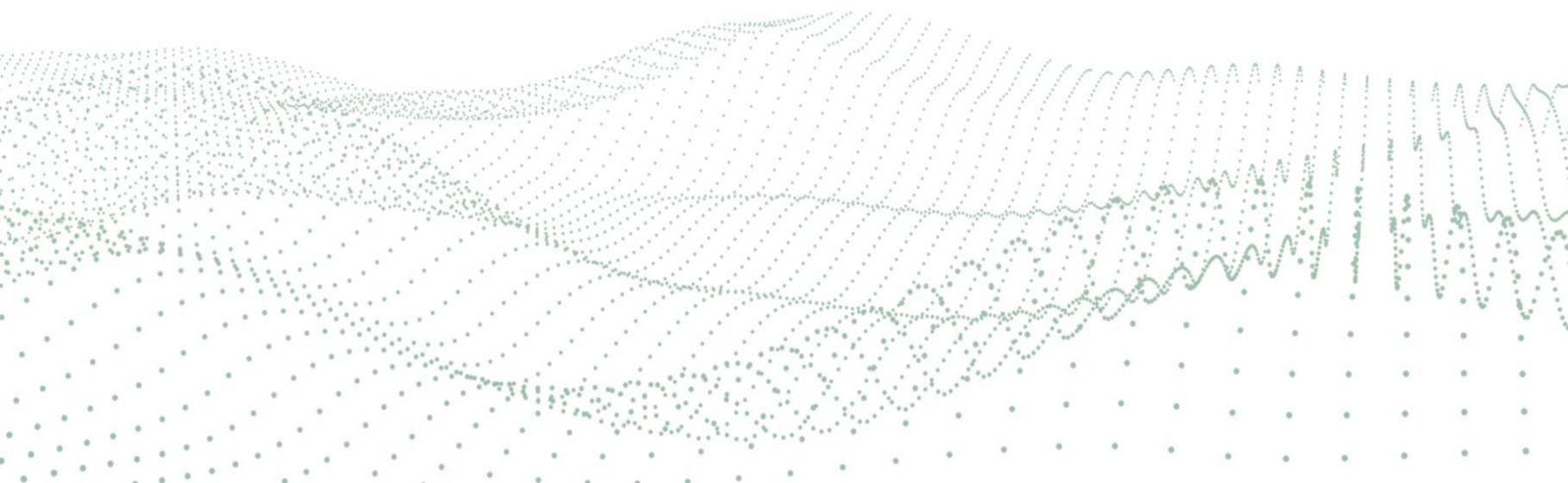


# 数据库发展研究报告

## (2025 年)

CCSA TC601 大数据技术标准推进委员会

2025年7月



## 编写委员会

本报告的撰写得到了数据库领域多家企业与专家的支持和帮助，主要参与单位与人员如下。

### ❖ 主要编写单位（排名不分先后）：

中国信息通信研究院、中国移动集团数智化部、金篆信科有限责任公司、中电科金仓(北京)科技股份有限公司、天津南大通用数据技术股份有限公司、甲骨文(中国)软件系统有限公司、上海运熹科技有限公司、中兴通讯股份有限公司、浪潮云信息技术股份公司、北京中亦安图科技股份有限公司、腾讯云计算（北京）有限责任公司、深圳计算科学研究院、深圳九有数据库有限公司、北京奥星贝斯科技有限公司、联想（北京）有限公司、天翼云科技有限公司、西安同星恒耀信息技术有限公司、成都虚谷伟业科技有限公司、北京新数科技有限公司、北京自然原数科技有限公司、蚂蚁科技集团股份有限公司、上海市数字证书认证中心有限公司、平凯星辰（北京）科技有限公司、北京海量数据技术股份有限公司、中国民航信息网络股份有限公司、中国第一汽车股份有限公司、新华三技术有限公司、北京酷克数据科技有限公司、浙江创邻科技有限公司、中移（苏州）软件技术有限公司、北京万里开源软件有限公司、杭州质变科技有限公司、上海天玑科技股份有限公司、天翼数字生活科技有限公司

❖ 编写组主要成员（排名不分先后）：

刘思源、齐丹阳、刘蔚、杨佳星、姜春宇、马鹏玮、康宸、王小玉、赵东明、吕伟初、朱业、秦延涛、左奇、董文、刘俊、胡一鸣、赵伟、冯文忠、白雪、温婧、李珈、杨鲁、杨元同、陈磊、赵衍衍、冷友方、孙路明、屠要峰、陈河堆、解海波、尚长军、邓光超、王阳、刘刚、黄远邦、郑成龙、高清华、徐晓敏、张展、郭文、欧伟杰、孟凡彬、李伟超、毛雨爽、蒋昀岂、何威、刘浔、全锦琪、李韬、苗旭成、向祥兵、王栩、王枫、李文娇、赵世凯、杨丹、钟华龙、郭立、张心怡、刘启荣、高新刚、张立、曹国顺、明玉琢、马攀、黄元霞、刘永利、朱浩、江晶、郑宇隆、刘孟占、何毓锬、刘小兵、戚仕鹏、林恒、郭智慧、杨晶、曹金龙、宋可欣、白玥、秦宁慰、田丰、刘中一、阎松柏、李莉、董骐、渠谨黛、郭宏伟、霍然、游根节、张德民、王子睿、李永平、张战防、徐田原、唐昕、周研、张晨、李欢、吴晓峰、陶捷、刘永德、张卓、徐爽、魏闯先、陆元飞、肖煜伟、朱伟、孙洪标、俞好龙、吴国新、朱奕飞、孙成、成思敏

# 前 言

人工智能已成为新一轮科技革命和产业变革的重要驱动力，数字经济高速发展为全球经济增长注入新动能。数字经济与人工智能日益融入到经济社会发展各领域全过程，也为数据库产业发展带来诸多机遇和挑战。以 AI 原生数据库、云原生数据库及数据库智能体等为代表的数据库产品及周边生态驱动多模态数据融合处理加速，推动向量检索、智能调优等能力跃升，助力企业应对智能时代业务变化和创新需求，加速企业数字化转型，推动产业创新发展。

2025 年，全球数据库产业、技术、应用呈现如下总体发展态势。

产业方面，全球数据库市场竞争激烈，形成多强格局，公有云数据库市场成为主导力量；我国数据库市场趋于理性，供给侧从“烧钱”向“赚钱”转型，产业侧从“数量型”向“质量型”转变。全球范围内，数据库市场规模持续增长，约为 1154 亿美元，企业数量 400 家，产品数量 535 款。我国数据库市场规模为 83.7 亿美元，占全球 7.3%，云数据库市场规模进一步扩大至 384.14 亿元，数据库供应商数量为 103 家，产品数量为 164 款。

技术方面，随着人工智能加速发展，数据库技术架构持续革新，迈入 AI 原生时代，呈现出十个细分发展方向。分别是交易分析一体化消除业务壁垒、存算分离架构提升资源利用率、云原生数据库实现高效资源管理、机器学习优化数据库应用效能、Text to SQL/Text to GQL 降低数据库交互门槛、RAG/Graph RAG 提升数据库查询准确性、向量数据库提升大模型检索效率、多模数据库赋能大模型多样化数据、

AI 原生数据库解锁场景新可能以及数据库智能体提供数据交互新模式。

应用方面，国产数据库应用规模不断扩大，用户体验不断提升。数据密集型行业引领数据库应用创新，传统行业数字化转型不断深入。金融、电信等数据密集型行业国产数据库应用规模不断扩大，形成成熟的国产数据库应用体系，引领行业发展。图数据库、多模数据库等新型数据库有力支持能源、交通运输业等传统行业数字化转型。

本报告是中国通信标准化协会大数据技术标准推进委员会（CCSA TC601）继《数据库发展研究报告（2021年）》、《数据库发展研究报告（2022年）》、《数据库发展研究报告（2023年）》、《数据库发展研究报告（2024年）》发布后的第五本数据库年度综合报告，内容涵盖数据库产业及市场、数据库产品及服务、数据库支撑体系、数据库技术发展趋势和典型行业数据库应用情况综述。由于水平所限，错误和不足之处在所难免，欢迎各位读者批评指正，联系 [qidanyang@caict.ac.cn](mailto:qidanyang@caict.ac.cn)。

# 目 录

一、 数据库产业发展情况综述 .....	1
(一) 数据库产业及市场 .....	1
1. 全球及中国数据库市场 .....	1
(二) 数据库产品及服务 .....	3
1. 全球数据库市场形成多强格局 .....	3
2. 全球数据库进入高质量发展期 .....	4
3. 全球数据库从业人员保持稳定 .....	6
4. 国内外产品类型数量分布各有侧重 .....	7
5. 非关系型数据库呈现百花齐放态势，向量数据库热度持续攀升 .....	9
6. 国外商用数据库占比略有上升，我国以商用为主 .....	11
7. 我国数据库市场规模持续扩大，市场格局逐步显现 .....	12
8. 重点行业头部用户运维投入逐年增长，研发投入有所回落 .....	13
(三) 数据库支撑体系 .....	16
1. 创新方面，多模融合成为研究重点，我国科研实力稳步增强 .....	16
2. 投融资方面，PG生态获得青睐，多云管理和AI成为融资亮点 .....	20
3. 标准方面，我国数据库标准体系日益完善，引领数据库产业高质量发展 .....	21
二、 数据库关键技术发展趋势 .....	24
(一) 技术架构持续革新 .....	24
1. 从分而治之到交易分析一体化 .....	24
2. 从本地部署到云上运行 .....	26
(二) 数据库迈入AI原生时代 .....	30
1. 人工智能与数据库双向赋能 .....	30
2. AI原生数据库解锁场景新可能 .....	42
3. 数据库智能体提供数据交互新模式 .....	46
三、 数据库行业应用情况综述 .....	50
1. 国产数据库在电信行业得到广泛应用 .....	50
2. 数据库技术驱动智能金融核心变革 .....	52
3. 数据库技术支撑能源行业数字化转型 .....	53
4. 数据库技术助力交通运输行业平安出行 .....	54
四、 总结与展望 .....	56

# 图 目 录

图 1	2024-2027 年中国数据库市场规模及增速 .....	1
图 2	2023-2025 中国公有云和本地部署数据库市场规模 .....	2
图 3	全球及我国数据库企业数量变化趋势 .....	3
图 4	全球数据库企业分布 .....	4
图 5	全球数据库企业开展业务时间 .....	4
图 6	我国数据库企业开展业务时间 .....	5
图 7	全球数据库企业人员数量分布 .....	6
图 8	我国数据库企业人员数量分布 .....	6
图 9	全球数据库产品类型分布 .....	7
图 10	我国数据库产品类型分布 .....	8
图 11	过去四年全球非关系型数据库数量占比排名 .....	9
图 12	过去四年我国非关系型数据库数量占比排名 .....	10
图 13	全球数据库产品商用开源对比 .....	11
图 14	全球活跃开源数据库开源时间 .....	11
图 15	我国活跃开源数据库的开源时间 .....	12
图 16	我国数据库产品数量 .....	13
图 17	2022-2024 年重点行业头部用户研发/运维人员数量和投入资金变化 ....	14
图 18	2024 年研发人员数量及投入资金区间分布 .....	14
图 19	2024 年运维的人员数量/投入资金区间分布 .....	15
图 20	2024 年 VLDB、ICDE 和 SIGMOD 论文分布情况 .....	16
图 21	2024 年 VLDB、ICDE 和 SIGMOD 论文关键词云图 .....	17
图 22	2022-2024 年中国高校及企业学术会议论文贡献情况 .....	18
图 23	2024 年中国高校三大会议论文数量前 20 名 .....	19
图 24	2024 年中国企业三大会议论文数量前 10 名 .....	20
图 25	CCSA TC601 数据库领域标准化工作体系 .....	22
图 26	GART 系统的总体架构 .....	25
图 27	Db2 LSM 架构图 .....	28
图 28	AzureSQL 数据库架构概览 .....	28
图 29	ProRP 架构示意图 .....	29
图 30	Text-to-SQL 发展历程 .....	33
图 31	NL2SQL 方法概览 .....	34
图 32	RAG 三种范式对比 .....	36

图 33	GraphRAG 框架在问答任务中的概览 .....	37
图 34	RAG 在向量数据库中的框架流程 .....	40
图 35	使用向量数据库的 GPT 的语义缓存概览 .....	40
图 36	统一抽象层表达异构数据结构示意图 .....	42
图 37	AI 原生数据库的架构 .....	43
图 38	AI 组织型数据库流程图 .....	45
图 39	大语言模型 (LLMs) 与图相互增强的总体框架 .....	48
图 40	DB-GPT 系统设计 .....	48
图 41	2024-2025 年电信行业上线数据库情况统计 .....	50
图 42	2025 年上半年电信行业部分国产数据库中标情况 .....	51
图 43	2025 年 1-5 月金融行业国产数据库部分中标/投产情况 .....	52

# 表 目 录

表 1	ML4DB 研究中查询计划表示方法总结 .....	31
表 2	基于预训练模型和大语言模型的 NL2SQL 方法分类 .....	35
表 3	支持向量数据的主流各数据库对比 .....	39
表 4	AI 原生数据库的五个阶段 .....	44
表 5	DB-GPT 与其他工具的比较 .....	47

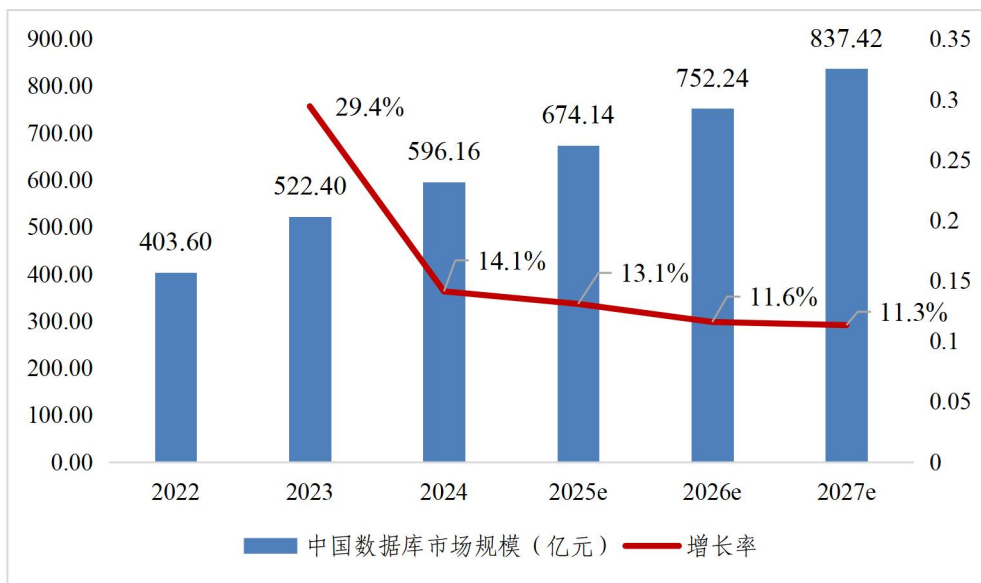
# 一、数据库产业发展情况综述

当前，全球数据库市场竞争激烈，形成多强竞争格局，公有云数据库市场成为主导力量。人工智能技术日新月异，助力数据库架构不断演进，数据库技术迈入 AI 原生时代。我国产业规模持续扩大，创新能力不断增强，市场格局逐步显现，非关系型数据库产品成为产业关注热点。我国数据库市场趋于理性，产业迈入高质量发展阶段。

## (一) 数据库产业及市场

根据中国通信标准化协会大数据技术标准推进委员会(以下简称:CCSA TC601)调研分析,我国数据库产业链包括数据库产品提供商、数据库生态工具提供商、数据库服务提供商、数据库安全供应商、数据库生态社区、数据库人才培养等多个环节,各领域参与者专攻术业,发挥竞争优势,积极拓展生态圈,为我国繁荣的数据库生态不断注入活力。

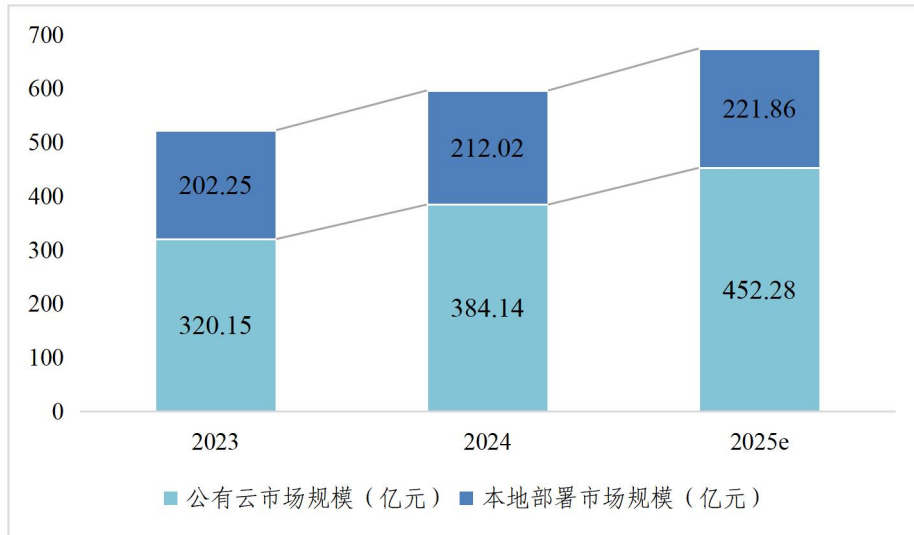
### 1. 全球及中国数据库市场



来源: CCSA TC601, 2025 年 6 月

图 1 2024-2027 年中国数据库市场规模及增速

据 CCSA TC601 测算<sup>1</sup>，2024 年全球数据库市场规模约为 1154 亿美元，其中中国数据库市场规模为 83.7 亿美元（约合 596.16 亿元人民币），占全球 7.3%<sup>2</sup>。预计到 2027 年，中国数据库市场总规模将达到 837.42 亿元，市场年复合增长率（CAGR）为 11.99%。



来源：CCSA TC601，2025 年 6 月

图 2 2023-2025 中国公有云和本地部署数据库市场规模

据 CCSA TC601 测算，按数据库部署方式划分市场规模，2024 年中国公有云数据库市场规模为 384.14 亿元，较 2023 年增速 20.0%，本地部署数据库市场规模为 212.02 亿元，较 2023 年增速 4.8%，公有云和本地部署模式市场规模分别占总市场 64.4%和 35.6%，预计 2025 年公有云市场占比将进一步扩大达到 67.1%，规模达到 452.28 亿元，本地部署模式市场增速降至 4.0%，规模为 221.86 亿元。

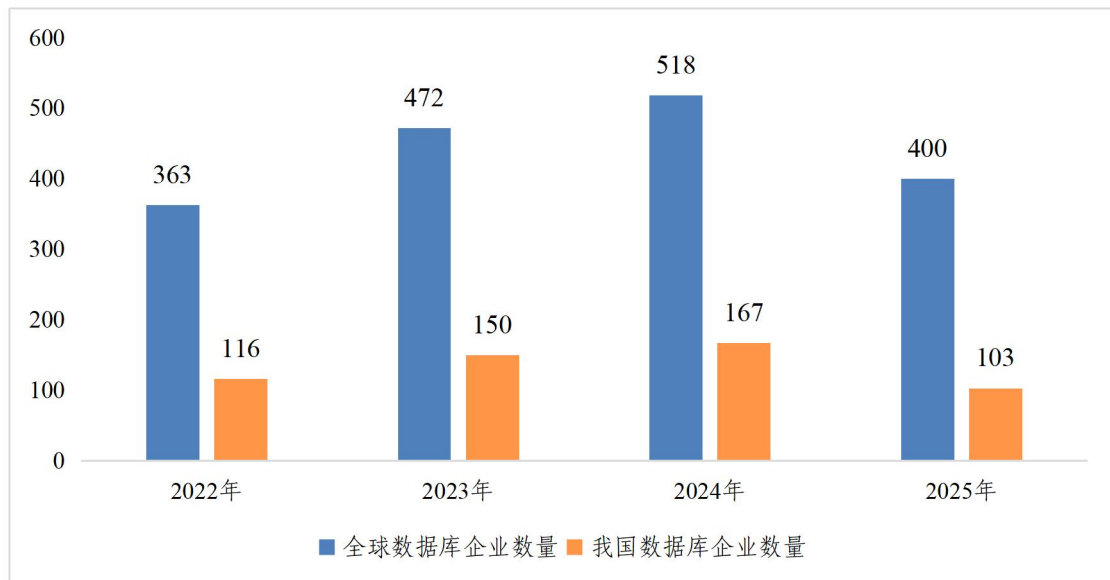
1 数据库市场规模包括数据库软件、服务和安全等周边生态。

2 《中华人民共和国 2024 年国民经济和社会发展统计公报》，国家统计局，2024 年全年人民币平均汇率为 1 美元兑 7.1217 元人民币。

## (二) 数据库产品及服务

### 1. 全球数据库市场形成多强格局

2024 年，全球企业数量有所收敛，市场形成多强格局，我国头部云厂商跻身前列。从国外看，谷歌、亚马逊、微软、甲骨文四家头部云厂商依旧维持在领导者象限的领跑地位，分析型数据库厂商 Databricks 和 Snowflake 也位于该象限；从国内看，阿里云和华为云分别入选领导者和挑战者象限，此外，OceanBase、PingCAP 和腾讯云均获“荣誉提及”<sup>3</sup>。

