应用生态、用户渠道等方面处于领先地位，云上数据库作为云基础设施的延续与发展，具备天然先发优势。2020 年，上述三家云服务商公有云数据库总营收约占中国公有云数据库市场份额 75.5%<sup>15</sup>。以电商、游戏、短视频等为主营业务的互联网公司是线上数据库的服务对象。出于对业务特性和生态兼容考虑，约 83% 的云上客户倾向选择 MySQL、Redis、MongoDB、InfluxDB 等开源数据库。

**存量市场替换空间可观，线下市场迎来激烈竞争。**2020 年数据库传统部署模式市场为 133.22 亿元，随着市场倾向的变化，传统部署市场替换国外数据库空间巨大。以关系型数据库为例，2017 年以前市场格局十分稳定，Oracle、IBM、Microsoft、Teradata 等为代表的产品占据数据库传统部署模式市场份额 90% 以上，以达梦、人大金仓、南大通用、神舟通用为代表的国产数据库，通常聚焦于军工、政务等封闭领域，整体市场份额较小，如今电信、金融等重要行业数据库改造变更需求不断，相关存量市场前景诱人。随着技术层面的分布式改造需求不断以及市场层面自发选择国产产品倾向，国产数据库市场份额有望得到大幅提升，各企业纷纷抢抓战略机遇，不断迭代打磨产品能力，抢占市场份额。据中国信通院大数据产品能力评测十二批<sup>16</sup>结果显示，国产数据库供给能力较几年前得到大幅提升，产品功能

<sup>15</sup>数据来源：中国信息通信研究院统计，主要依据人员访谈、评测评估、历史公开数据得出。

<sup>16</sup>中国信通院大数据产品能力评测是中国信息通信研究院下属的大数据及数据库评测品牌，每半年一批评测，至今已开展十二批。

逐渐完善，集群规模与日俱增，性能表现不断攀升，市场竞争程度较为激烈。



来源：DB-Engines 官方网站，2021 年 6 月

图 15 2021 年 6 月数据库开源与商业许可证流行度对比

**初创企业和巨头陆续投身开源市场。**开源已成为数据库产业的共识，2021 年 1 月，DB-Engines 官网显示<sup>17</sup>，开源许可证流行度首次超过商业许可证，开源数据库迎来新纪元。截止 2021 年 6 月，开源与商业许可证数量分别为 192 和 179 个，流行度分别占比 51.1%和 48.9%。近些年以巨杉、平凯星辰、涛思数据、欧若数网为代表的初创企业和以百度、华为、阿里云、蚂蚁金服为代表的巨头意识到开源有助于扩大人才规模及上下游生态影响力，通过运营开源社区快速获得反馈并加快产品开发、提升产品质量，同时反哺社区开发者及独立软件开发商（ISV）等生态伙伴，能够达到多方共赢目的。针对开源，企业纷纷采取不同的商业模式<sup>18</sup>。2014 年 12 月，巨杉数据库宣布开源

<sup>17</sup>[https://db-engines.com/en/ranking\\_osvsc](https://db-engines.com/en/ranking_osvsc)

<sup>18</sup>数据库领域催生三种开源商业模式：一是完全开源式，借助基金会托管，通过捐赠覆盖运营成本和获取利润，以 HBase、HAWQ 为代表；二是开源版本和商业版本分别运营，通过运营开源版本社区获取人才、应用场景、市场品牌、技术等，再通过售卖商业版本获取利润，以 MySQL 为代表；三是过渡式开源运作，

SequoiaDB，成为国内最早开源自研数据库项目；2017 年 10 月，平凯星辰开源 TiDB；2018 年，百度开源数据库 Doris 和 HugeGraph；2019 年，涛思数据和欧若数网分别开源 TDengine 和 NebulaGraph；2020 年 6 月，华为建立 openGauss 开源社区，并于 2021 年 3 月发布第一个 Release 版本；2021 年 5 月，阿里云宣布对外开放关系型数据库 PolarDB for PostgreSQL 源代码，同年 6 月，蚂蚁集团宣布开源 OceanBase。

### （三）数据库服务

很长一段时间内，我国数据库服务工作主要以附属技能的形式由应用开发商和硬件服务商提供保障。随着数据库对于企业的重要性越来越高，企业对于以数据库为核心的专业服务的需求也越加迫切，独立的数据库服务厂商开始崭露头角，并形成了一个专业化的数据库细分服务领域。

#### 1. 头部企业主体发展时间较长

数据库服务产业主体主要由多年来在电信、金融、政务等重要行业提供外包 IT 运维服务的企业构成，成立时间普遍十年以上，核心成员多为早期提供 Oracle、DB2 原厂或第三方服务的专家，由于企业数据库技术体系庞杂，需要服务提供商能够提供横向主流数据库产品和纵向多版本技术服务覆盖能力，服务行业技术壁垒较高。此外，由于一般与客户签订一至三年合同，服务提供商对客户系统非常熟悉，容易形成相对稳定的长期合作伙伴关系，市场壁垒较高，新兴初创公

