

证券研究报告 | 行业深度报告
计算机 | 计算机应用

筚路蓝缕，星火燎原

——大数据系列（一）：数据库深度复盘与展望

刘玉萍

liuyuping@cmschina.com.cn

S1090518120002

“慧博资讯”专业的投资研究大数据分享平台

点击进入  <http://www.hibor.com.cn>

CMS  招商证券

2022.2.24

核心要点

数据库作为信息化的核心环节，是底层硬件基础资源与上层应用之间的重要支撑。随着海量数据的涌现和分析能力需求增长，预计未来数据库存储量将出现严重错配，需求迫切性将显著提升。从信息产业链角度来看，数据库作为信息化系统承上启下的关键环节，向下调用硬件基础资源，向上为应用软件提供重要数据支撑，是信息化中间的核心环节；从国家政策角度来看，数字经济作为国家“十四五”建设的重中之重，数据作为新型关键生产要素，是推动数字经济发展的关键一环，同时，国内数据库长期由海外厂商主导，在国内数据库行业快速发展下，国产数据库将迎来重要发展机遇。

不断顺应市场需求变化，数据库行业正持续进化。本篇报告回顾了数据库70年发展历程，从数据模型逻辑、技术架构、需求功能、部署方式、商业模式、存储介质、治理模式等七大维度对数据库发展脉络进行详细梳理。我们发现，由于数据规模、读写需求量等指数级增长，为避免数据冗余与架构的复杂化，叠加分布式架构技术日趋成熟，将多技术、多模型融合成为数据库发展的重要方向。如将NoSQL非关系型的分布式拓展性与关系型数据模型融合诞生的NewSQL，保留了关系型数据逻辑的同时，大幅提升了系统的灵活拓展能力；数据模型方面，将结构化、半结构化、非结构化等多种数据融合的多模数据库正快速发展；需求功能方面，将OLTP与OLAP融合的HTAP数据库正逐渐成为企业首选等均显示出行业技术融合的发展特点。

在国内大力发展数字经济建设的大背景下，数据库行业规模将保持快速增长。根据信通院统计，2020年中国数据库市场规模为35亿美元（约合240.9亿元人民币），预计2025年市场总规模将达到688亿元，CAGR为23.4%。基于对数据库发展趋势与国家政策导向的分析，我们认为，未来数据库市场增量需求主要来自于以下三方面：

- 1) 国产化进程正逐步加速：**通过对金融行业国产化数据库建设案例梳理发现，国产数据库已经从边缘系统逐步实现核心系统国产化，显示出国产数据库产品功能、性能水平不断增强。预计随着行业信创逐步落地，对国产数据库需求将进一步提速。
- 2) 分布式架构凭借在经济性、安全自主、灵活性、可伸缩性等方面的优势，逐步实现对集中式架构替代：**目前，互联网、金融、电信等行业分布式升级进展较快，其中金融、电信行业已有分布式数据库在股份制银行、城商行、二三线城市运营商等实现了核心业务系统的改造落地，同时，墨天轮前十大热门数据库均为分布式架构，其应用场景涵盖各行业，显示出国产分布式数据库已经趋于成熟，正成为各行业数据库建设重要选择。
- 3) 随着工业互联网、互联网创新型业务、车联网等应用的快速发展，非关系型数据库应用将加速落地：**非关系型数据库应用场景不断丰富，如图数据库能够支撑社交网络、金融反欺诈等互联网新型应用场景的关联分析业务；时序数据库可以高效存储和快速处理海量时序大数据，在工业互联网快速发展下，传感器与设备产生的大量实时数据，将推动时序数据库应用快速发展等。