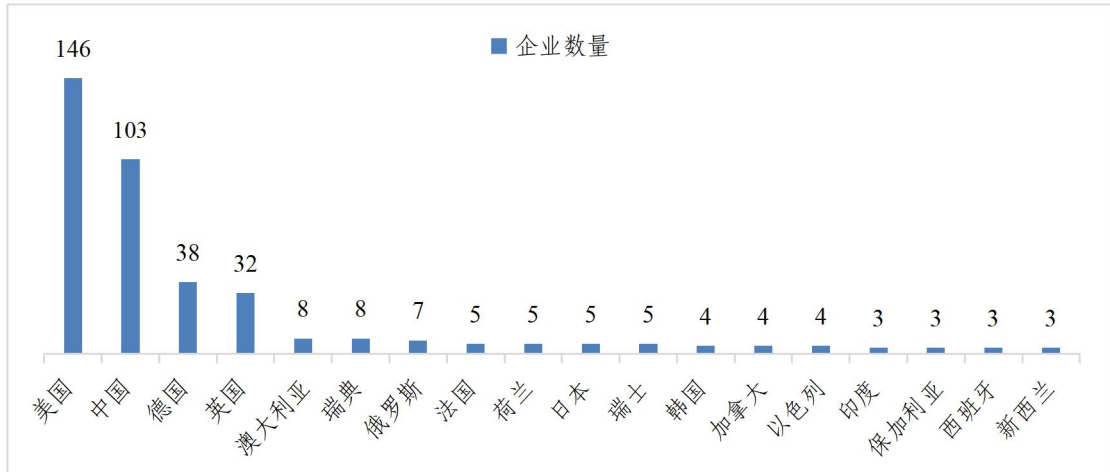


来源：CCSA TC601，2025 年 6 月

图 3 全球及我国数据库企业数量变化趋势

据 CCSA TC601 统计，2022-2024 年，全球数据库产品提供商数量逐年增加，由 363 家增加到 518 家，复合增长率约为 19.5%，我国数据库企业数量也呈现增加趋势，由 116 家增加到 167 家，复合增长率约为 20.0%；2025 年，全球及我国数据库企业数量有所收敛，当前共有 400 家企业。截至 2025 年 6 月，我国数据库企业数量为 103 家，占比 25.8%。

<sup>3</sup> 2024 年 Gartner 云数据库管理系统魔力象限。

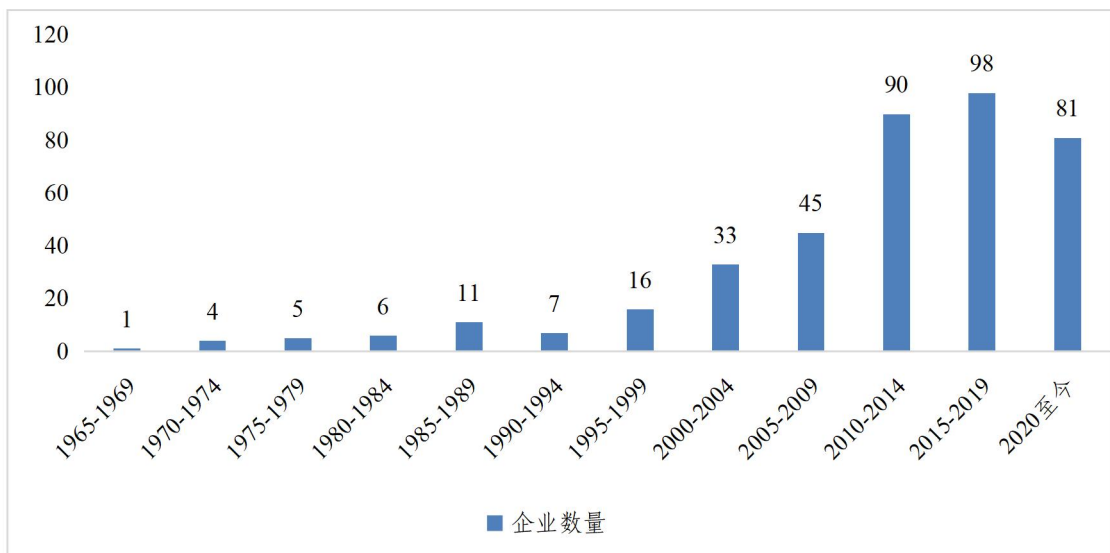


来源：CCSA TC601，2025 年 6 月

图 4 全球数据库企业分布

2014 年以来，我国数据库产业不断壮大，数据库产品提供商逐年增长，接近美国数据库产品提供商总数。近两年，由于市场竞争加剧等原因，我国数据库供给侧头部效应明显，市场格局日益清晰。此外，据 CCSA TC601 统计，截至 2025 年 6 月，德国、英国、澳大利亚、瑞典、俄罗斯等欧洲数据库厂商中仍正常运营的数量分别为 38、32、8、8 和 7 家，分别占比 9.5%、8.0%、2.0%、2.0%和 1.8%。

## 2.全球数据库进入高质量发展期

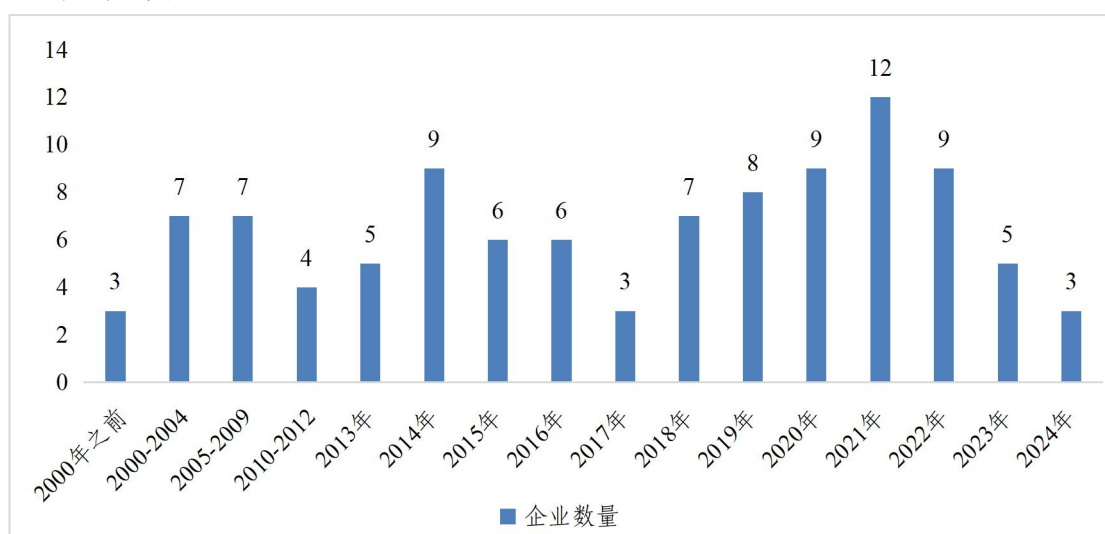


来源：CCSA TC601，2025 年 6 月

图 5 全球数据库企业开展业务时间<sup>4</sup>

<sup>4</sup> 当前正在运营的企业。

全球数据库发展经历两次热潮，21 世纪后进入蓬勃发展期，并于近两年趋于理性<sup>5</sup>。从企业开展数据库业务时间看，全球数据库企业起步于 20 世纪 60 年代，并于 80 年代迎来第一波发展热潮。21 世纪后全球数据库产业迅速发展，其中在 2010-2019 年进入发展高峰期，新兴企业不断成立。这十年间，一共出现了 239 家企业，全球 46.1% 的数据库企业成立于这一时期，其中 188 家企业发布的数据库产品持续市场运营至今，占比 78.7%。



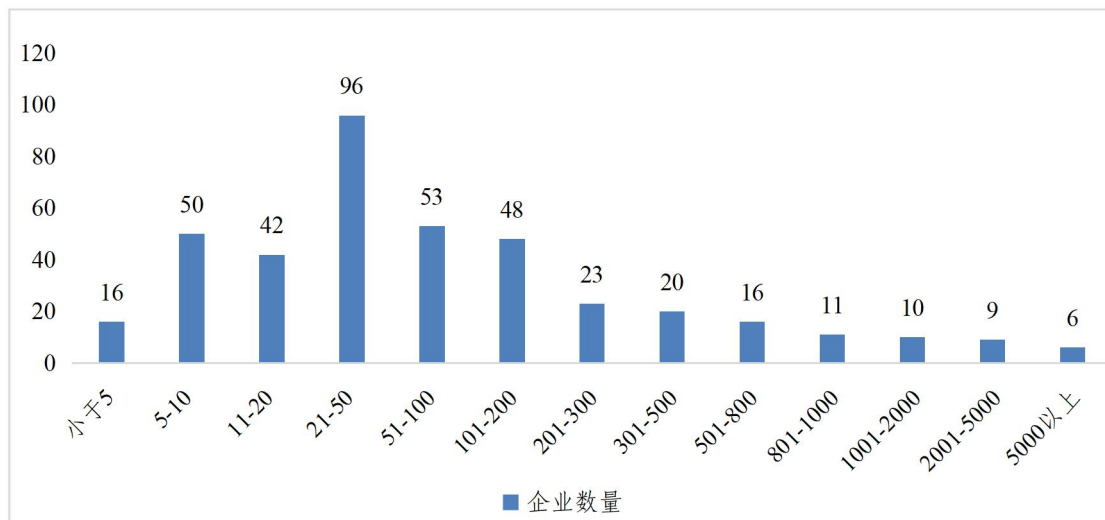
来源：CCSA TC601，2025 年 6 月

图 6 我国数据库企业开展业务时间

我国数据库产业始于 20 世纪末，并在 2014 年后迎来繁荣发展，2022 年以来，数据库供给侧日益回归理性，企业新增数量逐步放缓。据 CCSA TC601 统计，截至 2025 年 6 月，我国正常运营的数据库产品提供商共计 103 家，2024 年新增企业数量为 3 家。2022 年至今，新兴数据库企业数量逐年降低，市场头部效应明显，标志着我国数据库产业正式进入高质量转型期。

<sup>5</sup> 在《数据库发展研究报告（2024 年）》的基础上数据进行进一步修正。

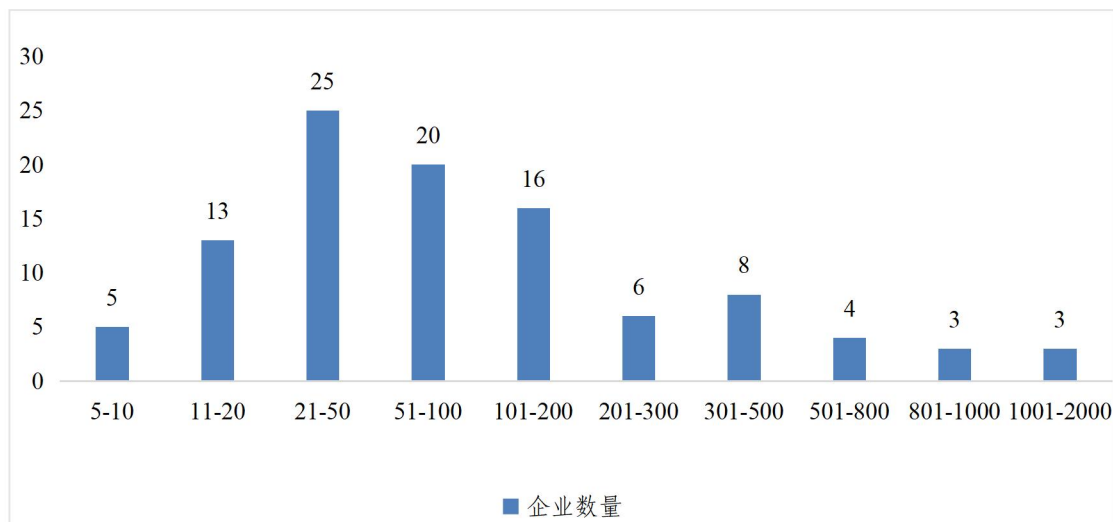
### 3.全球数据库从业人员保持稳定



来源：CCSA TC601，2025年6月

图 7 全球数据库企业人员数量分布

全球数据库企业从业技术人员已超 10 万人，接近四分之一企业技术人员位于 21-50 人数量区间。据 CCSA TC601 统计，截至 2025 年 6 月，全球数据库企业从业技术人员超 10 万人，员工数量平均不足 30 人。最高为 7000 人左右规模，最低不足 5 人左右规模。其中 21-50 人规模企业占比最高，数量 96 个，比例达到 24.0%，人数在 51-100 人左右规模次之，数量为 53 个，占比 13.3%，5-10 人位居第三，数量为 50 个，占比 12.5%。

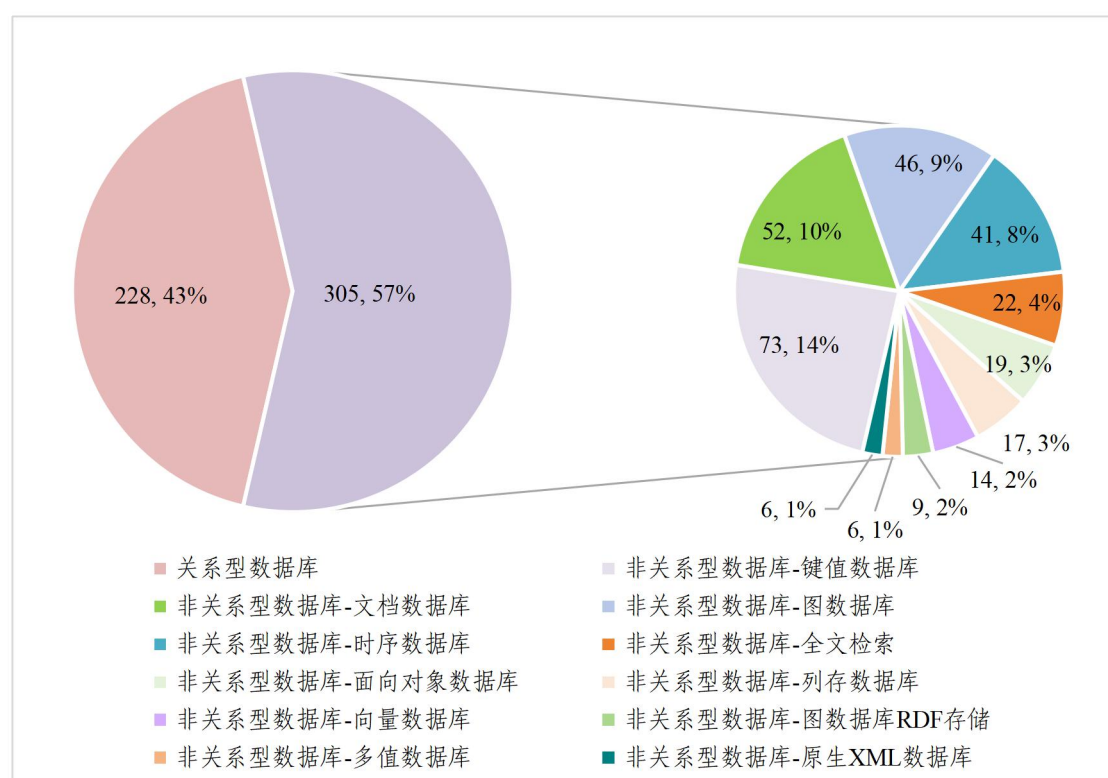


来源：CCSA TC601，2025年6月

图 8 我国数据库企业人员数量分布

我国数据库企业从业技术人员近2万人,员工数量平均约200人,我国数据库技术人才仍有较大缺口。据 CCSA TC601 统计,截至 2025 年 6 月,我国企业最高为 2000 人左右规模,最低不足 5 人左右规模。其中 21-50 人左右规模企业占比最高,数量为 25 个,比例达到 24.3%。人数在 51-100 人左右规模次之,数量为 20 个,占比 19.4%。101-200 人位居第三,数量为 16 个,占比 15.5%,前十家企业人员规模均在 500 人以上。

#### 4.国内外产品类型数量分布各有侧重

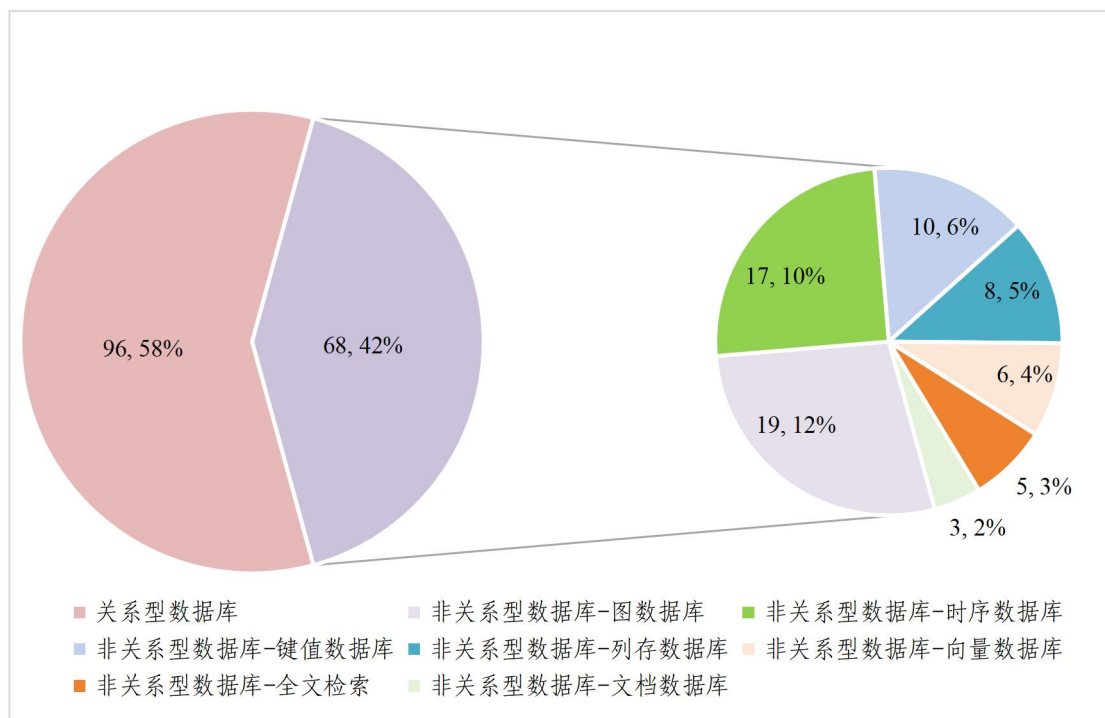


来源: CCSA TC601, 2025 年 6 月

图 9 全球数据库产品类型分布

全球数据库产品数量整体分布呈现以非关系型及混合型数据库为主,关系型为辅的局面。据 CCSA TC601 统计分析,截至 2025 年 6 月,已纳入统计的全球数据库产品共有 715 款,目前市场上共有数据库产品 535 款。除了早期的两款网状数据库和层次数据库,在剩余

的 533 个数据库产品中，关系型数据库 228 个，非关系型数据库有 305 个，占比分别为 42.8%和 57.2%。非关系型数据库中，键值型数据库 73 个、文档数据库 52 个、图数据库 46 个，在非关系数据库中依次占比 23.9%、17.0%和 15.1%。