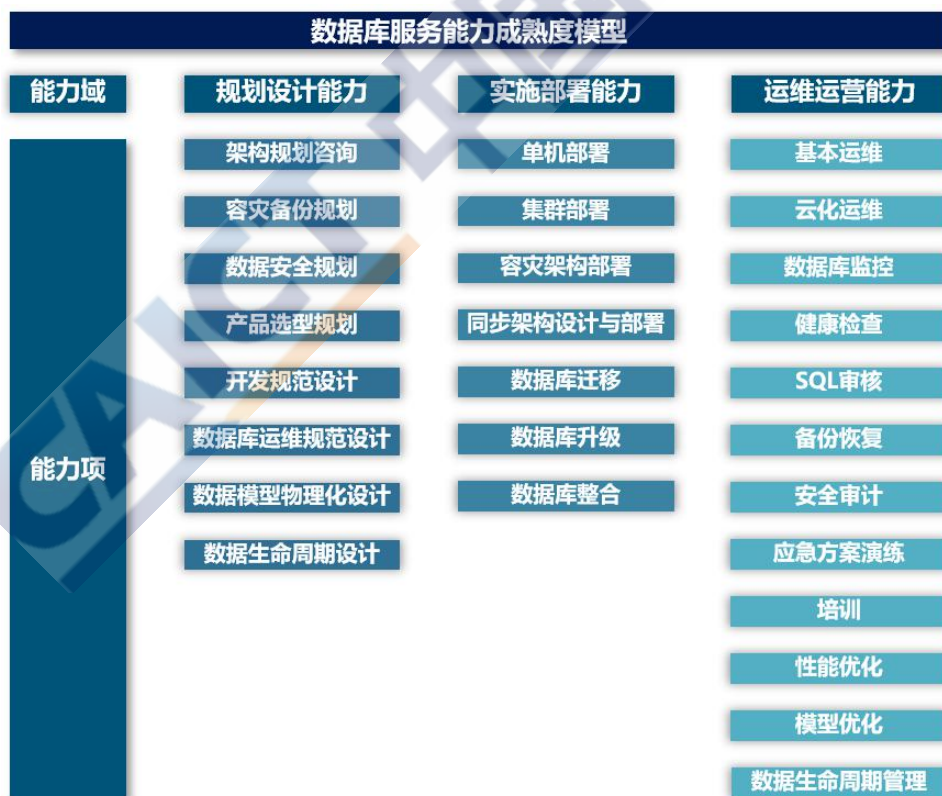
前期通过开放社区制造影响力和吸收外部技术力量，后通过修改商业许可证转为商业产品进行售卖，并停止社区的维护，为收购和独立上市铺垫，以 Neo4j 为代表。



司较少，巨头员工数量普遍在千人左右。头部典型企业有云和恩墨、新炬网络、海量数据、太阳塔、爱可生、中亦安图、万国数据、银信科技、天玑科技、新数科技、沃趣科技、迪思杰、九桥同步等。

## 2.服务工作范围广，缺乏行业规范和指引

数据驱动时代，企业开始利用海量实时数据分析业务发展、了解客户行为和优化配置资源，并据此制定企业发展战略。数据库系统作为数据存储的主要载体，数据库服务贯穿企业 IT 系统的整个生命周期。按照信息系统建设的不同阶段，数据库的服务范围主要覆盖规划设计、实施部署、运维运营三个方面，三个方面又细分多个服务工作内容，如下图所示。



来源：中国信息通信研究院

图 16 数据库服务能力成熟度模型框架

由于数据库服务产业正处于快速变革期，玩家众多，能力水平参

差不齐，服务过程缺乏行业规范和指引，导致众多数据库应用单位面临各类选型和实施问题。中国信息通信研究院联合国内数据库厂商和服务商，共同编制并发布了《数据库服务能力成熟度模型》团体标准，期望为国内的数据库服务生态体系提出更全面和专业的评估标准。

### 3.服务市场集中在重点行业，环境变革倒逼各主体转型升级

**服务市场主要集中在金融、电信、政府、制造、交通五个行业。**根据中国信通院统计分析<sup>19</sup>，各行业的数据库服务市场份额比例分别为金融 22.3%、电信 18.9%、政府 16.4%、制造 13.3%、交通 9.6%，这五个行业合计占比超过 80%。

**云计算改变传统服务市场格局。**在云计算逐渐成熟之后，与云上数据库市场份额迅速扩大。而云上数据库的咨询、部署、运维等服务工作则直接由云计算公司负责，所以云计算公司将改变传统服务市场格局。由于服务工作定制化程度较高，相对标准化云产品，属于劳动力密集型工作，所以从综合成本角度出发，未来云上数据库服务市场部分将由云计算公司依靠自身资源储备负责，另一部分云计算公司将与线下服务公司进行合作，形成优势互补，共同完成相关工作。

**服务企业向产品企业转型。**如今随着分布式云数据库兴起，数据库运维要求不断提升，数据库服务商除了提供传统的驻场与远程运维类服务外，围绕数据库开发、测试、运维等环节也提供多种类型的数据库周边工具。与此同时，以云和恩墨、新数科技、爱可生、海量数