风险提示：技术创新不及预期；政策落地不及预期；行业竞争加剧风险。

□ 数据库：信息系统承上启下的关键环节

- ✓ 数据库是信息化核心环节
- ✓ 七大维度全面梳理数据库发展历程
- ✓ 发展特点：顺应市场需求，行业持续进化

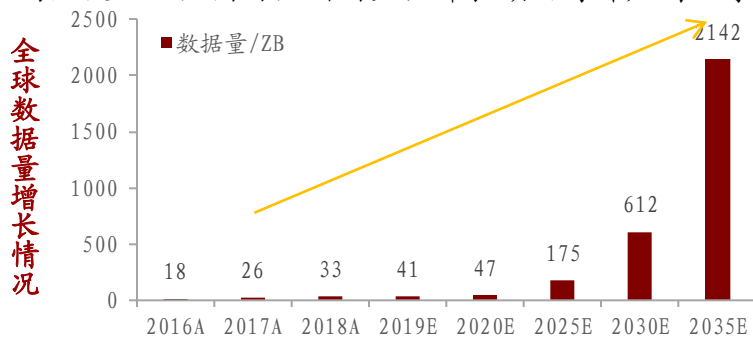
□ 国内数据库市场发展趋势分析

□ 国产数据库厂商详细梳理

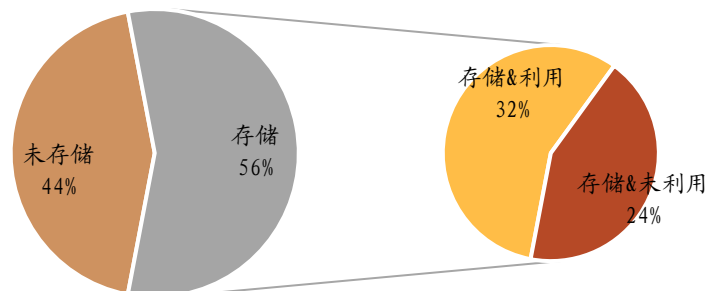
1.1 数据库是信息化核心环节

数据库作为信息化的核心环节，是底层硬件基础资源与上层应用之间的重要支撑。根据Statista、IDC与Seagate的统计调研，全球数据量将在2035年达到2142ZB，然而企业运营中的数据仅有32%能被存储利用，海量数据的涌现和分析能力需求提升，数据库存储量将严重不足，未来对数据库需求迫切性有望显著提升：

- 从信息产业链角度分析：数据库作为信息系统承上启下的关键环节，向下调用硬件基础资源，向上为应用软件提供重要数据支撑，是信息化中间的核心环节；
- 从国家政策角度分析：数字经济作为国家“十四五”建设的重中之重，数据作为新型关键生产要素，是推动数字经济发展的关键一环，同时，国内数据库长期由海外厂商主导，在国内数据库行业快速发展下，国产数据库将迎来重要发展机遇。

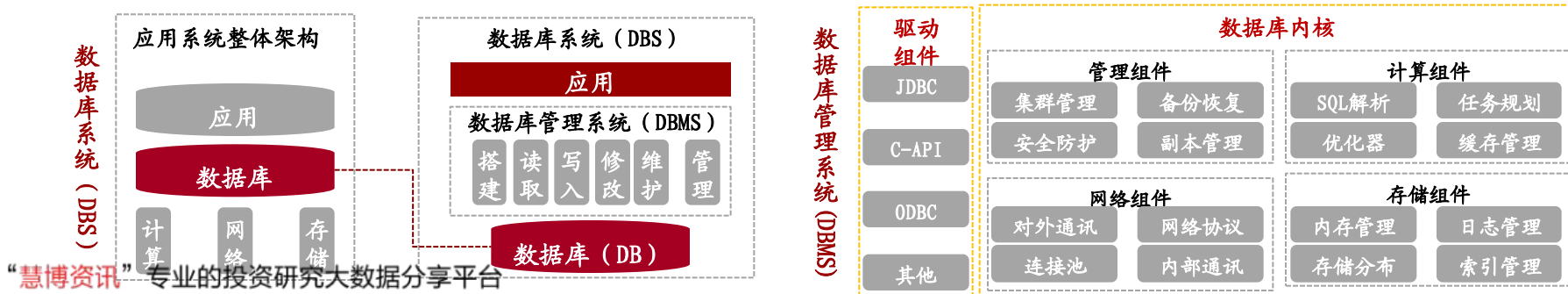


全球数据存储利用情况



广义的数据库通常指数据库系统，其包含数据库与数据库管理系统两部分，向上通过数据库管理系统支撑应用引擎，向下以数据库承接文本、图像、声音等数据源，调动计算、网络、存储等基础资源。

1. 数据库（Database，DB）是按照一定的逻辑结构组织、存储、管理数据的大容量电子文件柜；
2. 数据库管理系统（Database Management System，DBMS）是对数据库进行统一管理和控制的大型软件，主要由内核组件集和驱动组件构成，其中内核组件集按照功能模块划分为管理组件、存储组件、计算组件和网络组件。



1.2.1 从七大维度对数据库发展历程进行详细梳理

回顾数据库70年发展历程，分布式、云计算、人工智能等创新技术和基础设施的发展陆续为演化注入新活力。本章将从数据模型逻辑、技术架构、需求功能、部署方式、存储介质、商业模式、数据库治理模式等七大维度对数据库发展脉络进行详细阐述，理清行业演化逻辑与发展趋势。