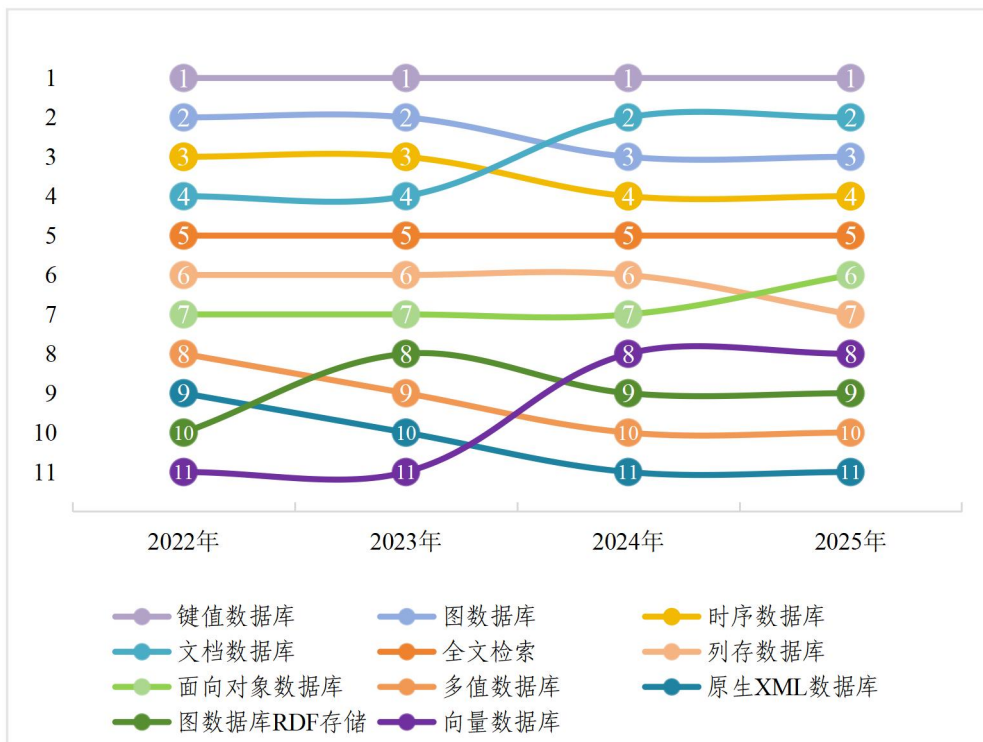
来源：CCSA TC601, 2025 年 6 月

图 10 我国数据库产品类型分布

与全球不同，我国数据库产品数量仍呈现以关系型为主，非关系型数据库为辅的局面。据 CCSA TC601 统计分析，截至 2025 年 6 月，目前我国数据库产品共有 164 款。其中，关系型数据库 96 个，非关系型数据库 68 个，占比分别为 58.5%和 41.5%。非关系型数据库中，图数据库 19 个、时序数据库 17 个、键值数据库 10 个，在非关系型数据库中依次占比 27.9%、25.0%和 14.7%。

## 5.非关系型数据库呈现百花齐放态势，向量数据库热度持续攀升

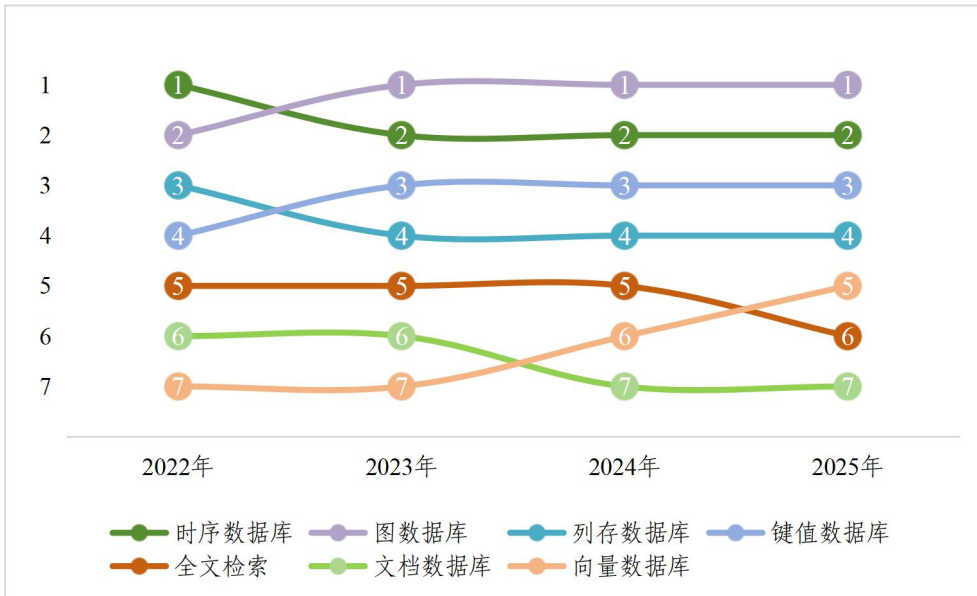
全球数据库产品呈现关系型数据库数量占比缓慢下降、非关系型数据库产品数量逐渐上升的趋势。非关系型数据库从2022年的40.6%上升到2025年的57.4%，产品热度持续走高。



来源：CCSA TC601, 2025年6月

图 11 过去四年全球非关系型数据库数量占比排名

全球来看，近四年非关系型数据库领域竞争激烈。键值数据库产品占比稳居第一，受到市场长期认可。图数据库、时序数据库、文档数据库等市场中占据相对领先地位，反映出市场对于特定场景下数据存储和处理的需求较为稳定且明确。向量数据库随 AI 浪潮兴起，其产品数量占比逐年上升。随着技术发展和业务场景的多样化，各类型数据库也在不断调整，以适应市场需求。

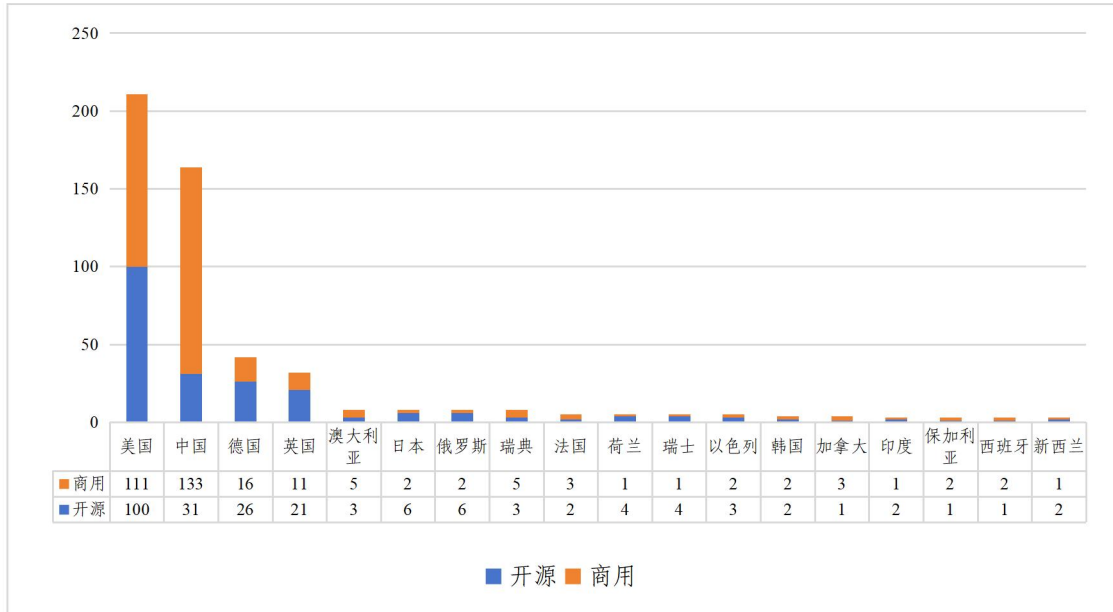


来源：CCSA TC601, 2025 年 6 月

图 12 过去四年我国非关系型数据库数量占比排名

我国来看，近四年非关系型数据库市场呈现出多样化的发展态势。图数据库和时序数据库凭借其在特定领域的强大支持，保持领先地位。与全球趋势一致，向量数据库 2023 年以来发展迅猛，排名不断提升。全文检索和文档数据库产品数量逐渐下降，其余数据库产品数量保持稳定。随着我国数字经济的不断发展，各行业对数据存储和处理的需求日益复杂和多样化，将推动非关系型数据库市场持续创新。

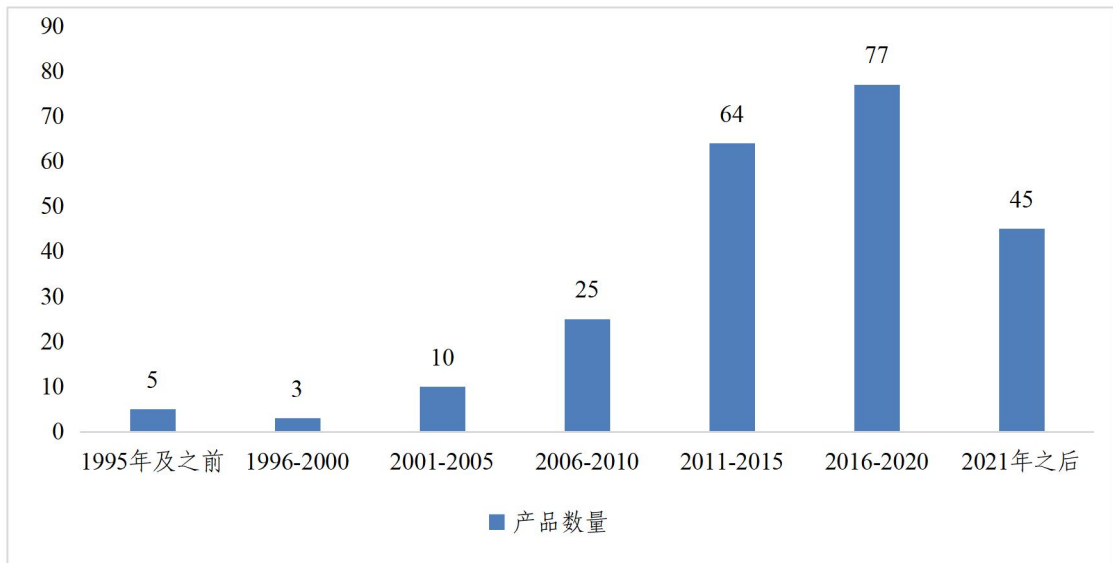
## 6.国外商用数据库占比略有上升，我国以商用为主



来源：CCSA TC601,2025 年 6 月

图 13 全球数据库产品商用开源对比

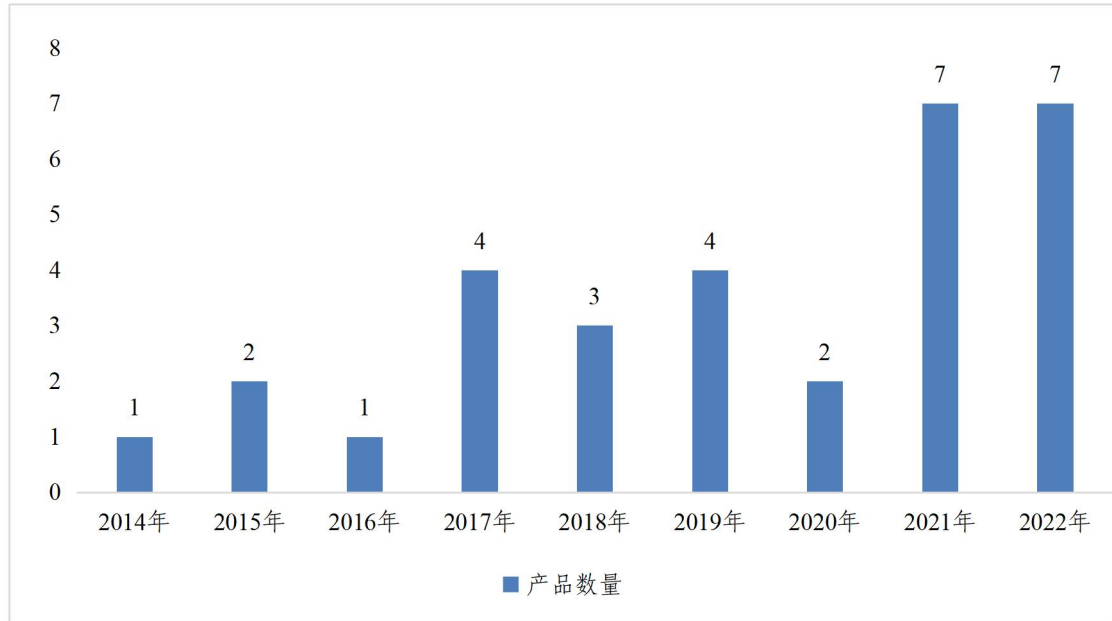
欧美等国商用数据库占比略有上升，我国商用数据库占比较大。美国开源与商用数据库数量基本持平，占美国全部数据库产品数量的比例分别为 47.4%和 52.6%。我国以商用为主，商用和开源占我国全部数据库产品数量比分别为 81.1%和 18.9%。



来源：CCSA TC601，2025 年 6 月

图 14 全球活跃开源数据库开源时间

全球开源数据库兴起于 20 世纪 90 年代。自 90 年代开源数据库不断推出，于 2006 年后迅速发展，产品数量每隔 5 年呈 2-3 倍增长。其中在 2011-2020 年，开源数据库进入发展高峰期，大量开源数据库产品不断推出。



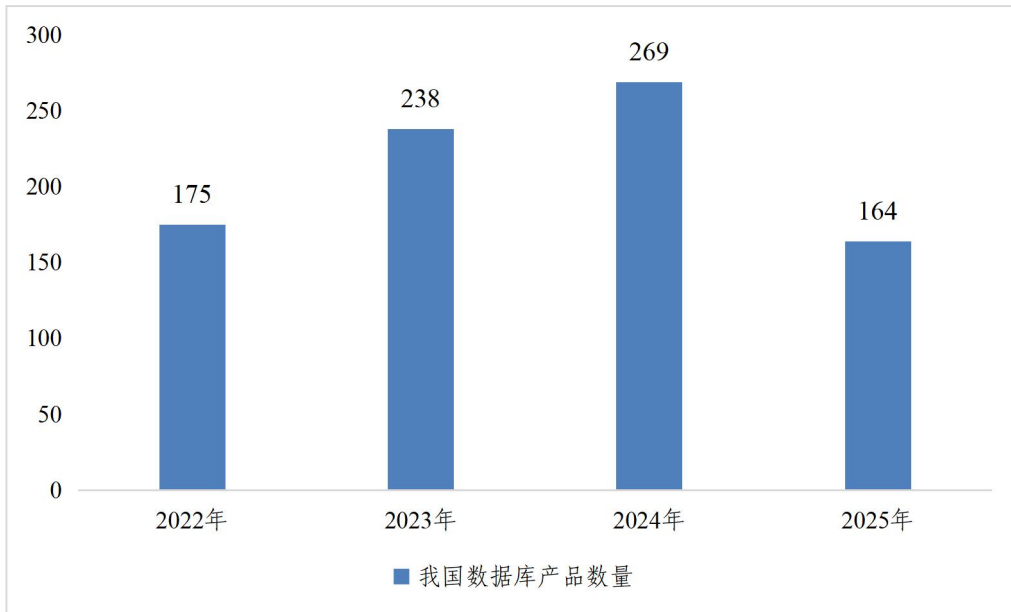
来源：CCSA TC601，2025 年 6 月

图 15 我国活跃开源数据库的开源时间

我国开源数据库产品始于 2010 年后，2017 和 2021 年开源产品数量激增。我国开源数据库整体起步较晚，目前正常运营维护的数据库产品基本开源于 2014 至 2022 年之间，2021-2022 年数据库开源生态蓬勃发展，新增开源产品数量达到新高。

### 7.我国数据库市场规模持续扩大，市场格局逐步显现

2024 年，中国数据库市场头部效应明显，产业规模持续扩大，市场格局逐步显现。国产数据库从高速增长期进入产业质变期，投资市场回归理性，供给侧从“烧钱”向“赚钱”转型，产业侧从“数量型”向“质量型”转变。



来源：CCSA TC601，2025年6月

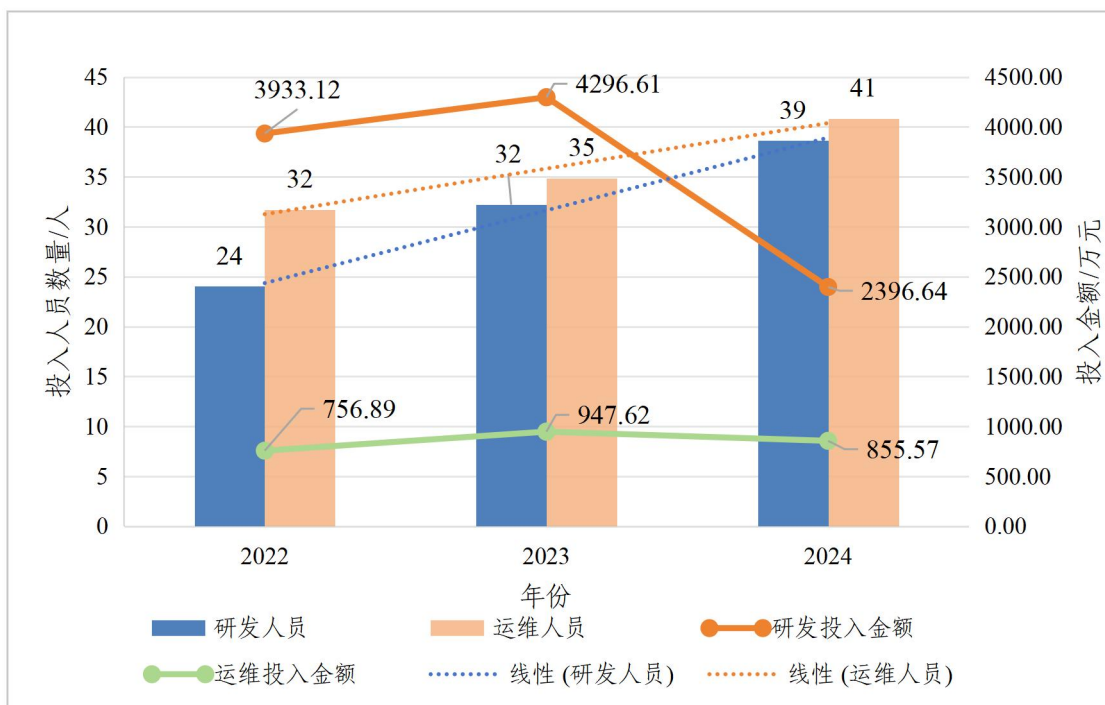
图 16 我国数据库产品数量

从我国数据库产品数量看，2022年至2024年，我国数据库产业规模逐步扩大，产品数量呈现迅速增长的趋势，由2022年的175款产品增加到2024年的269款产品，复合增长率约为24.0%；2025年，我国数据库市场规模持续扩大，但数据库产品数量逐步收敛，市场格局开始显现。

## 8.重点行业头部用户运维投入逐年增长，研发投入有所回落

自2022年以来，中国信息通信研究院联合CCSA TC601启动大数据“星河（Galaxy）”案例数据库方向征集工作，为我国数据产业发展和企业数智化转型树立了诸多高质量、示范性、可推广的行业应用创新案例。以下内容对过去三年“星河（Galaxy）”案例中数据库专项入选企业进行统计分析<sup>6</sup>。

<sup>6</sup> 统计范围为截至2025年6月30日已入选案例。



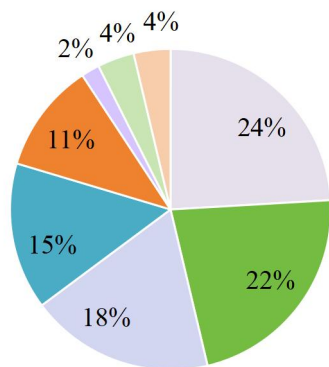
来源：CCSA TC601，2025 年 6 月

图 17 2022-2024 年重点行业头部用户研发/运维人员数量和投入资金变化

综合来看，过去三年入选企业人员规模均呈逐年增长趋势，数据库相关资金投入在 2023 年达峰值后于 2024 年显著回落，其中研发投入资金下降明显，运维投入金额保持相对稳定。

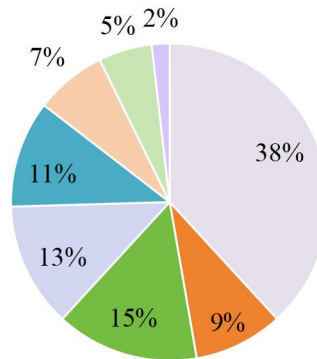
### 2024 年研发投入资金

- 0-100万
- 100-300万
- 300-500万
- 501-1000万
- 1000-2000万
- 2000-3000万
- 3000-5000万
- 5000万以上



### 2024 年研发人员

- 0-5人
- 5-10人
- 10-30人
- 15-30人
- 30-50人
- 50-100人
- 100-200人
- 200以上



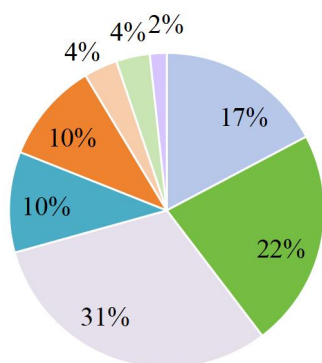
来源：CCSA TC601，2025 年 6 月

图 18 2024 年研发人员数量及投入资金区间分布

从企业研发投入资金看，超六成企业投入资金在 500 万及以下区间；从研发人员数量看，近四成企业研发人员数量少于 5 人，仅有 2%的企业拥有 200 人以上研发人员。