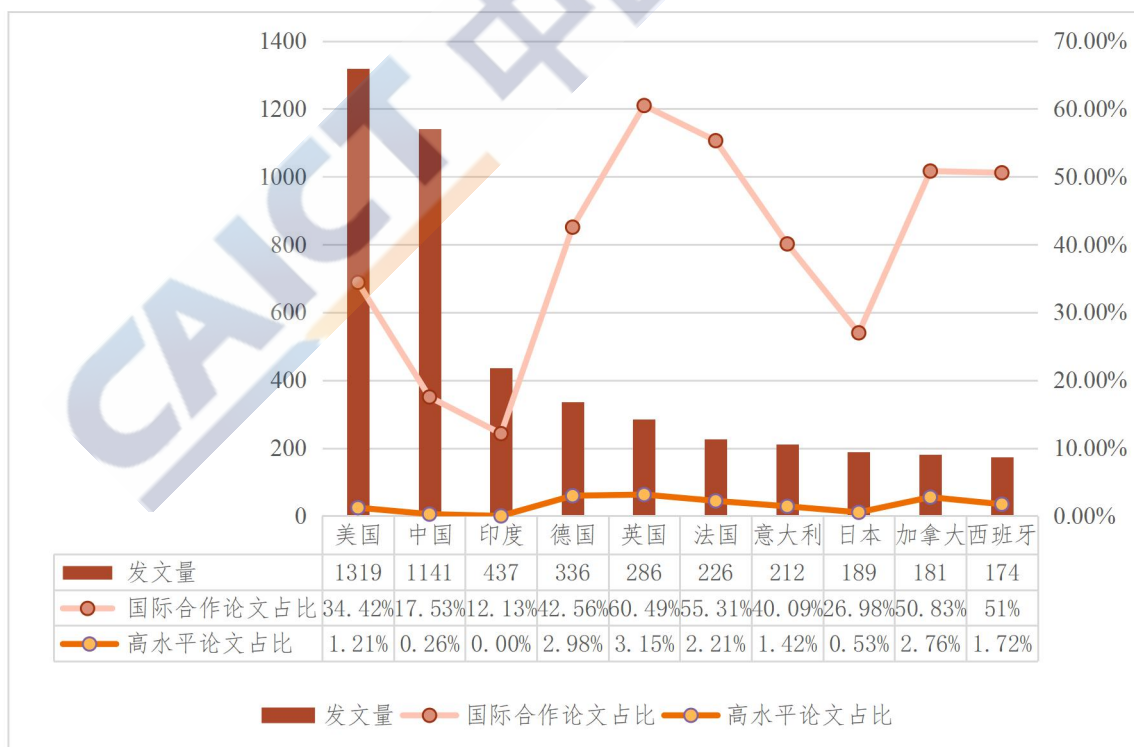
<sup>19</sup>数据来源：中国信息通信研究院统计，主要依据人员访谈、评测评估、历史公开数据得出。

据等为代表的数据库服务商为了拓展业务范围，提升企业利润总额，认识到可以利用自身服务能力积累与经验，对数据库产品供应商形成差异化优势，顺势推出自有数据库产品，进一步加剧了数据库产品市场竞争激烈程度。其中云和恩墨与海量数据分别基于开源数据库 openGauss 推出了企业级数据库 MogDB 和 Vastbase，爱可生和新数科技基于 MySQL 也发布了相关产品。

#### （四）数据库支撑体系

当前数据库支撑体系由于数据库技术路线不断演进，也正处于变革和创新的高发期。

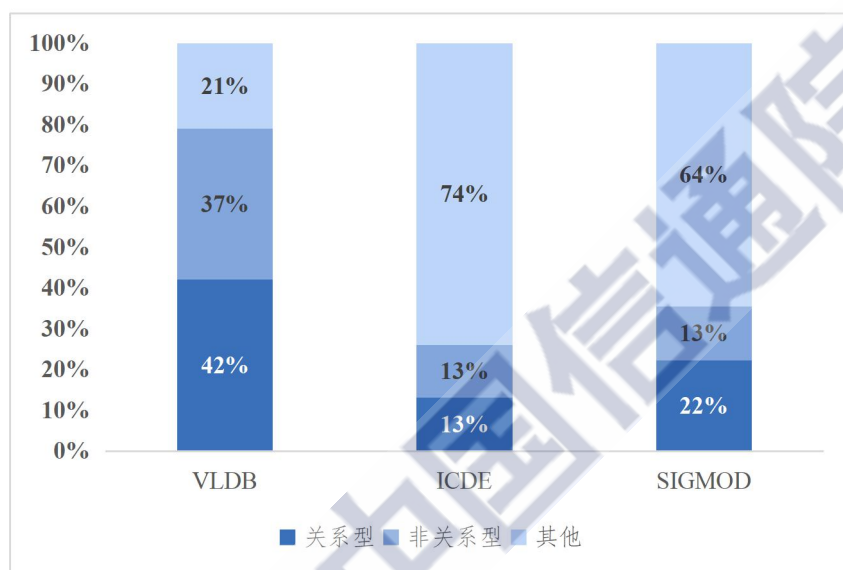
##### 1. 学术研究仍以关系理论为重点，国内研究水平逐渐提升



来源：中国信息通信研究院，2021 年 6 月

图 17 全球各国数据库领域发文量及质量

2016 至 2020 年，美国、中国、印度、德国和英国是全球数据库领域论文产出前五的国家，美国发文量最多，占全球总发文量 22.4%，之后依次为中国 19.4%，印度 7.4%<sup>20</sup>。从高水平论文数量分析，英国高被引论文数占 3.1%，中国占 0.3%。从国际合作论文的角度分析，英国、法国、加拿大、西班牙的国际合作论文较多，均超过 50%。