数据库发展历程

	-1970s	1970s	1980s	1990s	2000s	2010s-
互联网背景	基础技术阶段	基础协议阶段		基础应用阶段	Web 1.0	Web 2.0-3.0
数据模型	层状、网状	关系型			NoSQL	NewSQL
技术架构	单机		集中式			分布式
需求功能	联机事务处理过程OLTP			联机分析处理过程OLAP		HTAP融合
部署方式	本地部署					云部署
存储介质	磁盘数据库DRDB（数据保存在磁盘上，内存缓存磁盘内容作为临时存储）				内存数据库MMDB	
商业模式	商业			开源		
治理模式	自适应			自调优		自治
代表厂商 代表产品	IBM	IBM Oracle	微软 SqlServer Teradata	Mysql PostgreSQL	EMC、HP Greenplum Hadoop	AWS、google 阿里云 腾讯云

1.2.2 数据库发展特点：顺应市场需求，行业持续进化

顺应市场需求变化，数据库行业正持续进化：

- 随着大数据时代的到来，数据规模和逻辑关系复杂度进一步提升，其中非结构化数据的激增使得非关系型数据库迅速发展；
- 随着5G下的云计算、物联网等快速发展，对于数据分析的需求、响应速度、存储容量扩展、算力提升等均提出更高的要求，使得分布式数据库得到快速发展；
- 信息化水平快速提升下，企业业务规模快速扩张，数据库的运维和管理成为新的难题，数据库云部署模式与借助AI自治成为重要发展方向。

数据库发展里程碑事件梳理

标志事件		技术演变路径			
时间	事件	数据模型	技术架构	部署方式	治理模式
1964	世界上第一个数据库系统IDS(Integrated Data Storage, 集成数据存储)诞生	网状	集中式	本地部署	人工治理
1968	世界上第一个层次数据库系统IMS(Information Management System)诞生于IBM	网状 & 层次			
1970	IBM 的研究员Edgar F.Codd提出关系数据模型	关系型			
1974	IBM 推出关系模型的原型系统System R				
1974	IBM 的Ray Boyce和Don Chamberlin提出了SQL				
1979	第一个商业数据库版本 Oracle Release 1诞生，提供商用SQL				
1996	MySQL发布，开源数据库开始发展	非关系型			
2000s	非关系型数据库NoSQL开始盛行，主要包括4种类型：文档数据库、列簇式数据库、键值数据库和图数据库。				
2003-2006	Google发表奠定了业界大规模分布式存储系统的理论基础的三篇论文：Google File System、Google MapReduce以及Google BigTable	关系&非关系并行	分布式		AI原生、自治成为新方向
2011	结合SQL和NoSQL的NewSQL概念出现				
2011	微软发布 Windows Azure Storage				
2012	全球第一个 Global Database——Google Spanner 论文发表				
2016	Amazon 发布了代表性的云数据库 Aurora				
2018	Oracle发布自治数据库				
2019	谷歌联合麻省理工、布朗大学的研究人员共同推出了新型数据库系统SageDB			云端部署逐步兴起	
2019	华为公司发布了全球首款AI原生(AI-Native)数据库——GaussDB				