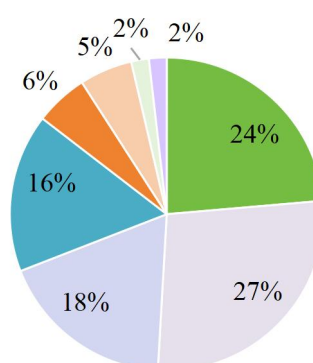
**2024运维投入资金**

- 0-100万
- 100-300万
- 300-500万
- 501-1000万
- 1000-2000万
- 2000-3000万
- 3000-5000万
- 5000万以上



**2024年运维人员**

- 0-5人
- 5-10人
- 10-15人
- 15-30人
- 30-50人
- 50-100人
- 100-200人
- 200以上



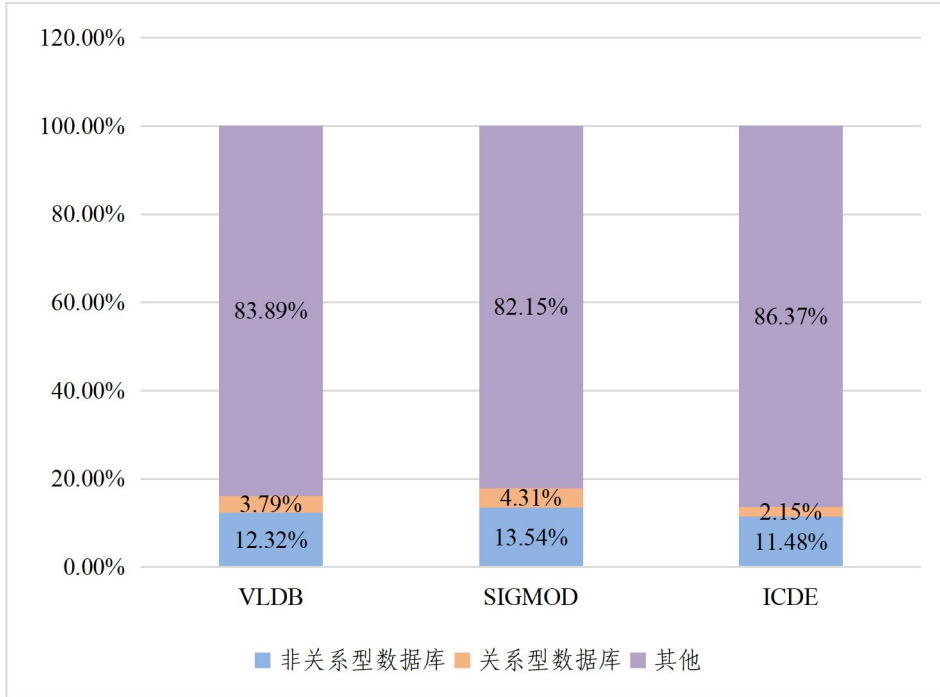
来源：CCSA TC601，2025 年 6 月

图 19 2024 年运维的人员数量/投入资金区间分布

从企业运维投入资金看，超七成企业投入资金在 500 万及以下区间，300-500 万区间占比最多为 31%；从运维人员数量看，近七成企业运维人员数量少于 15 人，规模为 5-10 人区间占比最多为 27%。

### (三) 数据库支撑体系

1. 创新方面，多模融合成为研究重点，我国科研实力稳步增强



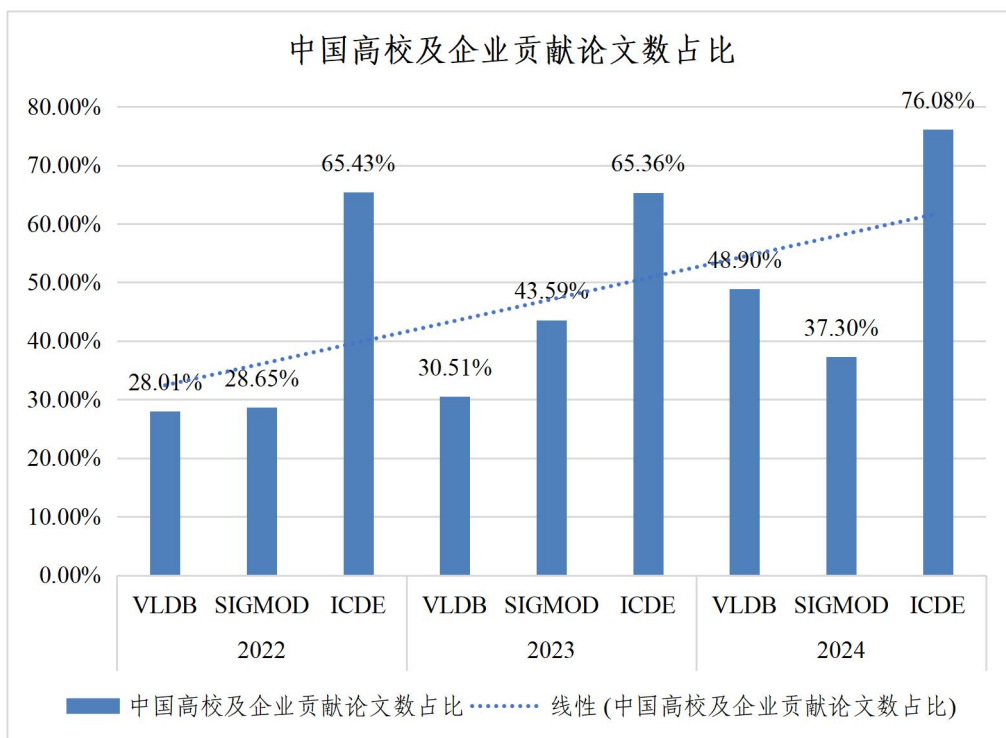
来源：CCSA TC601，2025 年 6 月

图 20 2024 年 VLDB、ICDE 和 SIGMOD 论文分布情况

从 VLDB、SIGMOD 和 ICDE 三个数据库领域权威的学术会议研究方向看，与近两年统计情况基本保持一致<sup>7</sup>，当前非关系型数据库研究内容数量占比完全超过关系型数据库。以 VLDB 为例，各类数据库论文的发表数量分别为：非关系型数据库论文 52 篇，关系型数据库论文 16 篇，以及其他类型论文 354 篇。在三年的论文总数中，非关系型数据库论文所占比例为 12.32%，关系型数据库论文所占比例为 3.79%。在 SIGMOD 会议中，各类数据库论文的发表数量分别为：非关系型数据库论文 44 篇，关系型数据库论文 14 篇，以及其他类型论文 267 篇。非关系型数据库论文占总论文数量的 13.54%，关

<sup>7</sup> 近两年统计情况可分别参考《数据库发展研究报告（2023）》、《数据库发展研究报告（2024）》

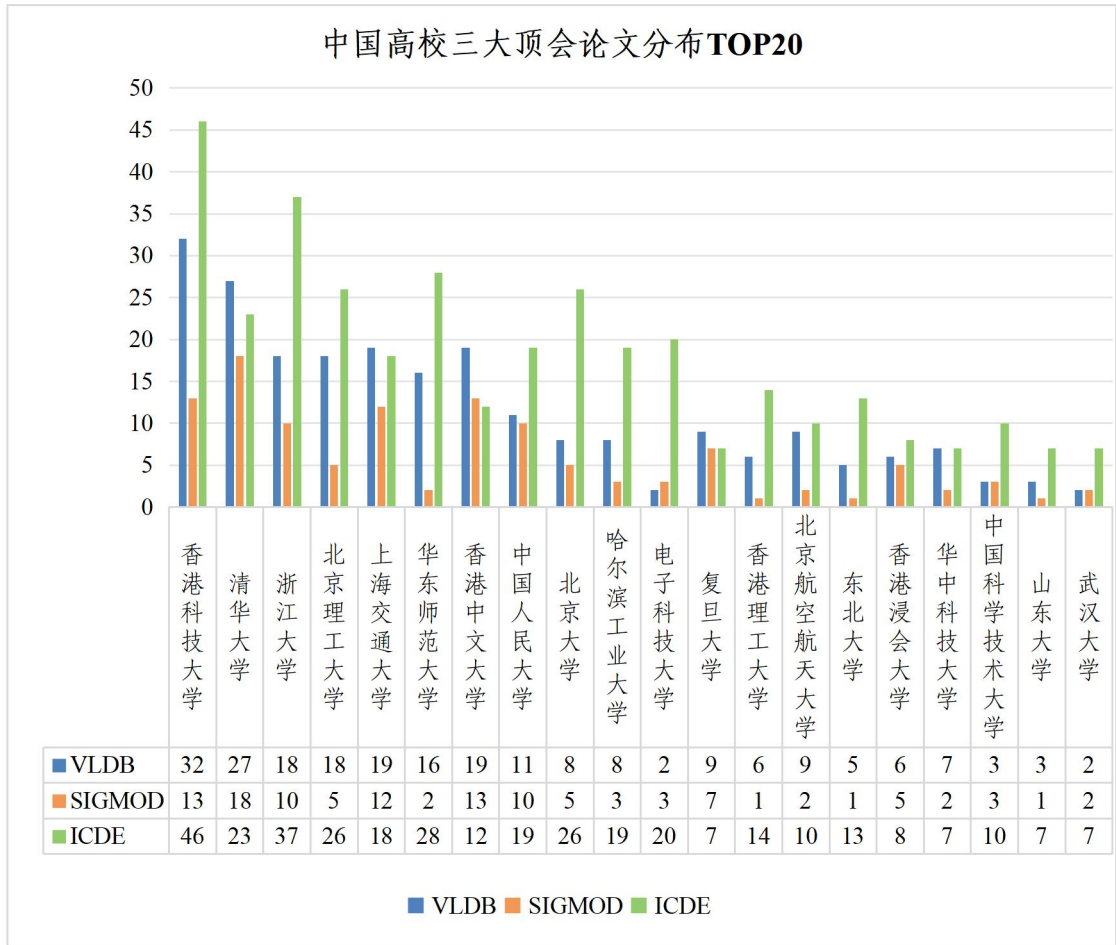




来源：CCSA TC601，2025年6月

图 22 2022-2024 年中国高校及企业学术会议论文贡献情况

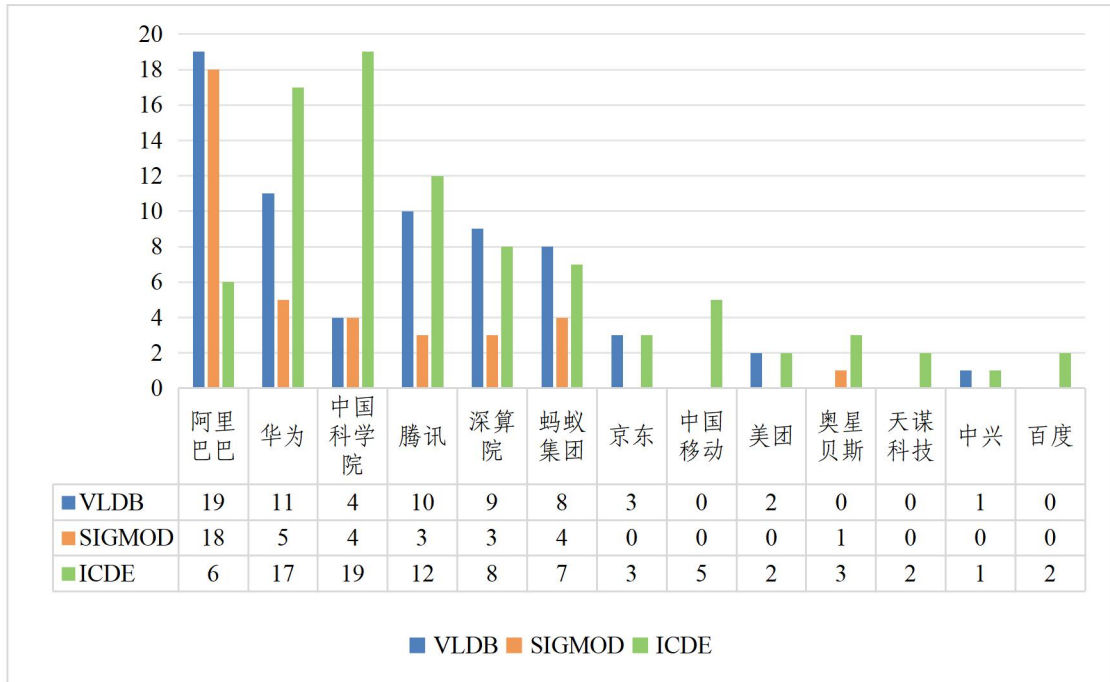
我国在全球三大数据库领域学术会议的影响力进一步提升。我国高校及企业在 ICDE 论文贡献占比最高，三年依次为 65.43%、65.36% 和 76.08%，三大会议每年贡献占比平均为 40.70%、46.35% 和 54.09%，数量呈逐年上升趋势，2022-2023 年中国高校企业在学术论文会议的贡献率涨幅为 5.65%，2023-2024 年涨幅为 7.74%。2024 年中 VLDB 与 ICDE 相较于前两年，中国高校与企业所贡献的论文占比有 10-20% 的大幅提升，主要源于近年来企业与高校不断重视数据库理论创新与应用实践相结合，有效助力了我国数据库技术水平快速发展，学术影响力不断扩大。



来源：CCSA TC601，2025年6月

图 23 2024 年中国高校三大会议论文数量前 20 名

2024 年入选三大顶会论文有将近 50 所中国高校，其中香港科技大学入选数量居于首位，共有 91 篇，在 VLDB 和 ICDE 顶会中分别发表了 32 篇和 46 篇文章，在中国高校中贡献率达到了 7.6% 和 11.0%。此外，超过 10 篇文章的高校包括清华大学、浙江大学、北京理工大学、上海交通大学、华东师范大学、香港中文大学、中国人民大学、北京大学、哈尔滨工业大学、电子科技大学、复旦大学、香港理工大学、北京航空航天大学、东北大学、香港浸会大学、华中科技大学、中国科学技术大学、山东大学、武汉大学、广州大学、国防科技大学、苏州大学、湖南大学。



来源：CCSA TC601，2025 年 6 月

图 24 2024 年中国企业三大会议论文数量前 10 名

我国企业不断加大数据库学术领域投入，通过持续创新将理论转化为应用落地，推动数据库技术的进步与发展。阿里巴巴、华为、中国科学院、腾讯、深圳计算科学研究院、蚂蚁集团、京东、中国移动、美团、奥星贝斯、天谋科技、中兴通讯、百度、KaiwuDB 等企业均在数据库三大顶会中发表了相关研究成果，其中阿里巴巴、华为、中国科学院、腾讯和深圳计算科学研究院入选三大顶会总数位列前五。

## 2. 投融资方面，PG 生态获得青睐，多云管理和 AI 成为融资亮点

过去一年，被广泛应用于云环境及集成诸多 AI 能力的 PostgreSQL 生态公司获得资本市场青睐，不仅彰显其易用性、兼容性带来的强大吸引力，也预示着多云管理和 AI 赋能将成为未来数据库投融资的焦点。

2025 年 4 月，多云数据库平台服务商 Tessell 获得 6000 万美元融

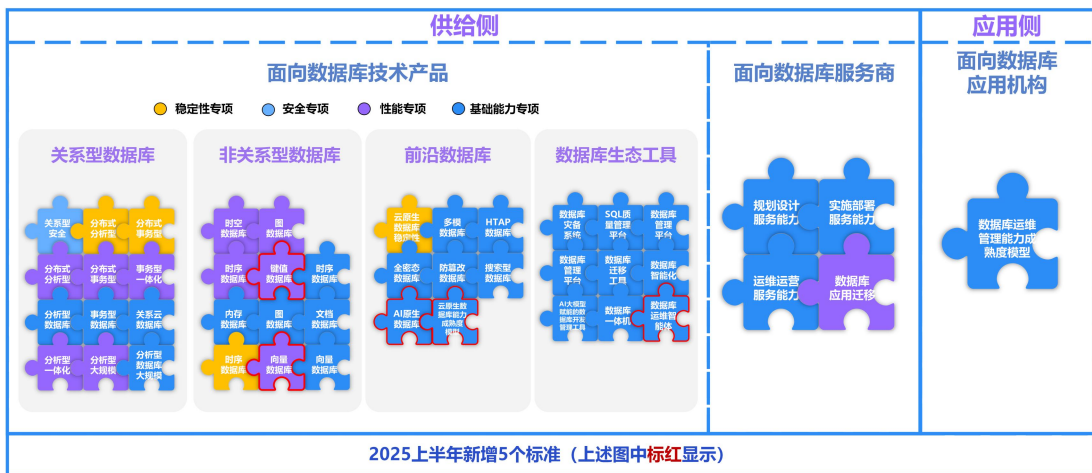
资，此次融资将用于进一步扩大其市场覆盖，并计划推出基于 AI 驱动的对话式数据库管理服务。同月，PostgreSQL 数据库云平台提供商 Supabase 宣布完成 2 亿美元 D 轮融资，凭借托管 PostgreSQL、身份验证、实时服务以及与 AI 大模型生态工具 Ollama 的友好集成能力，使得开发者能够在本地完全运行由 AI 驱动的完整开发环境，其推出的 AI 助手能让非开发背景的“提示词工程师”快速完成后端原型开发，极大提升 AI 原生应用的开发体验和效率。本轮投资后 Supabase 估值 20 亿美元，距离其上一轮 8000 万美元融资仅过去 7 个月，累计融资已达近 4 亿美元。此外，MySQL 的创始人之一，Vitess 之父 Sugu 于 2025 年 6 月正式宣布加入 Supabase，全面主导基于 PostgreSQL 的全新分布式数据库项目 "Multigres"。

2025 年 5 月，Databricks 宣布以 10 亿美元收购 Serverless Postgres 托管服务商 Neon，加强其在多云环境中数据处理能力的战略布局。同年 6 月，其竞争对手 Snowflake 宣布以 2.5 亿美元收购 PostgreSQL 的商业发行版公司 Crunchy Data，以扩大其 AI Data Cloud 的产品版图。这些收购行为不仅巩固了 PostgreSQL 生态在全球数据库市场中的地位，也显示了资本市场对于 PostgreSQL 技术发展方向和社区生态的高度认可。

综上所述，无论是从资本投入规模，还是从技术创新角度来看，PostgreSQL 生态均展现出强劲的发展势头。多云管理和 AI 赋能的数据库技术正成为领域投资热点。

3.标准方面，我国数据库标准体系日益完善，引领数据库产业高质量发展

2024年9月25日，国家发展改革委、国家数据局、中央网信办、工业和信息化部、财政部及国家标准委联合印发了《国家数据标准体系建设指南》（以下简称《建设指南》），旨在充分发挥标准在激活数据要素潜能、做强做优做大数字经济等方面的规范和引领作用，明确了到2026年底基本建成国家数据标准体系的目标，围绕数据流通利用基础设施、数据管理、数据服务等多个方面制修订30项以上数据领域基础通用国家标准。数据库作为数据基础设施的重要组成部分，其相关标准能够为技术研发、产品选型及应用开发等环节提供明确规范和准则，促进数据共享、交换与流通，为数字经济高质量发展奠定坚实基础。



来源：CCSA TC601，2025年6月

图 25 CCSA TC601 数据库领域标准化工作体系

中国通信标准化协会大数据技术标准推进委员会(CCSA TC601)紧跟国家战略，围绕数据库领域标准化工作，设立数据库与存储工作组（WG4）。自2015年起共推出40余项标准，逐步构建以数据库产品、服务和应用为目标的标准体系。产品能力方面，从关系型和非关系型，构建了基础能力、性能和稳定性的评估测试项目；服务能力

方面，围绕规划设计、实施部署和运维运营，推出国内首个《数据库服务能力成熟度模型》，围绕数据库应用迁移和 SQL 质量管理平台，推出能力分级标准；行业应用方面，面向数据库应用方内部运维管理团队，推出《数据库运维管理能力成熟度模型》，已在中国移动浙江、湖北、陕西、北京、上海、江苏六家省公司和上证所信息网络有限公司成功贯标。十年来，CCSA TC601 见证了我国数据库标准化工作有序有力进行，成为国家在数据库领域最重要的支撑单位，已搭建国内权威的第三方数据库评估测试体系，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，圈定了国内数据库厂商第一梯队，成为了数据库产业标准化发展的风向标。

## 二、数据库关键技术发展趋势

近年来，随着数智技术快速发展和赋能，数据库技术不断创新演进。一方面，数据库技术架构持续革新。从交易与分析分而治之到交易分析一体化，从本地部署到云上运行，再到如今云原生架构兴起，数据库部署和管理模式变得更加灵活高效。另一方面，数据库技术迈入 AI 原生时代，以生成式 AI 为代表的人工智能技术快速迭代，为数据库技术不断注入新动力。自然语言与数据库交互，机器学习和 RAG 技术，共同优化数据库应用，使得数据库的应用模式变得更加精准智能。

### (一) 技术架构持续革新

数据库技术已被广泛应用在电信、金融、能源等行业，在数据存储管理、查询分析等方面发挥了重要作用。过去五十年中，数据库从以 PostgreSQL 和 MySQL 等解决了数据存储、管理和查询问题的单机数据库，到以 Oracle RAC、IBM Db2 和 Microsoft SQL Server 等为企业关键业务提供高可用性和高可靠性保障的集群数据库，再到以亚马逊 Aurora、华为云 GaussDB 及腾讯云数据库等解决大数据时代弹性计算和动态数据迁移问题的分布式数据库与云原生数据库，不断完善、持续变革<sup>8</sup>。

#### 1. 从分而治之到交易分析一体化

OLAP 数据库擅长数据分析，OLTP 数据库专注于事务处理。随着数智时代的到来，图数据、时间序列数据、空间数据等新型数据类型不断涌现。新一代 OLAP 数据库不仅需要支持关系型数据管理，也需要支持多模数据并能够集成人工智能技术以应对新时代的挑战。而

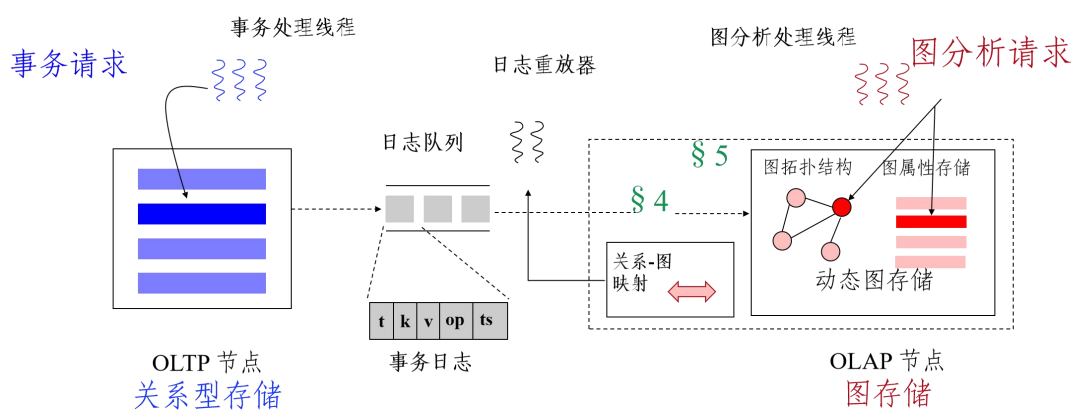
---

<sup>8</sup> 李国良,周焯赫. 轩辕:AI 原生数据库系统[J]. 软件学报, 2020, 31(3): 831-844.