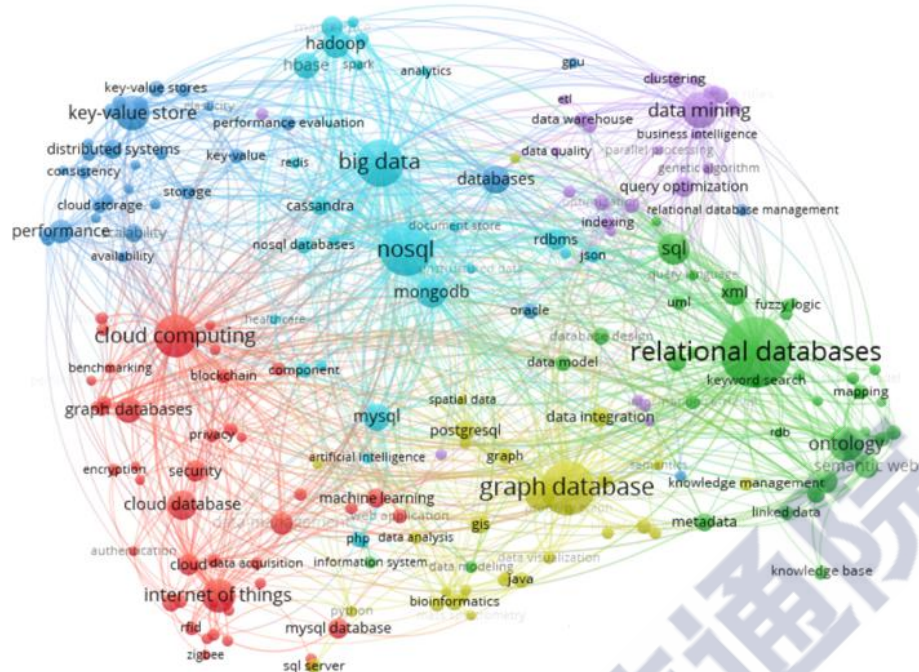
来源：公开资料整理

图 18 2018-2020 年 VLDB、ICDE 和 SIGMOD 论文分布情况

学术界公认的数据库领域顶级会议主要有 VLDB、SIGMOD 和 ICDE。从这些会议的研究方向看，当前以关系型数据库为主，非关系型数据库为辅。以 VLDB 为例，2018-2020 年，各领域论文总数分别为 146、151 和 95 个，关系型和非关系型数据库论文分别占三年论文总数量的 37%和 21%。SIGMOD 各领域论文总数分别为 90、88 和 144 个，关系型和非关系型数据库论文总数占三年论文总数的 22%和 13%。ICDE 各领域论文总数分别为 188、268 和 241 个，关系型和非关系型数据库论文总数占三年论文总数均为 13%。

<sup>20</sup>基于 Web of science 核心合集，选取 SCI-EXPANDED、CPCI-S 数据库。检索 2011-2020 年关于数据库领域的论文。检索时间为 2021 年 6 月。

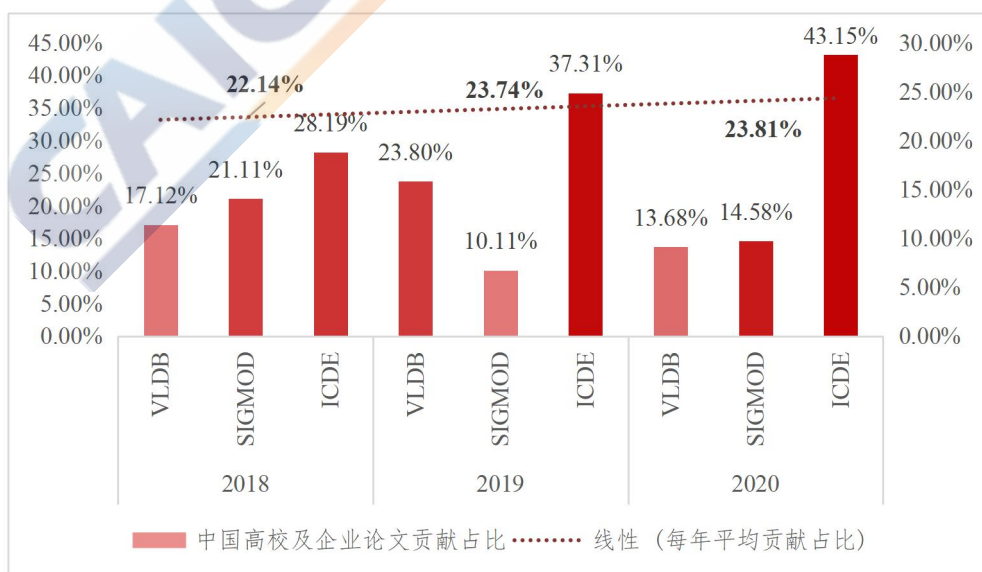




来源：中国信息通信研究院，2021 年 6 月

图 19 全球数据库领域主要研究内容热点图

综合分析全球论文研究主题，除了关系型数据库，图论、图数据库、查询优化、机器学习、分布式处理、时序数据、流数据、时空数据、云数据库等代表当前火热的技术方向。此外，数据安全、隐私保护也是每年不可或缺的研究主题。



来源：中国信息通信研究院根据公开资料整理

图 20 2018-2020 年中国高校及企业学术会议论文贡献情况

我国在全球数据库领域学术影响逐渐提升。高校及企业在 ICDE 论文贡献占比最高，三年依次为 28.19%、37.31%和 43.15%，三大会议每年贡献占比平均为 22.14%、23.74%和 23.81%，数量呈逐年上升趋势，研究方向以图论、图数据库、数据挖掘、机器学习、查询处理等方向为主。阿里巴巴、华为、腾讯、蚂蚁金服、百度、PingCAP 等企业和清华大学、香港科技大学、北京大学、香港中文大学、香港大学、浙江大学、华中师范大学、华东师范大学、中国人民大学、复旦大学、北京航空航天大学、华中科技大学、中国科学院、北京理工大学等高校论文纷纷入选三大顶会，显示我国数据库学术水平国际影响力不断扩大。

## 2. 领域内各类组织形成，产业热度不断提高

数据库支撑体系各类组织主要分为以下四类：

一类是由具备官方背景的研究组织，例如以中国计算机学会（CCF）数据库专业委员会为代表的学术组织和以通信标准化协会大数据技术标准推进委员会（CCSA TC601）为代表的行业组织，用于汇聚国内数据库理论研究头部力量；第二类是数据库从业人员牵头发起的面向数据库技术爱好者的用户组织，如面向 DBA 的 ACDU、面向 Oracle 用户的 ACOUG、面向 MySQL 用户的 ACMUG、面向 PostgreSQL 用户的中国开源软件推进联盟 PostgreSQL 分会等，用于进行各类专题技术交流和讨论；第三类是由数据库企业组建，针对自身特定产品讨论的官方技术社区，如阿里云开发者社区、华为云 openGauss 社区、PingCAP AskTUG 社区、PostgreSQL 中文社区、爱