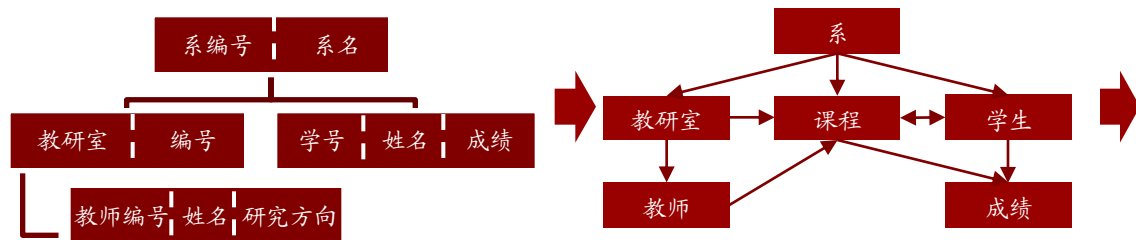
“慧博资讯”专业的投资研究大数据分享平台

点击进入<http://www.hibor.com.cn>

1.3.1 数据模型分类演化：NewSQL、多模引领新时代

数据模型的演化本质由底层数据规模、逻辑关系、类型等驱动，其路径是：层次->网状->关系->关系、非关系并行。

层次、网状、关系模型演化



学号	姓名	性别	年龄	专业
A0	赵亦	女	18	计算机
A1	钱尔	男	17	自动化
A2	孙珊	女	19	经济
A3	李思	男	21	统计

层次模型	网状模型	关系模型
采用树型结构来表示实体及其实体间的联系	采用有向图结构表示实体以及实体之间的联系	关系模型的数据结构是一张规范化的二维表

在当前关系、非关系并行时代，新型NewSQL与多模数据库快速发展。传统SQL受限于其拓展性，在大数据发展下逐渐面临瓶颈，而NewSQL提供了与NoSQL相同的可扩展性，而且仍基于关系模型，保留了极其成熟的SQL作为查询语言，从而保证了数据的强一致性；同时，基于分布式架构，对接多种数据模型引擎，避免搭建多种类数据库的多模数据库快速发展。

	层次 (1960s)	网状 (1960s)	关系 (1970)	关系、非关系并行		
				NoSQL (2000s)	NewSQL (2010s)	多模 (2010s)
优点	对层次逻辑的表达清晰	对于层次和非层次结构的事物都能自然地模拟	二维表逻辑更易理解；SQL语言操作方便；支持ACID，数据一致、稳定	易扩展；性能高；模型灵活；高可用	兼具NoSQL对海量数据的存储管理能力和传统关系数据库的ACID、SQL等特性	基于一套系统处理多种数据模型
局限	只能处理一对多的关系	只能处理少量数据；不易操作和定位	能很好地处理表格型数据，却无法处理文本、图像、视频等复杂类型	NoSQL不支持SQL查询，不支持数据的强一致事务处理，只能保证数据的弱一致性	不具有SQL系统的通用性，NewSQL对传统SQL系统的丰富工具仅仅提供部分访问	产品成熟度低
代表	IBM的IMS	基于CODASYL组织DBTG报告的DBTG系统	1) 商用数据库：Oracle、SQL Server、DB2； 2) 开源数据库：MySQL、PostgreSQL	1) 键值数据库：Redis； 2) 宽列数据库：HBase； 3) 文档型数据库：MongoDB； 4) 图形数据库：Neo4j	Google的Spanner模型、Amazon的Aurora	蚂蚁的OceanBase、阿里巴巴的Lindorm
扩展	纵向扩展	结构复杂，不易扩展	纵向扩展	水平扩展	水平扩展	水平扩展
性能	较差	较差	海量数据读写性能差	高性能处理海量数据	高性能处理海量数据	高性能处理海量数据
ACID	\	\	支持	不支持	支持	支持
OLTP	支持	支持	支持	效果差	支持	支持
OLAP	支持	支持	轻量查询	支持	支持	支持

数据库模型发展路径

“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

资料来源：墨天轮、CSDN、阿里云开发者社区、招商证券

1.3.1 数据库霸主——关系型数据库

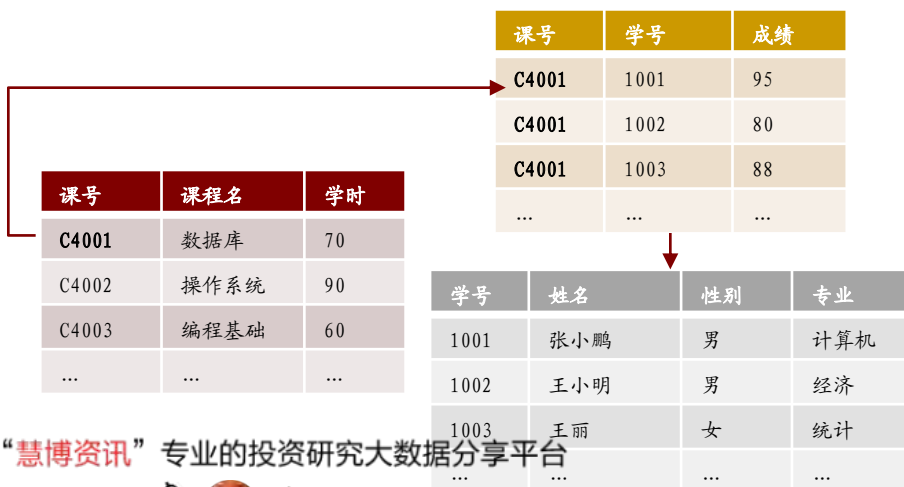
数据库前期的演化史本质上是关系型数据库的演化，关系型数据库已统治市场超过50年。关系型数据库以二维表形式存储实体和关系，具有易理解、易使用、易维护，自1970年诞生以来，逐步成为业界的霸主。时至今日，关系型数据库不断融入分布式、云原生、AI等新兴技术持续进化，仍是数据库的主流，据DB-Engines热度排名（截至2022/02），前四大数据库Oracle、MySQL、Microsoft、PostgreSQL均为关系型数据库，且分数遥遥领先。