新一代 OLTP 则可通过新型硬件如 AI 芯片、RDMA 以及 NVM 来改进事务处理。

随着应用场景的变化和数据量的积累，在事务处理的基础上，企业对于实时数据的分析需求不断提升，OLAP 和大数据技术栈不断分化也给企业运维管理多套系统带来挑战。HTAP (Hybrid Transactional and Analytical Processing) 架构应运而生，通过将事务处理与分析处理融合在一个系统中，消除了 OLAP 与 OLTP 之间的壁垒。HTAP 数据库按存储策略分可分为主行存与内存型列存、分布式行存与列存副本、单机磁盘型行存与分布式列存以及列存与增量型行存四大类<sup>9</sup>。

近年来，能够支持交易分析及图分析的图 HTAP 数据库也受到广泛关注。许多应用程序需要对关系型 OLTP 生成的数据集进行实时动态图分析处理任务。此外，在图数据库实际使用过程中，70%的任务包含事务和分析混合处理。为满足以上需求，上海交通大学、阿里巴巴及上海人工智能实验室的专家们提出了能够进行图分析处理的 GART 系统；来自创邻科技、香港科技大学等专家学者提出能够进行事务分析混合处理的原生分布式图数据库 Galaxybase<sup>10</sup>。



来源：上海交通大学等

图 26 GART 系统的总体架构（虚线框内的组件是 HTGAP 扩展的新功能）

9 张超,李国良,冯建华,等.HTAP 数据库关键技术综述[J].软件学报,2023,34(02):761-785.

10 Bing Tong, Yan Zhou, Chen Zhang, Jianheng Tang, Jing Tang, Leihong Yang, Qiye Li, Manwu Lin, Zhongxin Bao, Jia Li, and Lei Chen. Galaxybase:A High Performance Native Distributed Graph Database for HTAP. PVLDB,17(12): 3893 - 3905, 2024.

GART 系统是一个基于内存的混合事务处理和图分析 (HTGAP) 系统,该系统可以在 HTAP 系统的基础上通过图抽取结构定义关系型数据与图数据映射规则以适应多样化工作负载。基于对 HTGAP 工作负载特性的洞察,使用具备良好局部性的高效动态图存储结构以保障系统的图分析性能<sup>11</sup>。

## 2.从本地部署到云上运行

云计算自 2006 年开始迅速发展,各组织上云进程不断加速,传统数据库逐渐从私有部署转为云上部署。然而,仅改变数据库部署模式并不代表能够充分利用云为数据库赋能。传统数据库为了保障系统高性能,基本采用紧耦合设计,而云计算架构的理念则是通过计算、存储的高度解耦来实现各个层次资源的高效利用和弹性扩展<sup>12</sup>。为了更好地适应云计算的发展趋势,数据库架构也在不断变革。

### 2.1 存算分离架构提升资源利用率

从云就绪到云原生,尽管诸多数据库都可以在云上运行,但与真正生于云、长于云的云原生数据库却不尽相同。近年来,存算分离架构在云环境中得到了广泛应用,通过将计算与存储解耦,实现计算和存储资源的独立、弹性扩展,从而提高资源利用率并降低总体成本。Amazon Aurora、Microsoft Socrates、Google AlloyDB、Alibaba PolarDB 和 Huawei Taurus、Neon 等数据库产品,均充分利用了存算分离架构的优势<sup>13</sup>。

---

11 Shen, S., Yao, Z., Shi, L., Wang, L., Lai, L., Tao, Q., Su, L., Chen, R., Yu, W., Chen, H., & Zang, B. (2023). Bridging the Gap between Relational OLTP and Graph-based OLAP. In 2023 USENIX Annual Technical Conference (USENIX ATC '23) (pp. 1-16).

12 李国良,王磊,张金玉,等.GaussDB: 智能云原生分布式数据库[J/OL].电子学报,1-21

13 Xi Pang and Jianguo Wang. 2024. Understanding the Performance Implications of the Design Principles in Storage-Disaggregated Databases. Proc. ACM Manag. Data 2, 3, Article 180 (June 2024), 26 pages.

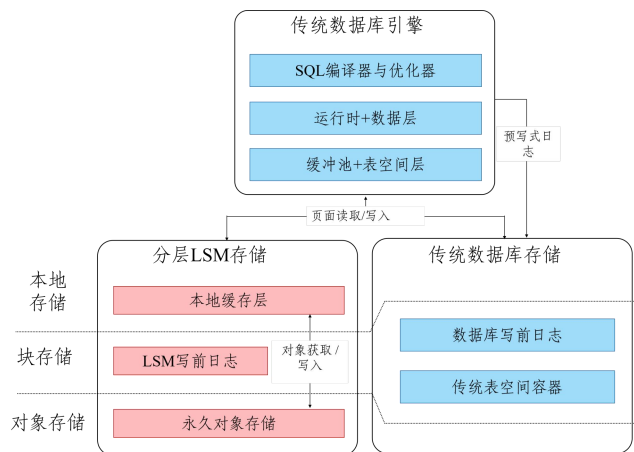
存算分离数据库通常采用软件级存算分离、日志即数据库以及共享存储设计三类创新设计原则。软件级存算分离是将存储引擎（如日志记录和存储）与计算引擎（如 SQL 层、缓冲区和事务）解耦，其中存储引擎运行在存储节点上，计算引擎运行在计算节点上。这种设计对于存算分离数据库引入新的优化能力至关重要，例如通过日志即数据库设计，以减轻网络 I/O 开销。日志即数据库设计是在事务提交时仅将写前日志发送到存储端，而不是像传统数据库那样发送实际数据页，从而减少网络中数据传输量，实际数据页随后会从存储端异步生成。共享存储设计则区别于传统的无共享架构，是多个计算节点之间共享存储设计，新添加的计算节点可以与现有计算节点直接共享相同的数据，从而消除了为新添加的计算节点复制或移动数据的需求，因此该设计也增强了系统弹性。

## 2.2 云原生数据库实现高效资源管理

诸多企业采用的存算分离架构以及应用其他相关技术也使得产品在云环境中能力得到提升：IBM Db2 Warehouse<sup>14</sup>、TiDB 等采用 LSM-Tree、列式存储等新型结构使得产品在存储引擎专业化方面得到突破，Microsoft Azure SQL、TDSQL-C 等采用资源动态调整和主动资源分配使得资源实现精细化管理，阿里云 AnalyticDB 采用 TP/AP 工作负载隔离实现混和负载智能调度，CockroachDB 采用存算分离实现跨洲多活使得全球分布式架构实现突破，AWS Redshift Spectrum 支持跨云数据访问实现数据服务生态重构，华为云 GaussDB 采用软件结合硬件实现软硬件协同创新加速。

---

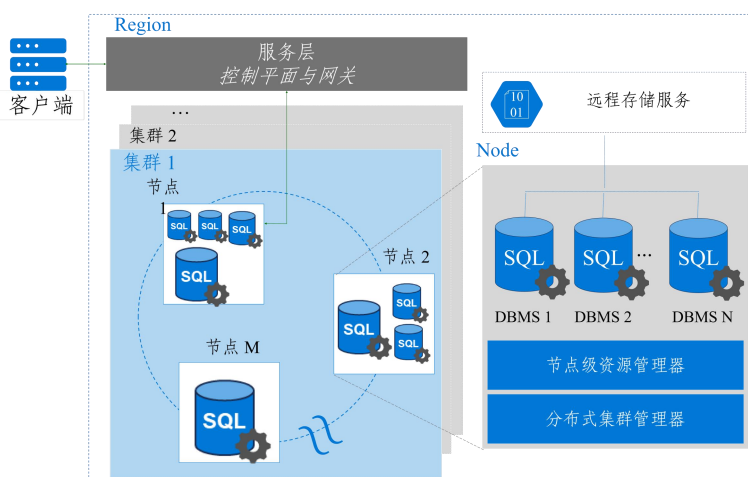
14 David Kalmuk, Christian Garcia-Arellano et al. 2024. Native Cloud Object Storage in Db2 Warehouse: Implementing a Fast and Cost-Efficient Storage Architecture.



来源：International Business Machines Corporation, IBM

图 27 Db2 LSM 架构图

传统数据库系统通常使用数据块或数据页存储数据，虽然适合随机 I/O 操作，但在云环境中并不具有经济优势。AWS、Azure 和 Google 等公有云服务商可以通过本身的架构优势提供更具有性价比的服务，而本地部署服务商在云环境下针对内核进行重构的成本又相对较高。为了弥补这一劣势，IBM Db2 通过 LSM 树，实现了在云对象存储中高效存储数据页，结合多存储架构、数据聚类优化等技术降低存储成本的同时提升了使用性能。此外，该架构未对数据库内核进行重大重构，从而保留了 Db2 强大的 SQL 和事务处理能力<sup>15</sup>。

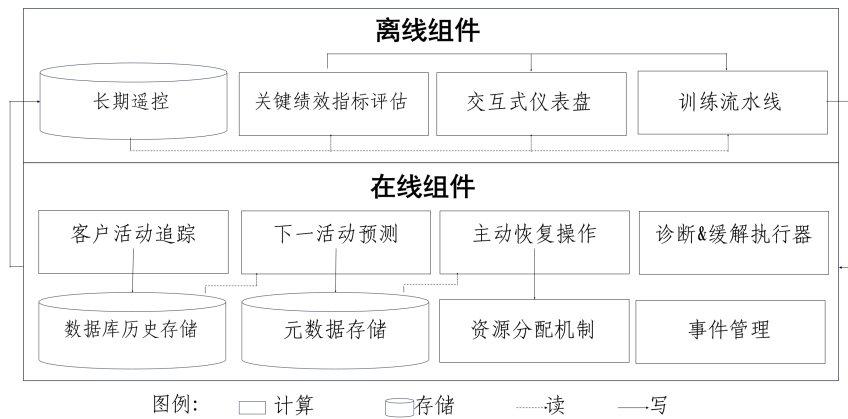


来源：Microsoft Corporation

图 28 AzureSQL 数据库架构概览

15 David Kalmuk et al. 2024. Native Cloud Object Storage in Db2 Warehouse: Implementing a Fast and Cost-Efficient Storage Architecture. In Companion of the 2024 International Conference on Management of Data (SIGMOD-Companion '24)

Azure SQL 数据库是微软公司提供的关系型数据库即服务（DBaaS），通常采用多租户模式，允许多个数据库共享物理资源。由于在实际使用过程中，工作负载不断变化，为了保障用户资源请求总能得到满足，服务提供商会给每位数据库租户静态预留最大允许资源。但由于成本原因，通常存在资源超订问题从而影响服务质量。为了寻找资源高可用和运营低成本之间的最优平衡，微软的工程师们提出了 Proactive Resume and Pause of Resources (ProRP)<sup>16</sup>以及灵活资源分配（Flexible Resource Allocation）的架构<sup>17</sup>。



来源：Microsoft Corporation

图 29 ProRP 架构示意图

ProRP 是一种针对 Serverless 数据库的主动资源分配基础设施，通过在线组件的方式直接集成到 Azure SQL 的代码库，能够利用数据库生命周期和历史数据预测用户未来活动，动态调整资源分配从而避免资源闲置和浪费。微软工程师提出的资源分配架构则可通过多租户内存代理（Multi-Tenant Memory Brokering）、基于利用率的 CPU 重平衡（Utilization-Based CPU Rebalancing）、租户放置优化（Tenant

<sup>16</sup> Olga Poppe, Pankaj Arora, Sakshi Sharma, Jie Chen et al. 2024. Proactive Resume and Pause of Resources for Microsoft Azure SQL Database Serverless. In Companion of the 2024 International Conference on Management of Data

<sup>17</sup> Pankaj Arora, Surajit Chaudhuri, Sudipto Das, Junfeng Dong et al. Flexible Resource Allocation for Relational Database-as-a-Service. PVLDB, 16(13): 4202 - 4215, 2023.

Placement) 以及故障转移影响缓解 (Mitigating Failover Impact) 实现在资源超分配情况下的高效资源管理。

中国信通院联合腾讯云、移动云、天翼云等多家企业研制了《云原生数据库能力成熟度模型》标准。当前云原生数据库能力成熟度分为四个等级：一是初始级，即数据库主要依赖传统数据库技术和运维模式，架构较为单一，缺乏灵活性和可扩展性；二是发展级，已经引入云原生数据库技术，但尚未形成完整的云原生数据库生态。运维和管理方式仍较为传统，缺乏自动化和智能化；三是成熟级，已经实现了云原生数据库的大部分自动化和智能化管理，采用了先进的云原生技术和架构，提升了数据库的灵活性和可扩展性，建立了完善的监控和反馈机制，能够及时发现和解决数据库运行中的问题；四是领先级，在云原生数据库领域处于行业领先地位，能够自主研发并发布创新的云原生数据库解决方案，满足复杂业务场景的需求，建立了完善的云原生数据库生态体系，包括数据库服务、开发工具、运维平台等。

## **(二) 数据库迈入 AI 原生时代**

### **1. 人工智能与数据库双向赋能**

近年来，人工智能迅猛发展影响着千行百业，数据库作为数据存储和管理的主要载体，与人工智能的融合发展正深刻地改变着数智应用的格局。一方面，人工智能技术为数据库的管理、优化和应用带来了智能化的手段；另一方面，数据库也为人工智能提供了海量、多样的数据支持，为模型训练和推理奠定了坚实基础。

#### **1.1 人工智能拓宽数据库应用场景 (AI for DB)**

随着人工智能技术的发展，数据库领域在查询优化、自然语言交互和智能资源管理等方面迎来了新的突破。机器学习技术被引入数据

库系统内部，为解决复杂查询优化问题提供了全新思路。大语言模型的兴起革命性地实现了通过自然语言与数据库系统的高效交互，大幅提升了用户使用数据库的便捷性。此外，数据库系统与检索增强生成技术深度融合，进一步拓展了大语言模型的知识边界与应用场景，为智能化数据处理开辟了更为广阔的前景。

### 1.1.1 机器学习优化数据库应用效能

随着机器学习技术不断发展，其在数据库管理系统中的应用逐渐增多，涵盖数据库的索引、数据布局、调度器、查询优化器等数据库核心组件，以及性能预测、参数调优、索引等数据库顾问的各个方面<sup>18</sup>。

表 1 ML4DB 研究中查询计划表示方法总结

来源：南洋理工大学

方法名称	应用场景	树模型类型
AVGDL	视图选择	LSTM
AIMeetsAI	索引选择	Feature Vector
ReJOIN	连接顺序选择	Feature Vector
BAO	优化器	TreeCNN
NEO	优化器	TreeCNN
Prestroid	成本估计	TreeCNN
E2E-Cost	成本/基数估计	TreeLSTM
RTOS	连接顺序选择	TreeLSTM
Plan-Cost	成本估计	TreeRNN
QueryFormer	通用用途	Transformer

ML for DB 的基础是查询计划表示和预训练模型。查询计划表示是将数据库查询计划转换为机器学习模型能够处理的形式，这一过程通常包括特征编码和树模型应用两个阶段。预训练模型是 ML for DB 的另一个重要基础，旨在解决训练数据收集成本高的问题，这些模型

<sup>18</sup> Gao Cong, Jingyi Yang, and Yue Zhao. 2024. Machine Learning for Databases: Foundations, Paradigms, and Open problems. In Companion of the 2024 International Conference on Management of Data.

能够在不同任务和数据集上进行迁移学习，从而减少新任务的数据收集和模型训练成本。

ML for DB 范式主要分为替代范式 (Replacement) 和机器学习增强范式 (ML-enhanced)。早期研究多采用替代范式，即直接用机器学习模型取代传统数据库组件，如递归模型索引 (RMI) 和递归空间模型索引 (RSMI)，通过机器学习模型替代现有组件以提升性能。尽管在某些场景下，替代方法优于传统组件，但也存在鲁棒性差和泛化能力有限等问题。为克服这些局限，机器学习增强范式应运而生，它在保留传统组件的基础上，利用机器学习优化性能，如 Bandit 优化器、DBG-PT<sup>19</sup>等，展现出更好的鲁棒性和泛化能力。

ML for DB 领域仍面临诸多挑战，一是模型效率有限，数据库系统对性能要求极高，因此机器学习模型需要在训练与推理速度方面进行优化。二是数据处理器和工作负载变化问题，数据库运行在数据和查询工作负载不断的动态环境中，大多数机器学习模型在面对这些变化时需重新训练。三是当前基础模型在单一任务和数据集设置中表现出色，但其有效性依赖于任务和数据集特定模式，严重限制了 ML for DB 系统的实用性。四是收集模型训练的数据集成本极高且质量难以保证。为应对这些挑战，专家们也从不同方面寻找新的思路。通过利用贝叶斯深度学习技术加速模型训练、借助机器学习的基数估计和数据生成技术提升对数据和工作负载变化的适应性、通过多样化训练数据学习通用特征使模型能够适应不同数据从而探索自监督学习技术、运用元学习技术实现模型跨任务泛化，以及开发低成本创建增强或模拟现实数据集的技术同时确保隐私合规。

---

19 Victor Giannakouris and Immanuel Trummer. DBG-PT: A Large Language Model Assisted Query Performance Regression Debugger. PVLDB, 17(12):4337 - 4340, 2024.