可生开源社区、移动云开发者社区等；**第四类是汇聚数据库整体行业信息的第三方技术社区**，如 ITPUB、墨天轮、DBAplus 等，用于搭建领域内线上交流平台。

### 3. 多层次数据库人才培养体系正在快速形成

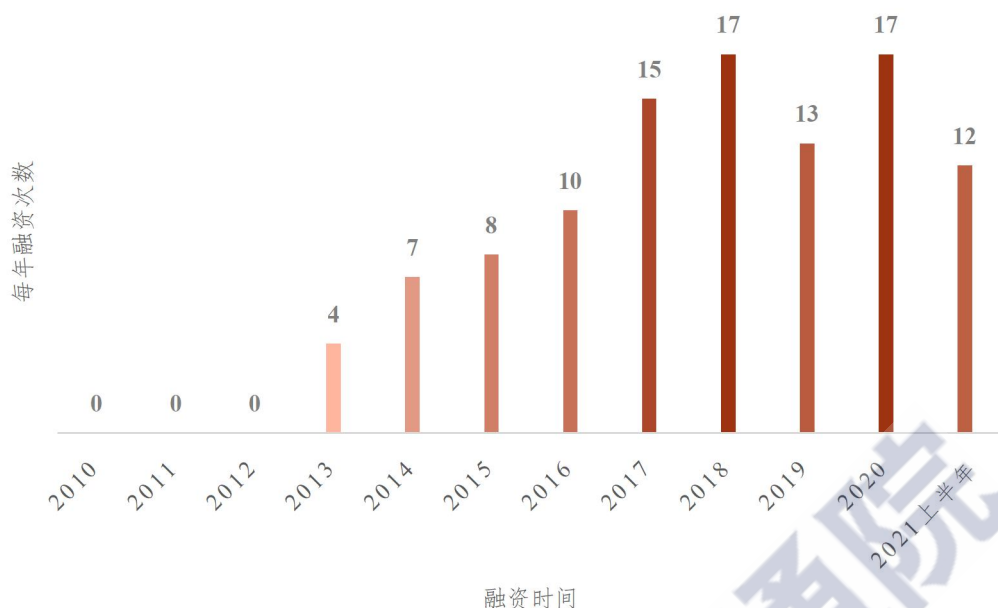
当前数据库人才培养渠道主要有三个：高校教育、培训机构和企业。各渠道分别具有不同的培训方式和培训目标。

**高校教育注重普适教育，重视社会人才发展大趋势需求。**通过原理性知识传授、数据库系统应用实践等教学方式，为数据库产业发展提供了大量储备人才。

**培训机构是数据库人才认证获取的主要途径。**培训机构累计为合作伙伴培训学员超 5 万人次，其中获得 Oracle、MySQL、PostgreSQL 认证学员数千人，为企业输送专业 DBA 万余人。培训方式为厂商授权培训中心或联合认证培训中心等，培训知识主要面向数据库工程实践和应用。除原厂培训，当前专业培训机构包括恩墨学院、新炬学院、盘古云课堂等。

**企业培训基于人才培养时间成本、人才可用性等考量因素，多渠道聚集人才。**一方面，数据库厂商开始建立自己的认证体系，并形成了不同级别的培训课程和认证考试；另一方面，一些企业开始加强与院校的产教融合尝试，通过与高校进行教材编撰、实训开发、专业共建、人才共建等合作项目，在高校提前培养数据库相关储备人才。

### 4. 数据库领域受资本市场高度追捧



来源：公开资料整理

图 21 我国数据库初创企业融资情况

中国信通院统计分析，自 2013 年至今，数据库企业累计完成约 42 次融资，根据披露金额显示，融资额度总计约为 78.6 亿元。自 2014 年成立的以数据库产品供应与服务提供为主营业务的企业为 29 个，其中 24 个企业先后获得单笔数百万元至最高 2.7 亿美元融资，仅 2021 年一、二季度期间，获得最新一轮融资的企业数就达 12 家，2020 全年获得融资的企业数量为 17 家，占比 59%，其中不乏高瓴创投、经纬中国、红点创投、红杉资本等知名投资方。由此看出，近些年随着国产数据库概念的火热与应用需求多样化带动的技术变革，国内外各路资本纷纷注入数据库产业，形成“百舸争流”的旺盛态势。

#### 四、数据库典型行业应用动态

金融、电信、政务、制造、互联网五个行业为数据库产品及服务采购份额前五的行业，采购总和占据全部市场份额的 80% 以上。所以本章重点梳理上述五个典型行业应用数据库的现状，并通过分析其业



务特点，研判各典型行业应用数据库的趋势。

首先通过 IT 监管环境、数据业务复杂性、核心业务数据类型、成本敏感性、科技能力储备情况五个维度对五个典型行业进行分析，对比结果如下表所示：

表 1 数据库典型行业应用特点

	金融	电信	政务	制造	互联网
IT 监管环境	强	强	强	一般	弱
数据业务复杂性	强	强	一般	一般	强
核心数据业务特点	强事务	强事务	分析实体 关联多	基于时间 变量分析	追求高速 处理
成本敏感性	弱	弱	一般	一般	强
科技能力储备情况	强	强	弱	弱	强

来源：中国信息通信研究院，2021 年 6 月

通过分析，金融行业和电信行业在五个维度方面表现相同，所以其数据库应用现状和未来发展走势类似，所以合并为一个小节进行论述。

### （一）金融行业&电信行业

传统金融机构（银行、证券、保险）和电信运营商作为支撑国民经济正常运行的关键行业，生产总值总和共占据我国 GDP 的 11%。两个行业的重要数据库系统由于支撑大量涉账业务，业务正确性和连续性关系国计民生，在强监管压力下，对数据一致性要求极高，所以主要应用以关系型数据库为主。据中国信通院统计分析，以业务系统数量为计数单位，我国金融行业各类数据库占比为 Oracle 55%、DB2 19%、MySQL 13%、PostgreSQL 6%，其他 7%。