关系型数据库的数据更新过程完全遵循事务ACID原则，保障数据库正确一致。如果将事务看做上层应用访问数据库的钥匙，ACID原则通过对钥匙的限制保证了数据库在断电、重启、并发读写等极端情况下的正确一致，保证数据库的合理性。

关系型数据库严格遵循ACID原则

ACID原则	详细
原子性（Atomicity）	原子性保证事务执行的完整性，只能全部执行或者失败回滚全部不执行
一致性（Consistency）	一致性通过约束保证数据和业务逻辑的一致性
隔离性（Isolation）	隔离性通过锁机制创造事务孤岛，保证并发事务互不干扰
持久性（Durability）	持久性保证已执行事务的效果可以永久保存

关系型数据库模型逻辑示意图



DB-Engines流行数据库排名前十（截至2022/02）

排名	数据库	主类型	分数
1	Oracle	关系型	1256.83
2	MySQL	关系型	1214.68
3	Microsoft SQL Server	关系型	949.05
4	PostgreSQL	关系型	609.38
5	MongoDB	文档型	488.64
6	Redis	键值型	175.80
7	IBM Db2	关系型	162.88
8	Elasticsearch	搜索引擎	162.29
9	Microsoft Access	关系型	131.26
10	SQLite	关系型	128.37

SQL语言逻辑简单易懂，使其快速流行，进一步巩固了关系型数据库的霸主地位。SQL是关系型数据库的标准语言，相比C、Java、Python等编程语言，SQL贴合口语的设计使得其代码复杂度大大简化、便于数据分析人员上手，同时嵌入式的设计也保证了SQL可以被其他编程语言轻松调用，据TIOBE统计，SQL在近20年稳居TOP10编程语言榜单。关系型数据库的二维表数据存储模式为关系型数据库的流行普及提供了环境基础；而SQL语言的流行也反哺了关系型数据库的推广，两者相辅相成，共赢于数据库市场。

1.3.1 星星之火——非关系型数据库NoSQL

随着互联网业务规模快速增长，拥有灵活拓展与高读写并发能力NoSQL数据库得到快速发展。由于传统的关系型数据库在保证数据强一致与系统可用的前提下，其水平扩展能力受到限制；非关系型数据库NoSQL则是基于分布式存储原理，放弃了传统的关系模型，在一致性与可用性之间寻求平衡，从而保证了系统地灵活拓展与高读写并发能力。随着社交媒体等互联网业务快速发展，数据量激增，传统关系型数据库性能逐渐达到瓶颈，因此，以保证拓展性与高并发性能为前提地NoSQL数据库得到快速发展。

分布式系统的CAP原则

CAP原则	一致性 (Consistency)	可用性 (Availability)	分区容错性 (Partition-Tolerance)
任何分布式系统必须在一致性与可用性之间寻求平衡	数据更新操作后，多个数据副本保持一致	系统在部分异常情况下仍可以正常服务	节点被划分为多个区域，区域内可以通信，区域之间无法通信

权衡一致性与可用性的BASE原则，确立了NoSQL数据库标准，为其发展提供了明确指向。BASE原则定义了NoSQL下的数据最终一致性，使得NoSQL开发更专注于海量数据分析和高并发负载，奠定了其更擅长处理数据规模高速增长在互联网等应用场景。

分布式系统的BASE原则

BASE原则	基本可用 (Basically Available)	软状态 (Soft state)	最终一致性 (Eventual consistency)
	数据库故障时，通过损失部分可用功能或者降低响应速度的方式保证核心功能可用	允许数据存在中间状态：不同节点的数据副本同步过程存在不一致性	无事务更新操作、故障发生时，经过延迟后，数据库最终的数据达到一致

非结构化数据存储和分析需求的涌现，进一步推动了NoSQL的发展。为应对不同种类数据的存储利用需求，NoSQL模式自由，不提供规范化，分类演化出键值、宽列、文档、图等不同的数据存储模型。据DB-Engines对数据库流程序度的统计，MongoDB、Redis和ES 3个NoSQL已经冲入TOP10榜单。此外，我国非关系型数据库主要以基于开源数据库如Redis、InfluxDB、CouchDB等产品进行二次开发为主，其产品已经步入世界前列，如浙江智史、阿里云TSDB、欧若数网Nebula Graph、百度HugeGraph、华为云GraphBase等均已在DB-Engines榜上有名。