### 1.1.2 Text to SQL/Text to GQL 降低数据库交互门槛

数据库操作是各行业从数据库中获取信息的关键环节，传统的数据库查询通常依赖结构化查询语言（SQL），需要用户具有一定的数据库知识和技能才能高效地构建查询。随着人工智能和自然语言处理技术的进步，Text-to-SQL 技术应用而生，研究者开始利用深度学习模型对文本数据进行训练，使得系统能够更准确地理解用户意图从而降低数据库使用门槛。

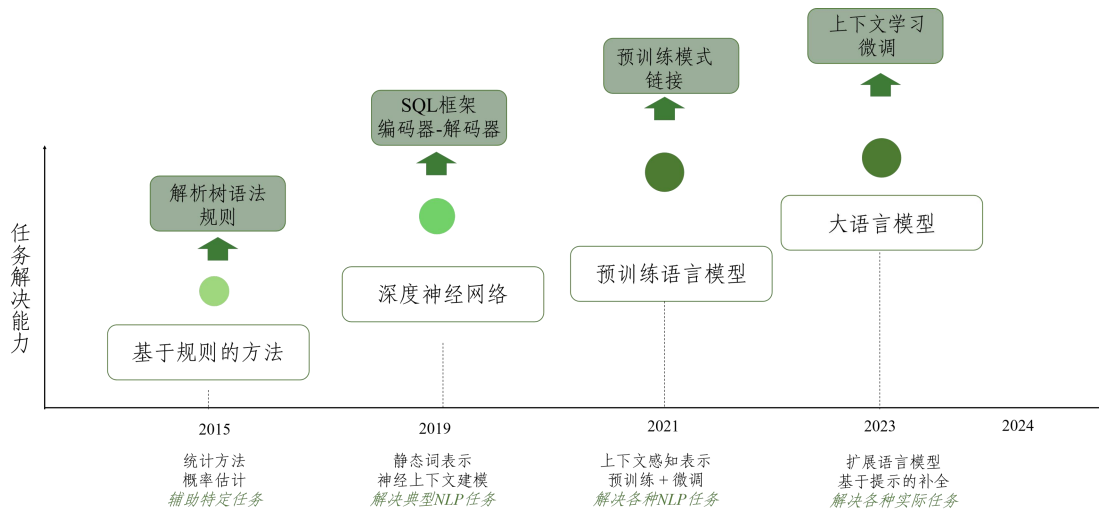
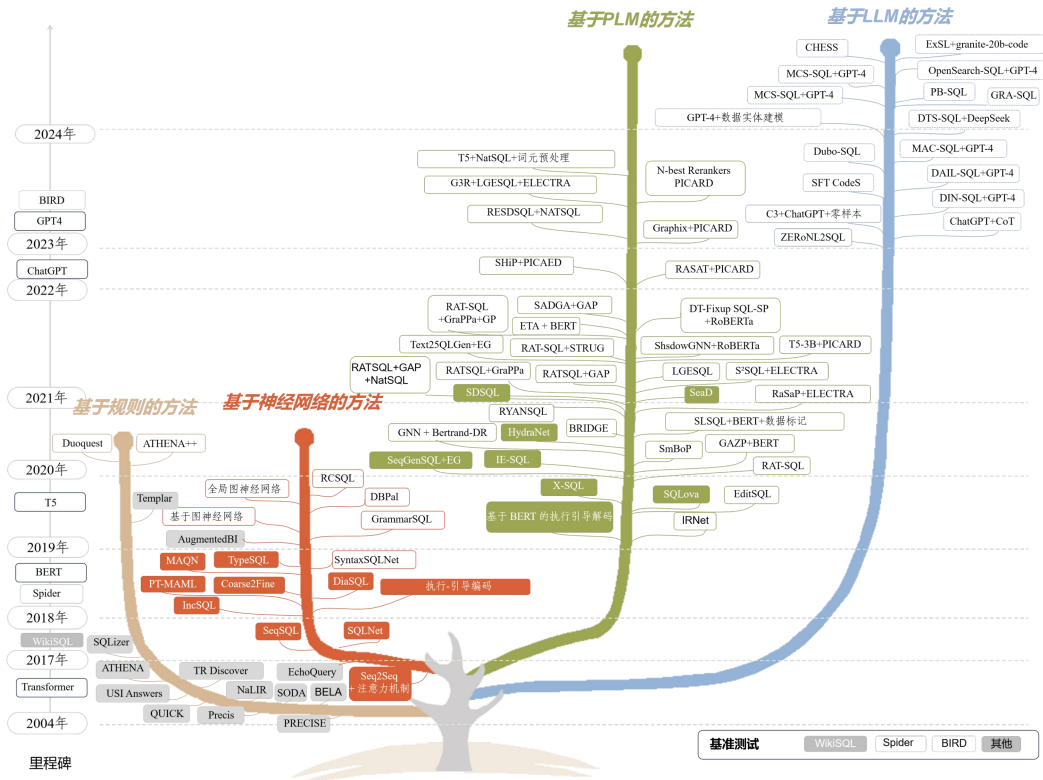


图 30 Text-to-SQL 发展历程

Text-to-SQL 的发展分为基于规则的 Text-to-SQL、基于深度学习模型的 Text-to-SQL、基于预训练模型（PLMs）的 Text-to-SQL 以及基于大型语言模型（LLMs）的 Text-to-SQL 四个阶段。传统的 Text-to-SQL 方法主要依赖于预定义规则、查询枚举或序列到序列模型。然而，这些方法在处理复杂的自然语言问题和 SQL 查询时，面临着可扩展性和泛化能力有限的困境。近年来，随着深度学习技术的不断发展，尤其是基于 Transformer 的模型（如 BERT）的出现，显著提升了 Text-to-SQL 解析的性能。这些模型通常需要大量的标注数据进行训练，并且在跨领域和复杂查询的处理上仍存在局限性。随着

大语言模型（LLMs）的出现，Text-to-SQL 迎来了新的突破。LLMs 如 GPT-4 和 GLM-130B，凭借其强大的语言理解和生成能力在 Text-to-SQL 任务中展现出了巨大的潜力。这些模型通过预训练学习大量语言知识和结构信息，能够在少量样本甚至零样本（zero-shot）的情况下生成准确的 SQL 查询。



来源：香港科技大学等

图 31 NL2SQL 方法概览

Text-to-SQL 解析主要的技术方法分为单轮文本到 SQL 解析和多轮文本到 SQL 解析两大类。单轮文本解析聚焦于独立的自然语言问题，将其转换为 SQL 查询，不涉及历史对话信息，代表性方法有 SQLova、RATSQL、SADGA、LGESQL 等。多轮文本解析是处理连

续的自然语言问题序列,结合上下文信息来准确生成当前问题的 SQL 查询,代表性方法有 EditSQL、CoSQL、CHASE 等<sup>20</sup>。

近年来 LLMs 在 Text-to-SQL 任务处理中也展现出巨大潜力,主要是通过提示工程 (Prompt Engineering) 来引导 LLMs 生成正确的 SQL 查询以及通过监督微调 (Supervised Fine-Tuning, SFT) 在特定任务数据集上进一步训练 LLMs,提升其在特定领域的性能。为了提升以 LLMs 为基础的 Text-to-SQL 的工作性能及准确性,诸多专家也在不断探索新的方法。Gao 等人于 2023 年提出的 DAIL-SQL 方法通过精心设计的提示工程,在 Spider 基准测试中取得了 86.6% 的执行准确率,显著优于之前的最佳方法。Zhou 等人提出了 DB-GPT-Hub 了标准测试套件<sup>21</sup>,专注于大模型微调,为 LLMs 在 text-to-SQL 任务中的研究和应用提供了重要的基准和工具<sup>22</sup>。

表 2 基于预训练模型 (PLM) 和大语言模型 (LLM) 的 NL2SQL 方法分类

来源: 香港科技大学等

类型	方法	基础模型	示例选择 (少样本)	模式链接	数据库内容	SQL生成策略			后处理策略
						多步	中间表示	解码策略	
大语言模型 提示	DIN-SQL	GPT-4	手动	√	×	分类分解	NatSQL	贪心搜索	自我纠错
	DAIL-SOL (with Self-Consistency)	GPT-4	相似度	×	×	×	×	贪心搜索	自一致性
	MAC-SOL	GPT-4	不适用	√	×	子问题分解器	×	贪心搜索	提炼器
	C3-SQL	GPT-3.5	不适用	√	×	×	×	贪心搜索	自一致性
	CodeS	StarCoder	相似度	√	√	×	×	集束搜索	Execution-Guided SQL Selector
	SFT CodeS	StarCoder	不适用	√	√	×	×	集束搜索	Execution-Guided SQL Selector
预训练语言模型 微调	RESDSOL + NatSQL	T5	不适用	√	√	结构解析	NatSQL	集束搜索	Execution-Guided SQL Selector
	Graphix + PICARD	T5	不适用	√	√	×	×	PICARD	×
	N-best Rerankers + PICARD	T5	不适用	√	√	×	×	PICARD	N-best Rerankers
	T5 + NatSQL + Token Preprocessing	T5	不适用	√	√	×	NatSQL	贪心搜索	×
	RASAT + PICARD	T5	不适用	√	√	×	×	PICARD	×
	SHiP + PICARD	T5	不适用	×	√	×	×	PICARD	×
	T5 + PICARD	T5	不适用	×	√	×	×	PICARD	×
	RATSOL + GAP + NatSQL	BART	不适用	√	√	×	NatSQL	×	×
BRIDGE v2	BERT	不适用	×	√	×	×	基于模式一致性的引导解码	×	

20 Qin, B., Hui, B., Wang, L., Yang, M., Li, J., Si, L., Huang, F., & Li, Y. (2023). A Survey on Text-to-SQL Parsing: Concepts, Methods, and Future Directions.

21 Gao, D., Wang, H., Li, Y., Sun, X., Qian, Y., Ding, B., & Zhou, J. (2023). Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation.

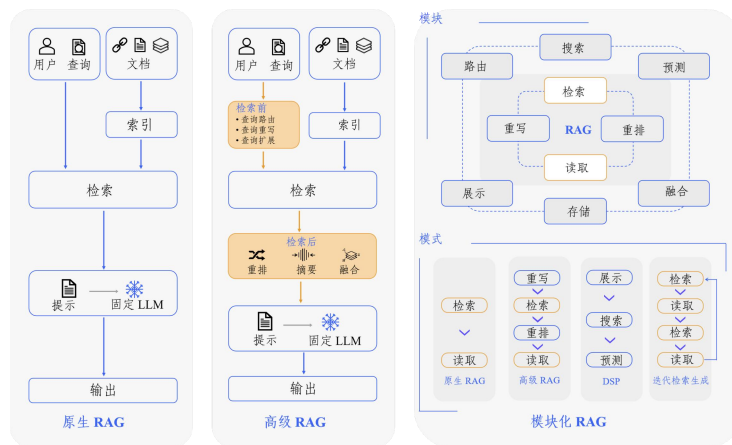
22 Zhou, F., Xue, S., Qi, D., Shi, W., Zhao, W., Wei, G., Zhang, C., & Chen, F. (2024). DB-GPT-Hub: Towards Open Benchmarking Text-to-SQL Empowered by Large Language Models.

尽管基于 LLMs 的模型在某些情况下表现出色，但在不同的使用场景中，没有一个单一的模型能够在所有情况下都取得最佳性能。在选择模型时需要考虑具体的使用场景和需求<sup>23</sup>。此外，随着图技术的不断发展，Text-to-SQL 技术延伸出了 Text-to-GQL 查询，即将自然语言查询直接翻译为图查询语句，以提高图数据的查询索引能力<sup>24</sup>。

### 1.1.3 RAG/Graph RAG 提升数据库查询准确性

在 Text-to-SQL 的实际应用中，面临诸多挑战，包括自然语言模糊性、业务上下文理解缺失、数据库类型语法规则复杂性、模型产生的幻觉以及查询复杂性等。为了应对这些挑战，研究者引入了检索增强生成（Retrieval-Augmented Generation, RAG）技术，尤其是解决大模型幻觉问题。

RAG 作为一种新兴技术，通过整合外部数据库的知识，有效解决了大型语言模型（LLMs）在特定领域或知识密集型任务中面临的幻觉、知识过时和推理过程不透明等挑战。该技术通过检索、生成和增强三个关键步骤，将 LLMs 的参数化知识与外部非参数化数据相结合，从而提升模型的准确性和可信度。



来源：同济大学等<sup>25</sup>

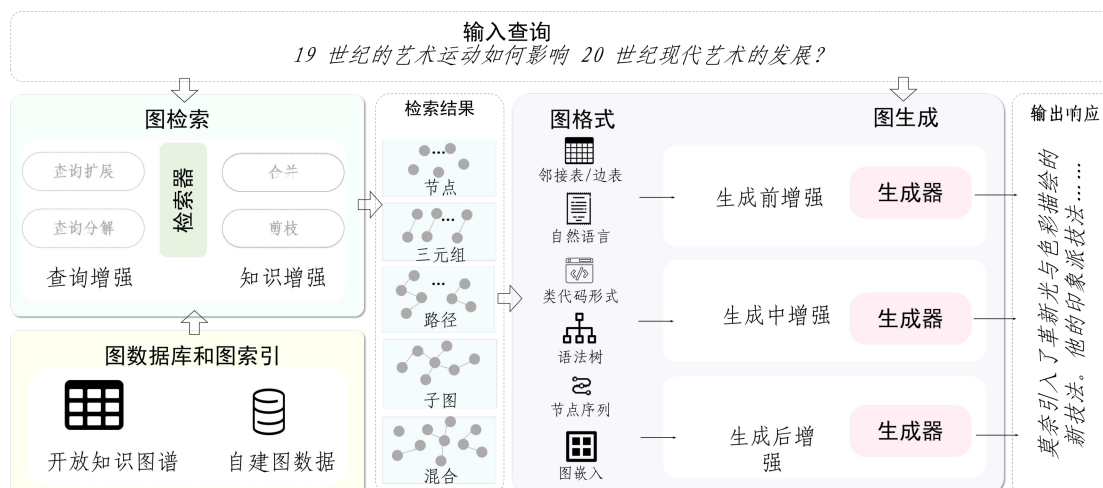
图 32 RAG 三种范式对比

23 Li, Boyan, et al. "The Dawn of Natural Language to SQL: Are We Fully Ready?." (2024).

24 <https://www.gqlstandards.org/>

25 Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. "Retrieval-Augmented Generation for Large Language Models: A Survey."

RAG 的研究范式不断发展演进，按照其特点分为三个阶段：基础 RAG（Naive RAG）、先进 RAG（Advanced RAG）以及模块化 RAG (Modular RAG)。Naive RAG 包括索引、检索和生成三个传统步骤，其局限性在于检索精度和召回率不足、生成内容可能出现幻觉或偏见，以及信息整合的挑战。Advanced RAG 则是在检索前和检索后引入优化策略，如查询优化、重新排序和上下文压缩等，以提高检索质量和生成内容的相关性。Modular RAG 通过引入新的功能模块（如搜索模块、内存模块等）和灵活的交互流程，增强了系统的适应性和灵活性，支持序列处理和端到端训练。



来源：北京大学等

图 33 GraphRAG 框架在问答任务中的概览

尽管 RAG 可以通过引入外部知识库来解决大模型幻觉等问题，但是在处理实体间复杂关系时仍存在不足。而 GraphRAG 的出现能够利用图数据库中的结构化信息，更准确地捕捉实体间的关系，减少文本冗余，提供更全面的信息以解决复杂任务。与传统 RAG 不同，GraphRAG 从预构建的图数据库中检索包含关系知识的图元素如节

点、三元组、路径或子图，能够利用图结构信息实现更精准和全面的检索<sup>26</sup>。

当前 RAG 技术发展仍具有一定的局限性，一是依赖于预构建的知识库难以处理动态更新的信息，二是当前大多局限于文本数据难以处理复杂多模态任务，三是在大规模知识图谱中检索效率和准确性可能下降。未来 RAG 技术将更加注重动态和自适应图的构建和更新，发展多模态 RAG 技术以处理复杂多媒体任务以及不断与 GNN、强化学习、微调等技术相融合，以提高在复杂任务中的处理性能等。

## 1.2 数据库技术支撑人工智能高效建模（DB for AI）

人工智能不仅促使数据库技术更加智能高效，数据库技术也在不断赋能人工智能发展。随着人工智能应用的持续快速发展，诸多数据库技术正在多方位满足人工智能对数据存储、检索和处理的要求，数据库领域不断创新也为人工智能发展提供强有力支撑。

### 1.2.1 向量数据库提升大模型检索效率

特征向量在多领域具有关键作用，如检索增强型大语言模型、电子商务、推荐系统及文档检索等。传统数据库管理系统并未针对特征向量特性进行优化，难以高效处理数十亿向量的查询需求，同时由于训练数据存在更新不及时、数据种类不足等问题，LLMs 在实际应用中存在幻觉问题。向量数据库（VDBMS）作为一种专为高维向量数据设计的存储与检索系统，为解决向量检索难题、LLMs 幻觉问题及知识库更新挑战提供了新路径。当前，向量数据库系统技术路线主要分为原生向量数据库产品、基于已有数据库产品进行扩展的向量数据库产品以及向量搜索引擎等。

---

26 Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. 2024. Graph Retrieval-Augmented Generation: A Survey. J. ACM 37, 4, Article 111 (September 2024), 41 pages.

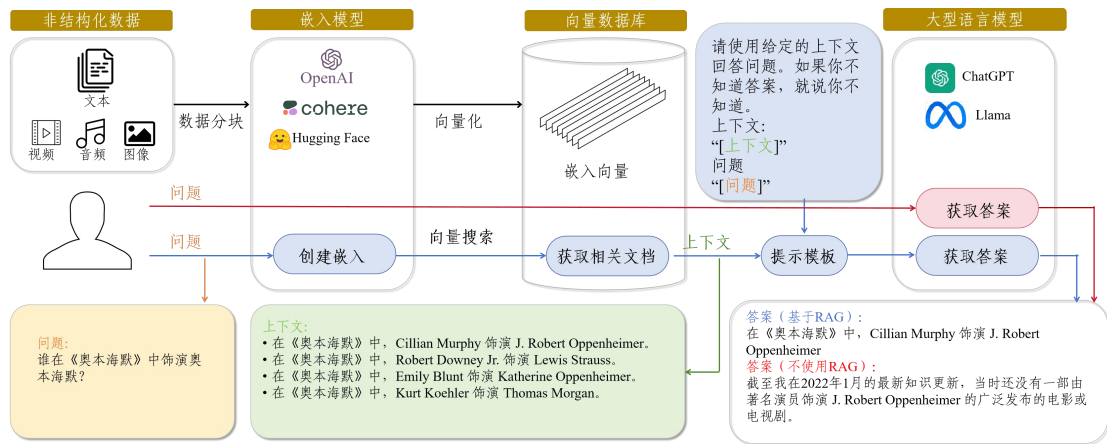
表 3 支持向量数据的主流各数据库对比

来源：卡内基梅隆大学等<sup>27</sup>

名称 (版本与年份)	支持的数据类型		支持的查询方式		向量索引	
	向量维度	类型	筛选	多向量	图索引	倒排索引
ChromaDB(2022)	1536	Vec.	✓		✓	
Manu(2022)	32768	Vec.+Ftx.	✓	✓	✓	✓
Milvus(2021)	32768	Vec.+Ftx.	✓	✓	✓	✓
Pinecone (2019)	20000	Vec.	✓	✓	✓	✓
Weaviate (2019)	65535	Vec.+Ftx.	✓	✓	✓	
Qdrant (2021)	4096	Vec.+Ftx.	✓	✓	✓	
Amazon OpenSearch(v2.9,2023)	16000	Frx.	✓	✓	✓	✓
ElasticSearch(v8.0,2022)	4096	Frx.	✓	✓	✓	✓
AnalyticDB-V(2020)	≥512	Rel.+Ftx.	✓		✓	✓
PostgreSQL-pgvector (2021)	2000	Rel.+Ftx.	✓	✓	✓	✓
MongoDB Atlas(v6.0,2023)	2048	NoSQL+Ftx.	✓		✓	
MyScale (2023)	1536	Rel.	✓		✓	

原生向量数据库如 Pinecone、Manu、腾讯云 VectorDB、拓尔思 TRS 海贝搜索数据库系统、华为云 GaussDB、百度智能云 VectorDB 等，主要能够针对高性能向量搜索应用。基于 NoSQL 或关系型系统进行扩展的向量数据库包括阿里云 AnalyticDB-V、火山引擎 ByteHouse、浪潮云岳向量数据库管理系统、PostgreSQL-pgvector 和 Vespa 等，能够适用于需要更复杂功能的应用。此外还有集成向量索引能够提供向量搜索功能的如 Apache Lucene、Elasticsearch 和 Meta Faiss 等产品。

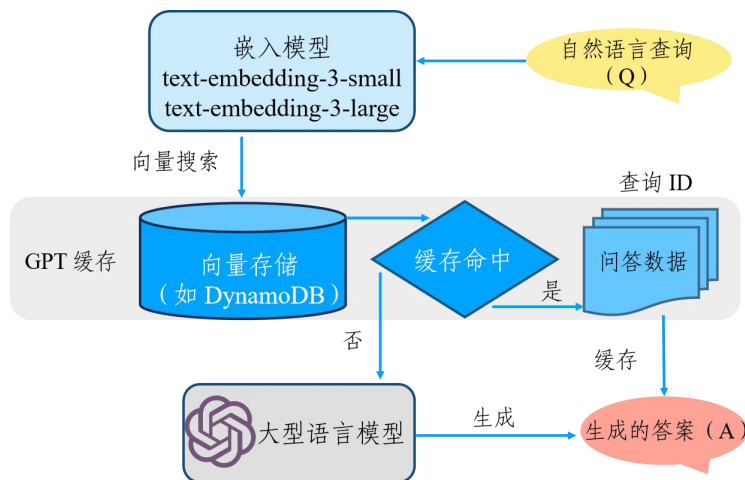
27 Jing, Zhi, Yongye Su, Yikun Han, et al. 2024. "When Large Language Models Meet Vector Databases: A Survey."