目前金融和电信行业在数据库应用方面呈现五个新需求：**一是**随着移动互联网的发展，数据量呈现急剧增长，这对数据存储和管理提出了更高要求；**二是**随着普惠金融落地、移动电话普及率达到 114 部/百人，其对业务连续性能力更加重视；**三是**随着电子支付的大面积普及，需要面临高并发业务和高用户量带来的系统压力；**四是**面临互联网金融对传统金融业务的冲击以及提速降费要求，业务创新要求进一步强烈；**五是**防止潜在供应链风险，技术层面存在国产化需求。根据上述需求分析，金融和电信行业在数据库应用方面正在呈现三大趋势：

**一是大部分存量数据库将向分布式架构升级。**金融和电信行业核心业务数据类型均为关系型数据库，在当前信息科技背景下，此类集中式数据库所面临的业务体量将呈现爆发性增长。同时由于单一硬件能力增长有限，所以无法依靠纵向升级硬件扩展存量数据库能力上限。而与集中式事务数据库相比，分布式数据库采用多种模式实现数据的分散存储，将压力分散到不同服务器上，并不断通过增加存储或计算节点来实现弹性升级，克服了集中式数据库的诸多缺点，典型案例如下：

表 2 金融和电信行业数据库架构升级典型案例

案例主体	案例描述	案例时间
中国农业银行	南大通用分布式分析型数据库 GBase 8a 助力农行完成首家 MPP 国产数据库替换的银行	2013 年
北京银行	平凯星辰分布式数据库 TiDB 替换传统集中式数据库，上线核心系统	2018 年

浙江移动	华为分布式数据库 GaussDB 作为业务大厅系统的数据库组件	2019 年
四川电信	云和恩墨助力完成 CRM 与账务两个核心系统上云，数据库替换为 TeleDB	2021 年

来源：公开资料整理

**二是应用大量非关系型数据库助力创新业务落地。**金融机构和电信运营商拥有庞大客户群体的各类数据，覆盖客户、账户、产品、交易等大量的结构化数据，以及海量的语音、图像、视频等非结构化数据。这些数据背后都蕴藏了诸如客户偏好、社会关系、消费习惯等丰富全面的信息资源，能够有效赋能其自身业务发展。而这些数据的价值挖掘，均需应用到图数据库、文档数据库等非关系型数据库，所以未来将会有大量非关系型数据库助力金融和电信创新业务落地，典型案例如下：

表 3 金融行业非关系型数据库替换典型案例

机构主体	案例描述	案例时间
民生银行	巨杉 SequoiaDB 数据库用于分布式影像平台	2019 年
华瑞银行	阿里 GDB 图数据库辅助智能风险管理	2021 年

来源：公开资料整理

**三是产品选型逐渐倾向国产数据库供应商。**统计结果显示金融和电信行业数据库产品类型以国外商用产品和国外社区开源产品为主，但当前受国际单边保护主义影响，国外商用数据库产品存在潜在供应链风险。面对强监管的要求，金融和电信行业对于风险的容忍度极低，所以各机构在选型方面均开始倾向国产数据库。同时目前多项案例、测评数据均显示国产数据库产品功能、性能水平不断增强，产业供给能力整体进入上升周期，增强了金融和电信机构选型国产数据库产品

信心。

表 4 金融行业核心系统数据库替换典型案例

机构主体	案例描述	案例时间
中信银行	中兴通讯分布式数据库 GoldenDB 替换传统集中式数据库，上线核心系统	2019 年
湖北银行	达梦数据库 DM7 承载新核心系统	2019 年
张家港银行	腾讯云 TDSQL 数据库上线核心系统	2020 年
邮储银行	开源数据库 openGauss 上线新一代核心系统	2021 年

来源：公开资料整理

## （二）政务行业

《第 15 届国际数字政府评估排名报告》显示<sup>21</sup>，中国处于第 37 位，仍处于较落后的水平。数字政府能力的提升，需依赖强大的现代化的智能治理基础设施，而数据库作为支撑数据存储和计算的核心组件，是智能治理基础设施的重要组成部分。

当前我国在提升社会治理的数字化治理水平过程中，主要呈现两大特点：一是个体、企业和社会等被治理对象数量庞大、日趋复杂，而当前我国智能治理基础设施仍以传统关系型数据库为主，效率较低，亟需变革更新；二是智能治理要求各层、各机构政府人员深度应用信息科技工具，而当前信息科技工具应用普遍需要较高门槛，政务行业科技能力储备情况较普遍较低，导致数据基础设施建设完成之后，应用效果不佳，难以达到预期效果。

当前政务行业在数据库应用方面正在呈现两大趋势：

**一是大范围应用空间型、关联型数据库等产品。**社会治理普遍基

<sup>21</sup>日本早稻田大学数字政府研究中心于 2020 年 9 月发布。

于空间位置和人、物、企业等实体的关联关系进行，而传统关系型数据库在面对空间位置和实体关联关系的分析和建模能力不足。当前空间型数据库和图数据库的快速发展提高了空间位置数据和实体关联关系分析的存储和分析效率，将是未来智能治理中数据库的重要应用方向，典型案例如下：

表 5 政务行业非关系型数据库典型案例

案例主体	案例描述	案例时间
自然资源部信息中心	国家基础地理信息中心承担建立的三维立体时空数据库系统接入自然资源部信息中心网络	2020 年
杭州市公安局	中奥科技的知识图谱产品助力智能警务发展	2019 年

来源：公开资料整理

二是利用各类工具组件，做到数据库应用“平民化”。为达到政务人员深度应用信息科技工具目的，一方面在加强政务人员信息化能力同时，数据库应用过程也需尽量做到“平民化”，包括但不限于具备可视化建模能力、支持特定政务领域解决方案模板、流程式指引等。