NoSQL数据库分类

典型NoSQL数据库分类				
	键值数据库	宽列数据库	文档数据库	图数据库
特点	以键值对的方式存储数据，键是唯一标识符	数据存储在记录中，行键唯一标识记录中的列，一行包含大量动态列	以文档格式（JSON、BSON、XML或YAML）存储和查询数据	存储节点、边、标签和属性
缺点	牺牲数据结构	牺牲存储空间	牺牲安全性和可靠性	较难适应分布式架构
优点	提高读写速度	解决数据压缩、过滤问题	架构灵活、横向可拓展	敏捷、可拓展、高性能
场景	web应用程序和会话、缓存、购物车	时间序列、历史记录	内容管理、APP、目录、日志文件	社交网络、知识图谱、搜索引擎
代表	Redis、Oracle BDB	HBASE、Riak	MongoDB、CouchDB	Neo4J、InfoGrid

1.3.2 技术架构分类演化：分布式架构逐渐成熟

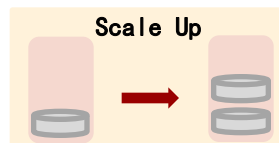
为了应对数据的海量增长并追求更低的扩容成本，数据库由单机、集中式向分布式架构快速发展。随着信息化水平快速提升，对数据库的存储、读写并发、扩容要求更高，集中式架构的“Scale Up”纵向扩容机制面临硬件性能瓶颈，因此“Scale Out”式的横向扩容分布式架构成为新趋势，通过网络将物理分散的数据库单元连接为逻辑上的统一整体。从集中式到分布式的演进降低了对硬件性能的要求，使得企业数据库拓展成本大幅降低。

分布式架构逐步由分库分表中间件向原生分布式发展，以应对不同应用场景。分库分表中间件早期主要以电商、社交起家的互联网公司自研为主，但由于其对复杂查询事务的支持较差，难以应对业务瞬时高峰需求，因此更加便捷、可靠的原生分布式数据库应运而生。按照具体实现方式可以分为两类：以Greenplum为代表的主流MPP架构和以Cassandra、HBase为代表的分区分片架构。

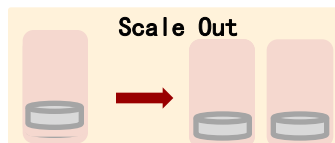
集中式、分库分表中间件、原生分布式对比

	集中式架构	分库分表中间件	原生分布式	
			MPP架构	分区分片
特点	集中式数据库是从多个点访问的单个位置的信息集合	数据存储在不同单机数据库上，通过中间件记录原数据	在硬件结构上“Shared Nothing”，不共享底层硬件资源，将任务并行分散到多个服务器节点上，待任务完成后，将各自部分的结果汇总得到最终结果	底层存储基于具备分区或分片功能的分布式文件系统，提高存储系统上限
优点	生态完善、适配性强	扩展能力高、效率高	保留ACID特性、数据从集中式到分布式的迁移更加简便；通过数据冗余提高集群可靠性	
局限	成本高，无法横向拓展	性能差、无法高并发	短板效应	数据同步迟滞
场景	传统金融业务系统	互联网公司	BI系统、决策支持系统	交易日志、状态跟踪、物联网