来源：卡内基梅隆大学等

图 34 RAG 在向量数据库中的框架流程

尽管大语言模型取得了显著成功，但是幻觉问题和计算资源有限的问题仍制约其发展。向量数据库能够在高维空间中表示现实世界的信息，有效管理和检索向量数据，很好地支撑大语言模型，优化计算资源，为其提供坚实的数据基础，主要通过以下三种方式。一是向量数据库可以作为大语言模型的外部知识库，支撑检索增强生成(RAG)，从而解决了大语言模型在集成和处理外部数据库中大量动态数据时所面临的挑战。二是向量数据库可以作为经济高效的语义缓存，显著降低数据操作成本。三是向量数据库可以作为大语言模型的可靠记忆，不断更新信息的存储解决方案，确保大语言模型的响应基于当前最相关的数据。



来源：卡内基梅隆大学等

图 35 使用向量数据库的 GPT 的语义缓存概览

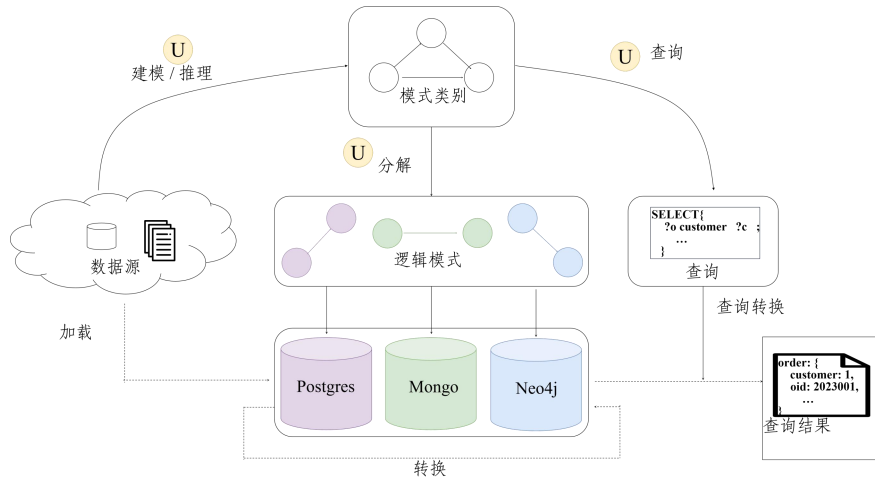
虽然向量数据库管理系统取得了显著进展，在人工智能技术发展中起到了重要作用，但仍面临查询延迟、安全与隐私等问题。诸多专家正在尝试通过多向量检索、增量检索、安全和隐私保护等技术推动向量数据库性能提升和广泛应用。

### 1.2.2 多模数据库赋能大模型多样化数据

非结构化和半结构化数据在人工智能应用中扮演着越来越重要的角色，如文本、图像、音频和视频等。传统的单一数据模型已逐渐难以应对复杂业务场景中多源、多结构数据的管理需求。多模数据库应运而生，能够在单一系统中同时支持多种数据模型（如关系型、文档型、图结构、键值存储、列式存储等）的数据库解决方案。通过统一的存储引擎、查询接口和事务管理机制，多模数据库在保证性能和一致性的同时，实现了对异构数据的原生支持，极大地简化了系统架构，提升了开发效率与数据处理能力。与传统多数据库系统相比，多模数据库有效规避了跨系统集成带来的维护负担和性能瓶颈，更适于处理跨模型联合查询和事务操作。在 DB-Engines 2025 年 5 月的排名中，前十的数据库管理系统有超过一半明确支持多模模型，在全球热度最高的 50 款数据库管理系统中，超过三分之二属于多模数据库类别，典型代表如 Elasticsearch、ArangoDB、OrientDB、MarkLogic、MongoDB 等。

尽管多模数据库展现出针对不同场景的应用能力，但是仍面临诸多挑战。一是数据模型差异性大，统一表示较为困难，关系型、文档型、图和时序等数据在结构、语义和约束机制上存在显著差异，难以构建通用的数据抽象层。二是查询语言与执行机制融合难度较高，多模数据使用 SQL、XQuery、Cypher 等不同查询语言，融合后执行计

划难以统一优化。三是事务一致性难以保障，不同模型的数据一致性需求与隔离级别差异较大，特别是图与文档模型往往不具备传统关系模型的 ACID 支持。



来源：查理大学等

图 36 统一抽象层表达异构数据结构示意图

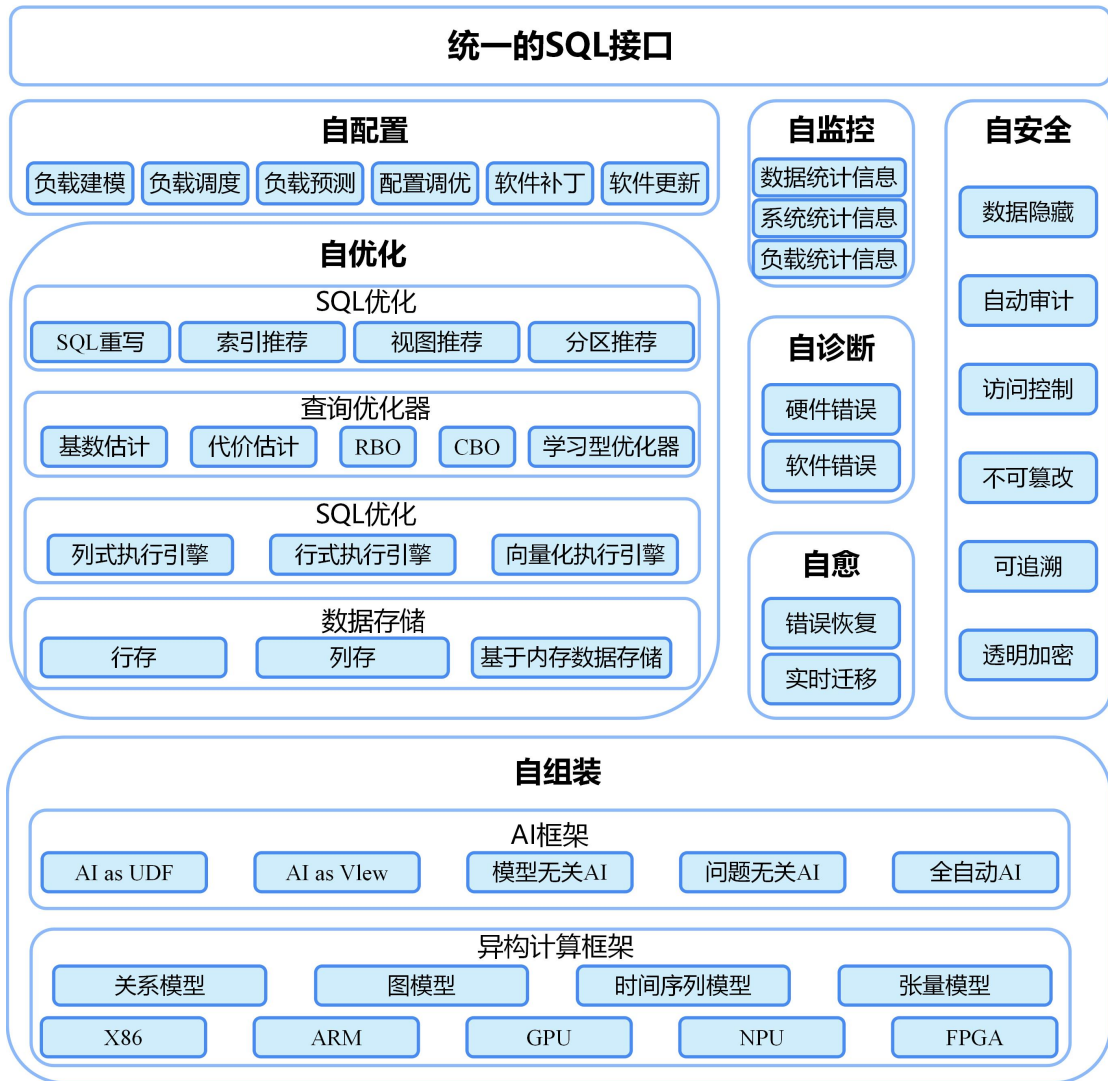
面对以上挑战，学术界与工业界正持续探索更具统一性与智能化的数据管理方式，多模数据库系统正朝构建语义统一的数据建模体系、统一的多模查询语言（MMQL）、引入数据库自治技术以及面向云环境建构演化以满足高并发与弹性扩展需求持续发展<sup>28</sup>。

## 2.AI 原生数据库解锁场景新可能

AI 时代，数据库系统面临三重挑战，一是传统的成本估算、连接顺序选择、参数调优等优化技术已无法满足大规模数据、多样化应用和多元化用户对高性能的要求，引入智能化技术能够使得数据库优化更智能。二是许多数据库应用需要使用 AI 算法，将 AI 算法嵌入数据库中，利用数据库技术加速 AI 算法并为数据库提供 AI 能力，但是传统数据库更多使用 CPU 等通用硬件，难以很好地利用 ARM、GPU、AI 芯片等新型硬件。AI 原生数据库将 AI 技术深度融合入数据库设计

<sup>28</sup> Jáchym Bártil. AI-Powered Orchestration of Multi-Model Data. VLDB 2024 Workshop: VLDB Ph.D. Workshop.

与功能中，实现数据处理的智能化、自动化，显著提升数据管理与分析的效率和价值。来自清华大学的专家学者提出一款原生支持人工智能的数据库系统，以应对人工智能时代对数据库带来的挑战<sup>29</sup>。



来源：清华大学

图 37 AI 原生数据库的架构

AI 原生数据库通过将 AI 结合到数据库的处理、运维和组装过程中使得数据库实现了自监控、自配置、自优化、自诊断、自愈、自安全和自组装的能力，并为人工智能和数据库服务提供了统一的调用接口。

29 李国良,周焯赫.轩辕:AI 原生数据库系统[J].软件学报,2020,31(3):831-844.

当前，AI 原生数据库依据智能程度依次分为五个阶段，AI 建议型数据库、AI 辅助性数据库、AI 强化型数据库、AI 自组装型数据库以及 AI 自设计型数据库。

第一阶段，AI 建议型数据库包括一个人工智能引擎，通过自动化建议，提供数据库的离线优化。这种外挂式的人工智能引擎与数据库松耦合。受到可用资源的限制，AI 建议型数据库主要从负载管理、SQL 优化、数据库监控与数据库安全四个方面提供辅助优化。

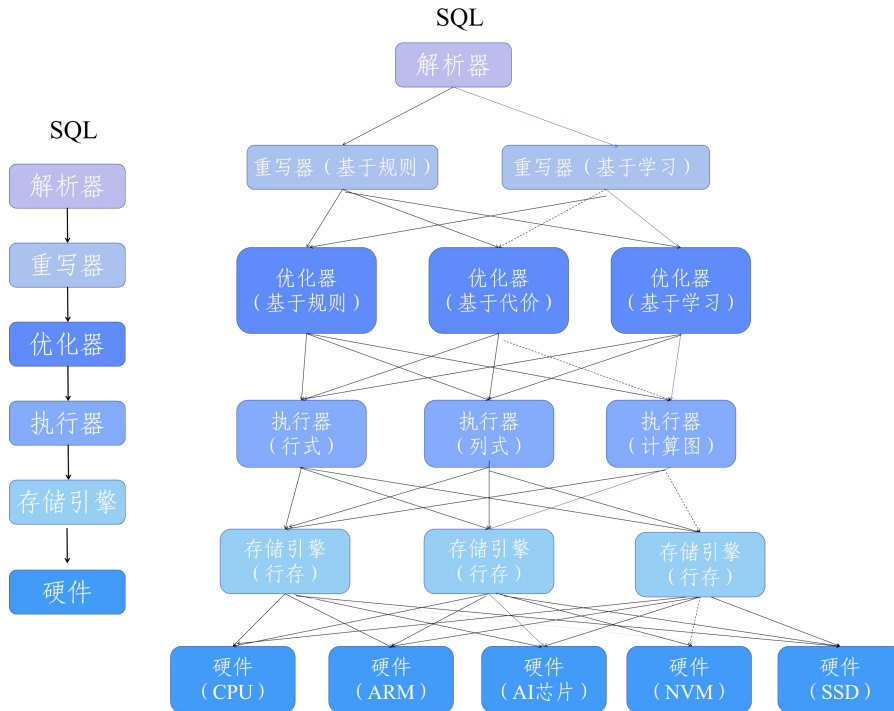
表 4 AI 原生数据库的五个阶段

来源：清华大学

阶段	特点	简介	举例
1	AI 建议型数据库	提供插件形式的 AI 引擎	<ul style="list-style-type: none"> <li>○ 负载管理（例如，负载调度）</li> <li>○ SQL 优化（例如，SQL 重写器、索引推荐）</li> <li>○ 数据库监视器（例如线下参数调优、系统统计）</li> <li>○ 数据库安全性（例如，自动审计/屏蔽）</li> </ul>
2	AI 辅助型数据库	提供基于数据库的 AI 引擎	<ul style="list-style-type: none"> <li>○ 自配置（例如，在线参数调优）</li> <li>○ 自优化（例如，SQL 优化、数据存储）</li> <li>○ 自监控（例如，监控系统状态）</li> <li>○ 自诊断（例如，发现硬件/软件问题）</li> <li>○ 自愈（例如，故障恢复，在线迁移）</li> <li>○ 自安全（例如，可回溯性、防信息泄露）</li> </ul>
3	AI 强化型数据库	提供统一的 AI、DB 引擎	<ul style="list-style-type: none"> <li>○ 基于学习的数据库组件                             <ul style="list-style-type: none"> <li>● 学习型查询重写器</li> <li>● 学习型代价估算器</li> <li>● 学习型优化器</li> <li>● 学习型执行器</li> <li>● 学习型存储引擎</li> <li>● 学习型索引</li> </ul> </li> <li>○ 声明型人工智能（UDF；视图；模型无关；问题无关；全自动）</li> </ul>
4	AI 自组装型数据库	提供异构处理架构	<ul style="list-style-type: none"> <li>○ 自组装</li> <li>○ 充分利用异构硬件（如 ARM、GPU、NPU）</li> </ul>
5	AI 自设计型数据库	基于 AI 的数据库生命周期	设计，编码，评估，监控和运维

第二阶段是 AI 辅助性数据库，它将 AI 引擎集成到数据库内核中，提供运行时优化。AI 工具如调优模型、工作负载调度、视图推荐等

可以合并到相应的数据库组件中。通过这种方式，人工智能被集成到数据库的工作过程中。人工智能辅助数据库的优势在于提供更细粒度的优化以及通过再内核中嵌入人工智能引擎，可以降低如通信成本等方面的开销。此外通过内置 AI 引擎，AI 辅助型数据库可以提供自配置、自优化、自监控、自诊断、自愈和自安全服务。



来源：清华大学

图 38 AI 组织型数据库流程图

第三阶段是 AI 增强型数据库，不仅用人工智能技术优化数据库设计，而且提供基于数据库内置的 AI 原生服务，一是通过人工智能算法对数据库组件进行增强，提供基于传统经验技术之外的替代策略；二是数据库内置 AI 服务，通过使用数据库技术降低人工智能门槛，一方面可以通过扩展 SQL 来支持 AI，另一方面可以通过数据库优化技术加速 AI 算法，例如索引、增量计算和共享计算；三是使用 AI 与 DB 混合引擎同时支持人工智能和数据库服务，通过查询优化器决定查询计划应发送给相应的数据库执行器或是相应的 AI 执行器，此

外，也需要设计新的模型既支持关系代数模型，又支持张量模型以使得用户可以利用统一的数据模型来支持人工智能和数据库服务。

第四阶段是 AI 自组装型数据库，不仅自动组装数据库组件来生成最适合给定场景的数据库，同时将不同任务调度到适合的硬件上。第五阶段是 AI 自设计型数据库，在这个阶段，数据库完全可以由人工智能设计，包括设计、编码、评估、监控和维护等各个阶段，将人工智能技术集成到数据库全生命周期中，使数据库和人工智能都获得最佳性能。

AI 原生数据库是一种深度融合人工智能技术与数据库系统的下一代数据管理平台，AI 原生数据库技术包括 AI4DB 和 DB4AI 两部分内容：其中 AI4DB 指数据库通过集成 AI 技术，提供自配置、自优化、自监控、自诊断、自愈、自安全和自组装等功能；DB4AI 指通过直接在数据库中运行向量搜索、RAG、AI 算法、机器学习等 AI 模型，无需数据迁移及服务集成，同时为 AI 应用提供优化支持，从而为 AI 应用搭建强大的基础设施保障。基于产业需求，中国信通院联合华为云、浪潮云、南大通用、电科金仓、极限科技和中兴通讯等多家企业编制了《AI 原生数据库技术要求》标准。《AI 原生数据库技术要求》标准覆盖基本能力、AI 数据存储与检索、AI 数据库服务能力、AI 算法底座、智能化自治管理能力、兼容能力、安全能力、扩展能力、高可用能力等 9 大能力域。

### 3.数据库智能体提供数据交互新模式

数据库智能体是数据库与人工智能深度融合的代表，正逐步改变人们与数据库交互的方式，整合了自然语言处理、机器学习、知识图

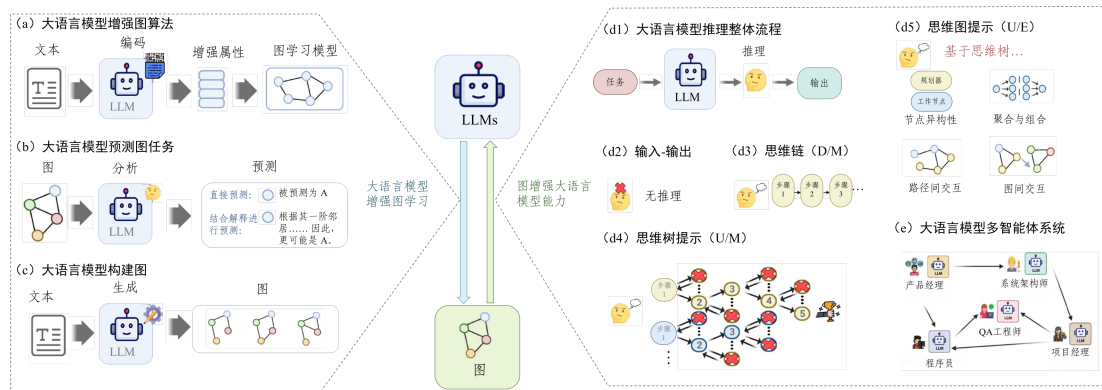
谱和多模态数据处理等多种技术，能够理解用户意图并高效地与数据库进行交互，为用户提供智能化、个性化的数据服务。

表 5 DB-GPT 与其他工具的比较

来源：蚂蚁集团等

	LangChain	LlamaIndex	PrivateGPT	ChatDB	DB-GPT	
系统组件	多智能体框架	✓	✓	✗	✗	✓
	支持多种大语言模型	✓	✓	✗	✓	✓
	多数据源的检索增强生成	✓	✓	✗	✗	✓
	智能体工作流表达语言	✗	✗	✗	✗	✓
	微调的文本转 SQL 模型	✗	✓	✗	✗	✓
数据交互功能	微调的文本转 SQL 模型	✓	✓	✗	✓	✓
	Chat2DB / Chat2Data / Chat2Excel	✓	✓	✗	✓	✓
	数据隐私与安全	✗	✗	✓	✗	✓
	多语言交互	✗	✗	✗	✓	✓
	生成式数据分析	✗	✗	✗	✗	✓

随着分布式、多模态等数据库技术的不断发展和人工智能技术的崛起，数据库正朝着智能化方向迈进。数据库智能体技术主要基于自然语言处理与数据库交互、机器学习优化数据库应用，以及多模态数据处理技术。数据库智能体可采用单智能体或多智能体架构。单智能体负责完成所有任务，结构简单所以处理复杂任务的能力有限。多智能体架构则由多个智能体协同工作，每个智能体负责特定的任务，如规划、执行、监控等，通过不同智能体协作，高效地完成复杂任务以实现数据库自动化安装配置、数据库性能优化及数据智能分析等工作。



来源：蚂蚁集团等

图 39 大语言模型（LLMs）与图相互增强的总体框架: (a)-(c)为大语言模型增强图学习的三种途径。(d)-(e)为图结构增强大语言模型推理的技术。技术名称后的括号表示图的类型。D、U、M 和 E 分别代表有向图、无向图、同构图和异构图。

近年来，以 ChatDB<sup>30</sup>、DB-GPT<sup>31</sup>、Chat2graph 为代表的数据库智能体也在持续涌现，以协助用户完成复杂任务。ChatDB 框架包含输入处理、Chain-of-memory 和回复总结三个部分，将数据库作为 LLMs 的符号记忆来增强其能力。DB-GPT 是集成 LLMs 的 Python 库，包括多智能体框架、Agentic Workflow Expression Language（AWEL）和 Service-oriented Multi-model Management Framework（SMMF）三个部分，支持自动提示生成、特定数据库模型微调以及模型设计与预训练。Chat2Graph 主要是面向图数据库的多智能体系统，通过已有大模型或智能体能力帮助用户降低图数据库的使用门槛。



来源：蚂蚁集团等

图 40 DB-GPT 系统设计

30 Hu, C., Fu, J., Du, C., Luo, S., Zhao, J., & Zhao, H. (2023). ChatDB: Augmenting LLMs with databases as their symbolic memory.