### （三）制造业

工业场景中，80%以上的监测数据都是实时数据，且都是带有时间戳并按顺序产生的数据，这些来源于传感器或监控系统的数据被实时地采集并反馈出系统或作业的状态。

随着政策环境的铺垫和工业互联网基础设施的逐步完善，工业大数据迎来重大发展机遇。当前物联网技术正逐步渗透工业领域，不断增长的传感器、飙升的数据量以及更高的大数据分析需求对原有的数据库系统提出了新的需求，包括增加扩展性、实现与大数据技术生态



的友好对接、降低大规模应用价格、充分利用边缘计算设备能力。

未来工业行业在数据库应用方面将呈现两大趋势：

**一是应用大量时序数据库。**随着物联网场景的丰富以及人们对信息全面掌控的需求，基于时间序列数据类型的应用越来越多，工业业务对于查询的要求已经不仅仅是满足于一些基础的条件查询或是插值查询，查询会大量扩展到基于时间的维度，而时序数据库更契合此类场景，典型案例如下：

表 6 制造业时序数据库典型案例

案例主体	案例描述	案例时间
3M 公司	使用微软 Azure SQL Edge 数据库打造智慧工场 4.0	2020 年
三一重工	腾讯时序数据作为设备数据分析引擎	2020 年
睿信世达	涛思数据 TDengine 进行睿信物联网平台改造	2021 年

来源：公开资料整理

**二是逐步向边缘计算发展。**工业领域是 IoT 的重要试验田。当数据过于庞大，集中化的处理方式就很难响应实时的数据分析需求，这就带来了数据计算向边缘的发展，需要实时响应的监控就通过边缘设备及时的处理并反馈，需要用于大规模分析的数据再进行集中存储，这种分级的处理方式能够有效地提升时效性数据的价值，同时减轻存储系统的负担，当前许多工业数据库正在研发边缘计算版本。

#### (四) 互联网

2020 年，我国规模以上互联网和相关服务企业（简称互联网企业）完成业务收入 12838 亿元，其中电子商务、社交、游戏、音视频、

搜索引擎五类业务合集收入占总收入的比重达 85%。这五类业务均需要极致快速的用户体验作为竞争优势基础，而极致快速的用户体验离不开底层 IT 系统中数据库的科学建设与持续优化。

当前我国互联网行业发展进入调整期，上市企业营收政府大幅放缓，总市值持续波动，投融资低位徘徊，行业格局出现分化。互联网行业为充分发挥新要素、新模式、新动能等方面的优势，对底层 IT 系统中的数据库提出了多项新要求，以性能好、造价低、迭代快为核心。

未来互联网行业在数据库应用方面将呈现三大趋势：

**一是利用内存数据库加速业务效率。**近年来，随着动态随机存储器（DRAM）容量的上升和单位价格的下降，使大量数据在内存中的存储和处理成为可能，Redis、Memcached 等内存数据库逐渐成熟，应用范围越来越广。未来几年，随着非易失性存储器件（NVM）逐步投入商用，新硬件将会给内存数据库带来更大的发展机遇。

**二是开源数据库应用更加广泛。**开源数据库具有获取方便、成本低等优势，同时互联网企业科技能力强、对于成本敏感等特点，所以开源数据库产品在互联网行业已被广泛应用。当前越来越多的数据库企业将开源作为一种商业运营策略，所以开源数据库市场将更加丰富，为其进一步在互联网领域广泛应用提供了条件。

**三是初创公司利用云数据库促进其快速发展。**互联网创业公司一般具备人数较少、资金有限、场地不足等特点，无法投入大量资源用于建设 IT 基础设施。当前各大云计算公司推出众多云数据库产品，

由于利用了云资源的优势，云数据库具备弹性好、计费模式便捷、套件生态好等特点，符合互联网创业公司的应用需求。互联网创业公司未来将更大面积应用云数据库。

## 五、总结与展望

数据库作为提供数据存储与处理能力的软件产品，是各机构信息系统的关键部件，是助力数据价值释放的核心引擎。随着数据跃升为生产要素，数据重要性进一步提高，我国数据库产业也迎来新一轮变局。

**从产业角度看**，宏观政策利好推动了存量数据库市场上行，我国数据库产业进入蓬勃发展的初期，产品供应商、服务提供商、支撑产业从业者均积极行动，各自发挥技术、渠道、运营等优势，寻求对于自身最优的发展路径；另一方面云基础设施的发展成熟将接近一半的传统数据库市场转移到了线上，云计算企业利用既有基础设施优势，成为最大获益者。

**从技术角度看**，一方面数据应用的变化倒逼数据库具备更大数据存储容量、更多数据计算模型、更快数据业务响应能力，整体技术发展进入后关系型阶段，架构设计逐渐分布式化、模型构建逐渐场景化；另一方面，人工智能、新型硬件、区块链、密态计算等关联技术的创新正在催生新型数据库设计模式，传统数据库功能边界正被逐渐突破。

在数据库产业和技术的变局过程中，供给侧、应用侧、生态侧均处于机遇与挑战并存的状态。未来一段时期，我国数据库行业将围绕两个核心命题进行发展：

一个是缩小“高要求的存量数据应用需求”与“仍处于发展初期阶段的供给能力”之间的差距。在存量数据应用需求方面，国际数据库产品已经为我国数据库行业树立了发展的目标与方向，相比之下，我国数据库产业供给能力仍处于发展初期，与应用方既有的使用预期仍存在差距。预计我国数据库行业未来围绕“技术”和“生态”两条主线，一方面利用资金、开源、人才、场景等资源完成技术成熟度达标，另一方面补足运营、服务、渠道短板，完成生态成熟度达标。