系统拓展方式：scale up VS. scale out

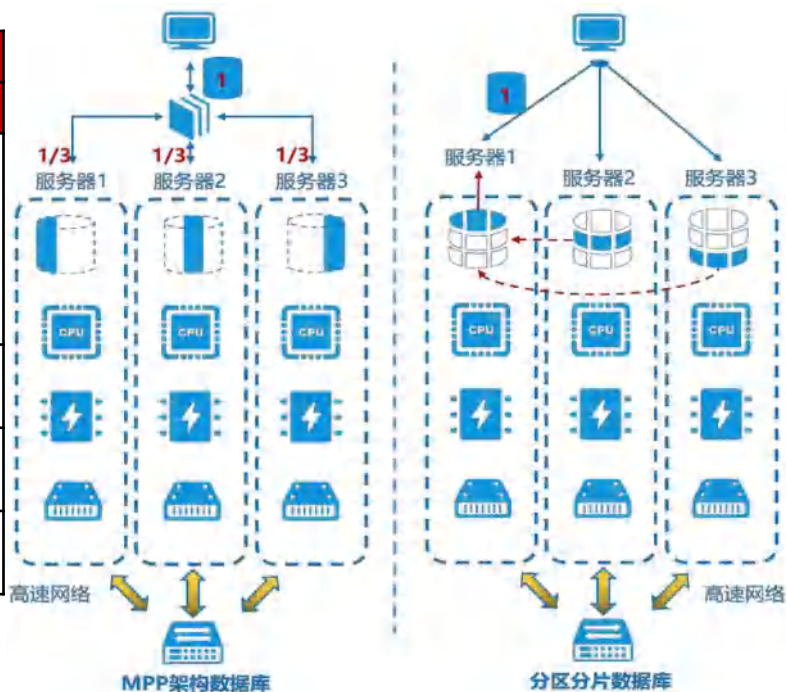


纵向扩展：利用现有存储系统，通过不断增加存储容量来满足数据增长的需求。
(容易达到性能瓶颈)



横向扩展：以节点（包含容量、处理能力、I/O带宽）为单位增加，系统的容量增长和性能扩展同时进行。
(按需扩容，扩展性强)

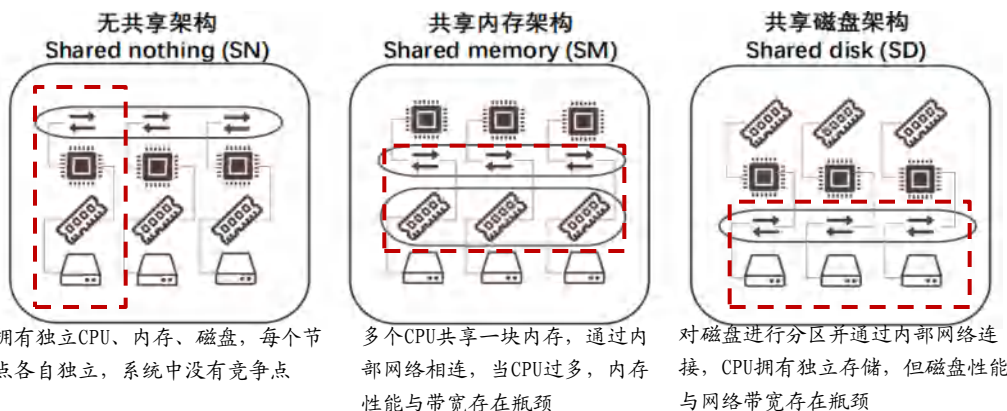
MPP与分区分片架构示意图



1.3.2 技术架构分类演化：AI技术赋能分布式架构

无共享架构凭借实现高可用、带宽要求低、易横向扩展和远程部署四大优势，逐步成为分布式主流架构。分布式架构由共享磁盘、共享内存架构，逐步向纯无共享架构发展，在无共享架构中，每个处理器私有内存和磁盘空间，利用网络通信，“Scale Out”式横向增加处理器和相应的内存、磁盘，从而避免事务对内存访问和网络带宽的竞争，提高处理性能。国产方面，无共享分布式架构已经趋于成熟，如PingCAP的TiDB、华为的GaussDB、蚂蚁的Oceanbase、达梦的达梦+、阿里云的PolarDB、腾讯云TDSQL、南大通用的GBase、人大金仓的KingBase、中兴通讯的Golden DB等分布式数据库均为MPP无共享架构。

架构	详细	优势	劣势	典型代表
无共享架构	每个CPU都有一个专用存储区和专用磁盘空间，两个CPU无法访问相同的磁盘空间，并且CPU之间的通信通过网络连接	无共享可以保证数据库的稳定增长，当事务数量继续增加时，添加额外的CPU和主内存可以确保每个事务的处理时间保持不变；且由于各自独立，性能不受硬件上限与网络带宽影响	架构设计较为复杂	DB2 DPF Hadoop
共享内存架构	多个CPU共享同一块内存，并且这些CPU通过内部通信机制（内部互连网络）进行通信	通过增加节点可以提高并行处理能力，扩展更便捷	共享内存和共享磁盘的基本问题是干扰，当添加更多CPU时，由于内存访问和网络带宽的竞争加剧，系统速度变慢，且存在性能瓶颈	PostgreSQL
共享磁盘架构	每个CPU使用其自己的专用存储区，并通过内部通信机制直接访问所有磁盘系统			Oracle RAC



无共享架构在可用性、拓展性等方面具有明显优势

	高可用性	带宽要求	横向扩展	远程部署	并行控制	灾备恢复	设计易度	负载均衡
无共享架构	强	强	强	强	中	中	弱	弱
共享内存架构	弱	弱	弱	弱	中	强	中	强
共享磁盘架构	中	中	中	中	弱	弱	中	中