31 Siqiao Xue, Danrui Qi, Caigao Jiang, Fangyin Cheng, Keting Chen, Zhiping Zhang, Hongyang Zhang, Ganglin Wei, Wang Zhao, Fan Zhou, Hong Yi, Shaodong Liu, Hongjun Yang, Faqiang Chen. Demonstration of DB-GPT: Next Generation Data Interaction System Empowered by Large Language Models. PVLDB, 17(12): 4365 - 4368, 2024

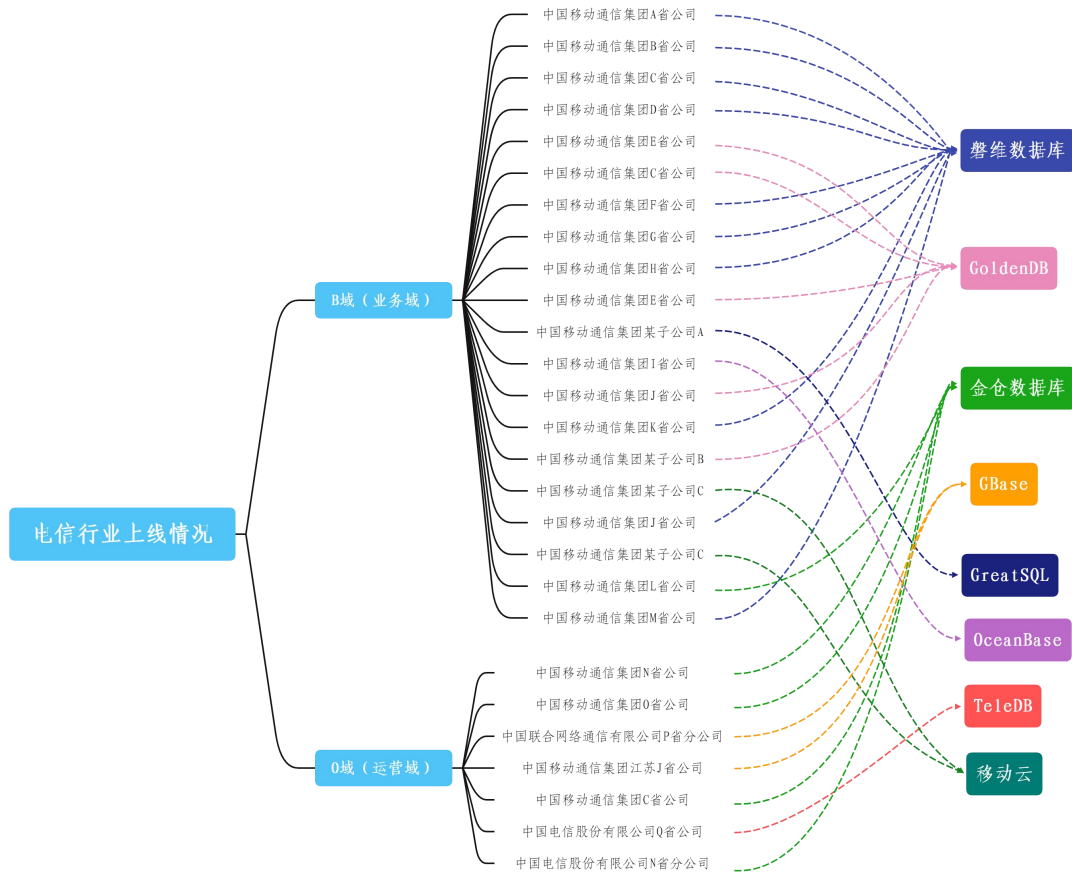
此外，为了更好地使 LLMs 与外部数据库、API 和工具进行高效交互，Anthropic 提出一种专为大语言模型和多智能体系统设计的标准化交互框架 MCP (Model Context Protocol)，旨在解决 LLMs 在复杂应用场景中上下文共享与协作的问题。GaussDB、TDSQL、PolarDB、OceanBase、IotDB、TDengine、KaiwuDB 等数据库产品已具备基于“数据库+MCP+LLMs”技术方案快速搭建智能体应用的能力。2025 年 7 月 4 日，腾讯云数据分析智能体 TCDataAgent 在“NL2SQL 全球最难榜单” BIRD-Bench 中获得全球第三、国内第一的优异成绩，超越 IBM、Meta 等众多国际厂商，其引入的数据库约束验证机制，能够自动识别并修正 SQL 中潜在结构性或语义性错误，相关研究论文已被 2025 年 VLDB 会议接收。

当前，数据库智能体仍处于发展早期，还未能够进行规模化落地应用，在数据同步与一致性、大规模数据处理以及只是表示与推理能力等方面仍需要不断优化。未来，数据库智能体将深化语义理解，更好地处理复杂和模糊的自然语言表达，形成自适应智能体，更好适应不同行业应用中的差异，以满足不同场景与用户需求为数据库技术发展带来更多便利与创新。

### 三、数据库行业应用情况综述

#### 1. 国产数据库在电信行业得到广泛应用

随着 5G/6G 网络的普及、物联网和云计算等技术发展，电信行业数据规模持续增长，业务日趋复杂，数据库系统面临响应速度、并发处理和智能化能力的多重挑战。

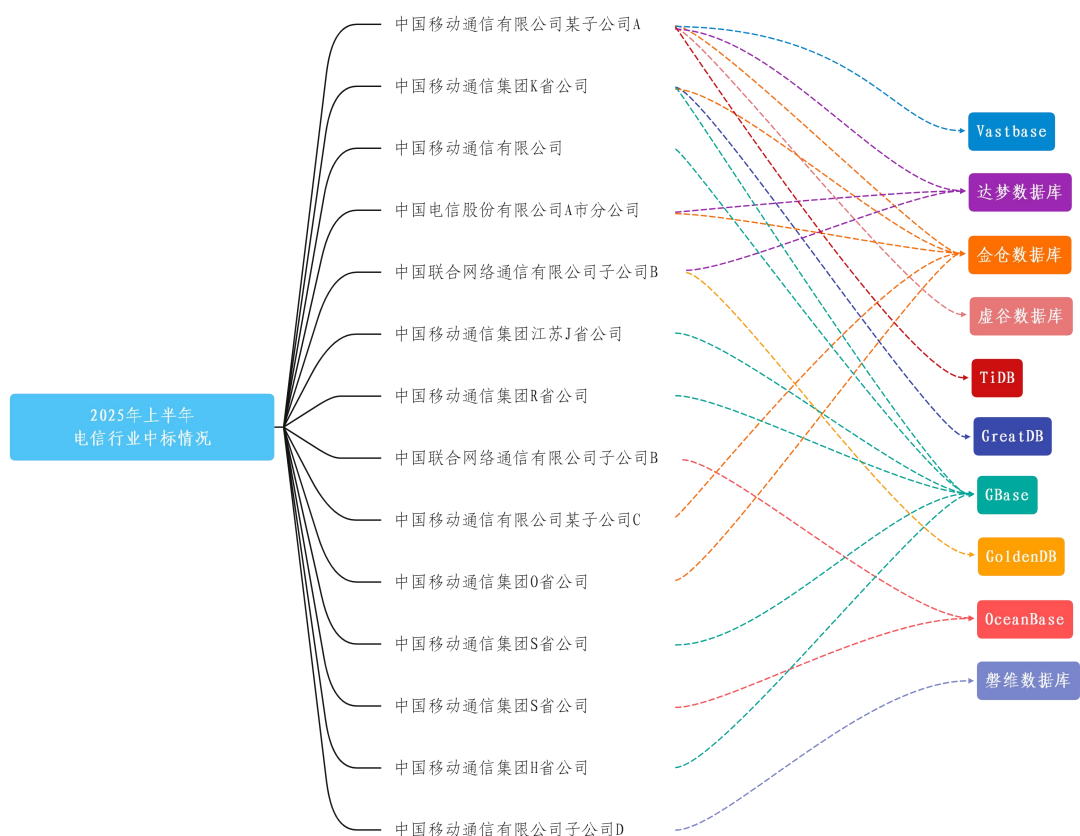


来源：CCSA TC601 根据公开信息整理

图 41 2024-2025 年电信行业上线数据库情况统计

2024 年至 2025 年上半年，电信行业在运营域(O域)和 B 域(业务域)上线了诸多国产数据库。在用户业务管理方面，海量用户账户信息、话单数据(CDR)、套餐规则与计费记录构成电信系统的数据核心。此前，传统关系型数据库如 Oracle、IBM Db2 等是运营商采用的主要数据库产品，随着数据规模不断扩大，应用场景愈加丰富，系

统面临高昂的运维成本与性能瓶颈。为了更好地保障业务稳定运行，运营商逐步引入以 GoldenDB、OceanBase、磐维数据库、TeleDB、GBase、金仓数据库、TiDB 等为代表的产品，以应对高并发、高可用、横向扩展及多中心部署等复杂业务需求。在网络运维方面，由于日志与告警数据需实时处理，时序数据库与图数据库有效支撑了设备监控与网络拓扑分析。在客户服务方面，面对结构化与非结构化数据的融合需求，多模数据库被用于支持用户画像、智能推荐等场景，实现智能化营销与服务响应。



来源：CCSA TC601 根据公开信息整理

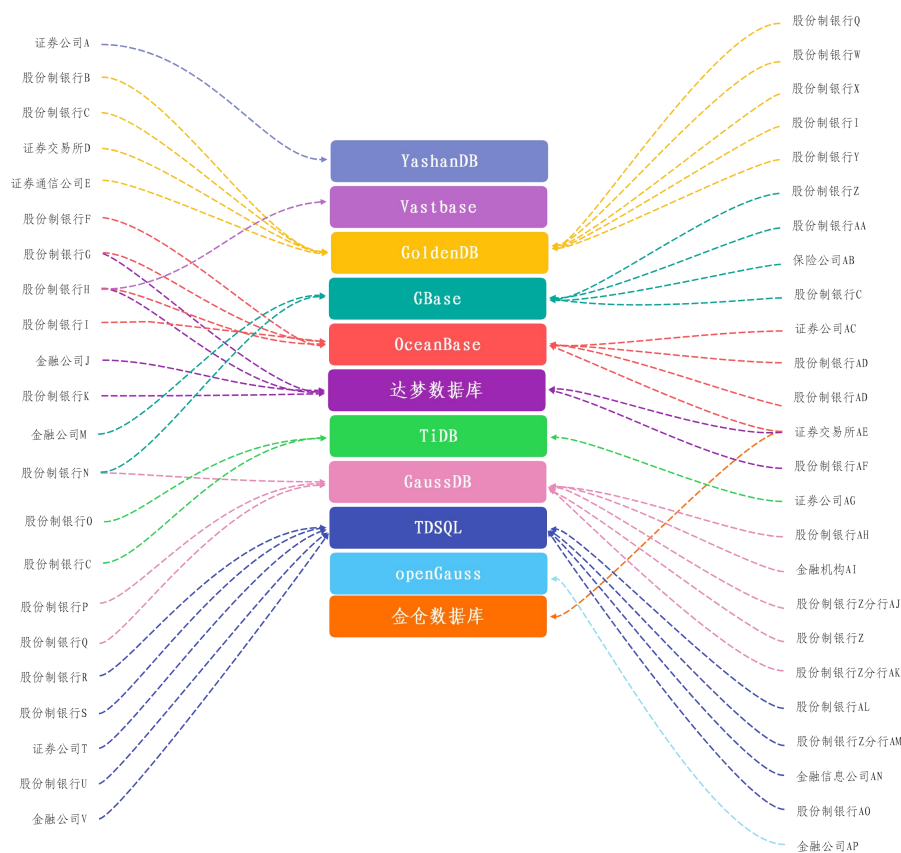
图 42 2025 年上半年电信行业部分国产数据库中标情况

数据库技术在电信行业正经历深刻变革，从集中式向分布式演进以满足高并发与多中心部署需求，从单一数据模型向多模融合扩展从而应对结构多样的业务场景，从传统静态存储向智能推理升级用以支

撑实时分析与自主决策。作为通信系统的核心支撑平台，数据库系统正成为电信行业实现业务敏捷化、网络智能化、服务个性化和管理自主化的关键基础，构筑起新一代“数字能力底座”，为行业高质量发展提供坚实保障。

## 2.数据库技术驱动智能金融核心变革

随着金融行业数字化转型的深化，海量数据处理、高并发交易支持及自主可控等需求，正推动数据库技术以前所未有的速度革新。许多金融机构每日数据增量超 100GB，传统集中式数据库难以应对庞大的存储与计算压力。同时，实时交易、风控校验等关键场景对高并发处理能力需求也日益迫切。在多重挑战下，分布式架构、国产化数据库及智能化运维成为行业应对的关键。



来源：CCSA TC601 根据公开信息整理

图 43 2025 年 1-5 月金融行业国产数据库部分中标/投产情况

2024 年大数据“星河”案例数据库专项入选的案例中，中信信托基于海光 CPU 与银河麒麟系统，采用南大通用 GBase 8a MPP Cluster 分布式分析型数据库，构建横跨资金端与资产端的一体化平台，压缩数据处理时间、有效扩大数据处理规模、保障数据安全。中信信托还依托数据库构建起普惠金融数据集市，整合多源数据形成统一模型，更好为业务赋能。邮储银行依托鲲鹏硬件，以 GaussDB 分布式数据库为核心打造统一查询系统，通过 DRS 数据迁移工具实现在线平滑扩容，解决了多年以来的历史数据查询问题，构建起大型同城双活集群。此外，智能化运维也成为技术创新趋势，智能预警，自动化故障预测等技术也在金融行业得到广泛应用。

数据库在金融行业中作为“数字能力底座”的组成部分，具有重要的战略价值。金融行业数据库应用正从“可用”向“好用”跃迁，不仅能够满足日常数据处理需求，更能助力通过数据释放业务价值。随着 AI 与数据库深度融合，金融行业将构建更敏捷、智能、安全的数据库体系，为业务创新提供更强支撑，推动金融科技、智慧服务等多元场景纵深发展。

### 3.数据库技术支持能源行业数字化转型

在“双碳”目标引领下，能源行业正加速数字化智能化转型，数据库技术在此过程中扮演着关键角色，助力能源行业提升数据管理效能、推动业务创新与可持续发展。能源行业数据来源广泛，涉及结构化、半结构化与非结构化数据，数据库需支持多源异构数据的高效接入与整合。

能源行业产生的数据类型繁多，包括生产数据、设备运行数据、地质勘探数据、图像视频等数据。多模数据库能够高效存储和解锁不

同类型数据。此外对于地质勘探数据、管道布局数据等，支持空间索引和地理信息系统功能的空间数据库能够实现对地理位置数据的精确查询和分析。在油气勘探和开发过程中，空间数据存储引擎帮助工程师快速定位地质构造和油气藏，提高勘探效率和成功率<sup>32</sup>。能源生产过程中，如电力系统的发电、输电、配电环节，以及石油、天然气的开采和输送等，会产生大量实时数据，如设备运行状态、温度、压力、流量等。数据库通过与传感器、监控设备等的连接，能够实时采集这些数据，并快速存储到数据库中，为后续的分析 and 决策提供数据基础。

2024 年大数据“星河”案例数据库专项中，青海黄河智慧能源科技有限公司的光储大数据平台充分利用星环科技 Timelyre 企业级分布式时序数据库有效满足光伏、能源领域海量时序数据的高效存储与分析。太原向明智控科技有限公司以浪潮 KaiwuDB 多模数据库为数据底座结合向明智控煤矿综采自动化数据分析平台，支撑采煤工作面数据应用需求。能源行业数据库应用正朝着技术创新与融合、应用场景拓展与优化、性能与安全需求提升、新兴技术拓展的方向发展，以更好满足能源行业的数字化转型需求。

#### 4.数据库技术助力交通运输行业平安出行

近年来，随着交通运输业飞速发展和用户出行需求上升，车辆制造、路网调度、计价售票等多个环节产生海量多元数据以及对系统实时响应和安全稳定运行需求，为交通运输业数据管理带来了新的挑战。

在道路运输业，车辆数据来源多样且结构复杂，从车辆制造的零部件数据到车辆运行数据，再到用户驾驶习惯以及智能网联道路等信

---

<sup>32</sup> 董文, 张俊峰, 刘俊, 等. 国产数据库在能源数字化转型中的创新应用研究[J]. 信息通信技术与政策, 2024, 50(10): 68-74.

息，数据来源广泛复杂，数据量增长迅速。同时，由于数据涉及大量用户个人信息，系统在信息安全保护方面也面临着极高的要求。数据库技术在应对以上挑战中起到了重要的作用。分布式数据库提高数据处理效率与可扩展性、多模数据库应对异构数据统一管理挑战、密态数据库保障信息安全，以及库内机器学习预测分析需求从而提升自动驾驶系统安全性等，诸多数据库技术不断创新发展，为道路运输业提供更强大的数据处理能力，促进产业链上下游数据协同，推动行业高效、安全与智能发展。

在航空运输业，民航运价计算系统是航空收益管理的核心引擎，承担着实时处理全球航班运价规则、动态计算最优票价的关键任务。为了更好适应市场需求，航空公司运价产品不断丰富，也为运价计算系统带来了诸多挑战。系统面临着数据规模不断增大、业务响应延迟要求严苛，硬件资源依赖过重等问题。此外，航班调度需实时分析天气、客流以及机组资源等问题，需要引入 AI 算法助力优化航线网络，提升客座服务精准度。传统集中式架构由于扩展性不足等问题，难以支撑海量数据实时分析与高并发需求。为了应对以上挑战，民航运价系统引入新一代分布式对象数据库 MMDB，通过数据分片、跨节点传输优化与一致性保障等技术提升了运价计算效率。同时，民航业也通过多架构一体化、多工作负载一体化以及多模一体化等技术，以保障复杂场景下安全的运行、关键业务链路毫秒级的响应以及航班调度数据秒级分析等业务场景，以一体化数据库为支点，构筑民航新一代“数据智能底座”。

## 四、总结与展望

人工智能正成为推动行业转型升级的重要驱动力，我国积极探索人工智能与各行业深度融合，为产业升级和经济高质量发展赋能。数据库作为数据存储与计算的核心基础软件，在信息产业生态中扮演着至关重要的角色。近年来，我国数据库产业加速发展，市场竞争激烈，在技术创新和市场竞争能力方面不断迈向新高度，已正式进入产业质变期，为我国数字经济的高质量发展和全球科技竞争格局的重塑提供了有力支撑。

**产业层面看**，全球数据库市场增速平稳，竞争进入深水区；我国数据库产业规模稳步增长，市场格局头部效应明显；产业生态日益完善，数据库产品愈发成熟；人才生态逐渐丰富，开源和学术影响力逐步扩大，多模数据处理、“AI+DB”等成为学术研究的热点方向。

**技术层面看**，数据库技术架构持续向着一体化、云原生等方向不断革新，推动数据库部署和管理模式向灵活高效的方向发展，融合型数据库成为发展方向。数据库技术迈入 AI 原生时代，数据库开发、运维模式变得更加精准智能，为数据库创新应用提供了更加丰富的场景和广阔的空间。

**应用层面看**，我国数据库应用创新场景日益丰富，应用规模持续扩大，自主可控程度不断加深，高效赋能传统产业转型升级，促进数字经济和实体经济深度融合，助力新质生产力加快形成和发展。

“聚沙成塔，积水成渊”。近年来，通过汇聚各方力量，我国数据库产业稳中求进。未来，我国数据库产业发展需继续以技术创新、人才培养、生态建设为核心驱动力，促进实现关键技术自主可控，推动产业链上下游协同发展，为产业长远发展奠定坚实基础。



**大数据技术标准推进委员会**

**地址：** 北京市海淀区花园北路 52 号

**邮编：** 100191

**邮箱：** TC601@CCSA.org.cn

**网址：** [www.tc601.com](http://www.tc601.com)