另一个是探索“创新型数据应用需求”与“数据库技术产品演进路线”的合理映射关系。创新型数据应用需求是更长周期内数据库产业发展的牵引力，但是数据库技术产品如何演进、从而满足创新型数据应用需求并不存在成熟的路线图，所以未来数据库技术产品将在一段时间内呈现较强的随机演进状态。为解决这一核心命题，数据应用方与数据库供给方必须利用产业平台、人才互通等手段，完成深度融合与绑定。

“明者见于无形，智者虑于未萌。”当前我国数据库产业的发展格局，是紧跟时代步伐顺应历史规律、着眼全球提升国际综合竞争力、立足国情推动新旧动能接续转换的外在表现。我们相信，以数据库为代表的新型数据基础设施不断创新发展，对于全面建设社会主义现代化国家的征程，将起到重要的推动作用。未来已来，拭目以待。

## 六、附录

### （一）数据库管理系统细分类别

表 7 按数据模型分类的数据库管理系统

类别	定义	适用场景
关系型数据库管理系统 (Relational DBMS)	支持关系型的表数据（行与列）管理方式的数据库。通常访问方式为 SQL 语言。	适用于以行列方式进行数据组装的关系型数据应用。其关注事务(交易)ACID 特性，且拥有复杂 SQL 查询诉求。
文档数据库 (Document DBMS)	以“文档”存放数据的没有特定数据模型的存储模型。其往往采用带有具体含义的内定符号进行数据单元切割。基于文档存储的数据可以直接通过上层应用进行操作。	适用于对关联查询、事务要求不高且数据模型不明确的场景。例如会话保存、购物车等。
时序型数据库管理系统 (Time Series DBMS)	专门针对时序型数据（一种任何操作都必须带有时间戳的数据类型）进行优化的数据库。	适用于更关注时间点的工业及物联网场景，例如各类传感器、仪表等无需对数据变更的数据采集场景。
图形数据库管理系统 (Graph DBMS)	数据以图结构存放的数据库。其中图数据结构指的是像“节点”“边”（节点之间的关系）的形式。往往用于计算“图”的特定属性，例如一个节点到另一个的跳数。	适用于类似社交网络、金融风控、推荐系统等专注于构建关系图谱的业务场景。
空间型数据库管理系统 (Spatial DBMS)	用于例如“点”“线”“面”这样的典型空间数据类型的存取与运算的专用数据库。此类数据库往往支持专门用于加速空间数据操纵的空间索引。	适用于空间地理数据操作场景，如卫星遥感、自然资源等。
混合型数据库 (Hybrid Transaction and Analytical Process DBMS)	兼容存储和计算模型，可以同时存储和处理关系型和非关系型数据的数据库。	适用于对异构数据互通共享有需求的业务场景，以及同时有事务类和分析类需求的业务系统。

来源：中国信息通信研究院根据公开资料整理



## （二）数据库企业人才发展计划

表 8 部分数据库企业人才培训一览表

时间	企业	人才培养举措
2008 年	人大金仓	成立金仓学院
2014 年	达梦	成立达梦大学
2015 年	阿里云	联合 8 所高校，成立 AUCP
2016 年	南大通用	成立 Gbase 学院
2018 年	星环科技	成立星环大学，联合大数据分析与应用技术国家工程实验室共建星瀚大数据联合实验室
2019 年	华为	华为和教育部夯实“智能基座”项目，覆盖 72 所高校
2019 年	腾讯云	发布“诺亚计划”，招募 50 个重点合作伙伴，提供超 100 场定向培训
2019 年	巨杉数据库	与太原理工大学成立联合实验室
2020 年	蚂蚁金服	与华东师大共建“分布式数据库联合实验室”

来源：中国信息通信研究院根据公开资料整理

## （三）中国信通院数据库应用创新实验室

2019 年 10 月 31 日，中国信息通信研究院数据库应用创新实验室在 2019（第二届）金融科技产业峰会正式揭牌。当前已有来自数据库企业、金融机构、数据库服务商、系统集成商、硬件服务商共 50 余家企业加入。数据库应用创新实验室已经围绕金融领域开展数据库产品的功能测试、性能测试和第三方验证测试，助推数据库产业应用创新。

## （四）中国信通院数据库评测体系

中国信息通信研究院作为国家高端专业智库，一直深耕数据库领域测试工作。自 2015 年起，针对数据库基础软件领域，依托中国通

信标准化协会大数据技术标准推进委员会(CCSA TC601)，已经开展了分布式分析型数据库、分布式事务型数据库、时序数据库、关系云数据库、图数据库、键值型内存数据库和文档数据库七个方向的工作，发布行业报告十余本，牵头金融、工业、电信等行业多项标准。经过十二批测试，已累计完成国内 46 家数据库企业 73 款产品的 103 次评测工作，测试项目包括功能测试、性能测试和稳定性测试等。经过 6 年发展，该评测体系见证了国内数据库产品由弱变强、产品生态逐渐丰富的过程，圈定了国内数据库产品厂商第一梯队，成为了数据库产业发展的风向标。

## 参考文献

- [1] TM Connolly, CE Begg ,Database Systems: A Practical Approach to Design, Implementation and Management 4th Edition (England: Addison Wesley, 2005).
- [2] 吴鹤龄. 关系数据库的标准语言——SQL[J]. 计算机研究与发展, 1989(06).
- [3] 金融分布式事务数据库白皮书. 中国信息通信研究院.
- [4] 关系型云数据库白皮书. 中国信息通信研究院.
- [5] 图数据库白皮书. 中国信息通信研究院.

CAICT 中国信通院

中国信息通信研究院 云计算与大数据研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：13691032906

传真：010-62304980

网址：[www.caict.ac.cn](http://www.caict.ac.cn)