AI有望解决分布式架构划区难、运维难等痛点，成为新的发展方向：分布式数据库最核心的任务是根据合理的维度划分数据，然而面对分布键不明显，难以sharding的政企类数据，如何找到完美的划分维度、避免跨节点事务成为最大的难题。随着AI技术的发展，通过算法自学习数据模型以智能推荐分布键的技术成为新的研发方向。同时，分布式架构将模型复杂化，运维难度的大幅提升也迫切需要智能诊断、智能调优、智能运维等技术的发展。华为的GaussDB是业界首个将人工智能技术融入分布式数据库全生命周期，实现自运维、自管理、自调优、故障自诊断和自愈的数据库，目前已部署在金融、消费、游戏等行业，发展前景广阔。

慧博资讯 专业的投资研究大数据分享平台

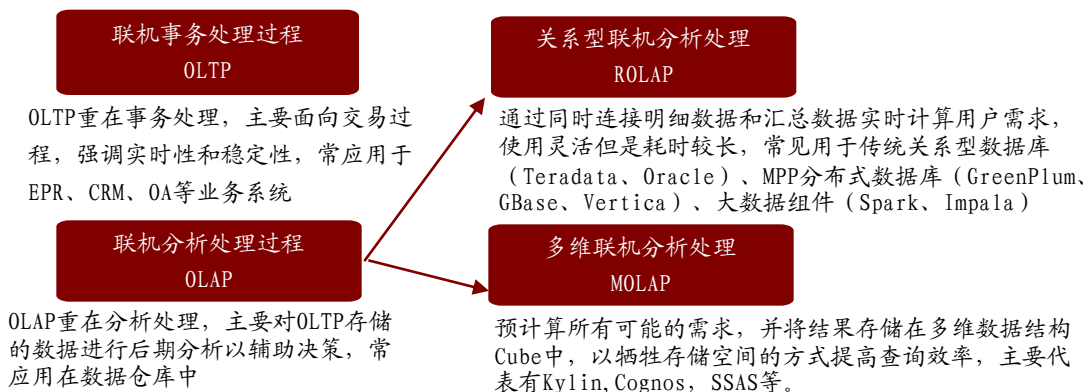
点击进入 <http://www.hibor.com.cn>

资料来源：沙利文市场报告、信通院、招商证券

1.3.3 需求功能分类演化：HTAP一站式解决方案

从OLTP到OLAP，分析功能引入对数据库要求进一步提升：

OLTP、OLAP特点对比



	目的	数据处理	标准化	数据量	存储介质	特点	场景
OLTP	处理订单、合同等业务	增删改查	是	小	传统DBMS	方便创建、维护	关注实时数据
OLAP	业务支持决策	数据分析	否	大	数据仓库	适合创建视图分析	分析长期数据规律走势

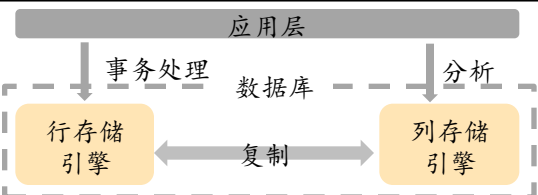
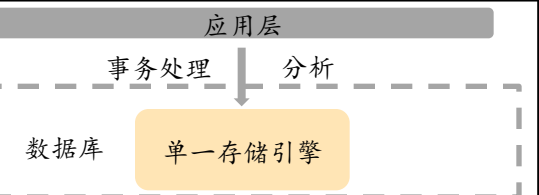
分布式架构的经济性、高并发性能等优势，使得OLTP与OLAP融合的HTAP逐步成为主流：

	OLTP和OLAP独立（-1990s）	OLTP+OLAP（1990s-2010s）	HTAP混合事务/分析（2010s-）
原因	上世纪90年代以前，数据存量小、增量小、事务处理为主、分析型需求少	随着互联网发展和数据规模增大，分析型需求涌现，解耦成为大势所趋，企业往往通过ETL过程将OLTP结果数据同步到OLAP系统数据库再加以分析，维护成本大大提高，数据同步延时、纰漏、冗余不可避免	随着分布式架构的出现，存储扩容成本大大降低，且高并发性能使得OLTP与OLAP并行成为可能
特点	OLTP和OLAP基于同一数据库实现	通过ETL将OLTP结果同步到OLAP系统数据库加以分析	2014年Gartner定义HTAP：打破OLTP和OLAP的壁垒，基于一套数据同时支持实时交互和数据分析需求
优缺	无法处理海量数据	1) 运维成本过高； 2) 同步延时、纰漏频出	基于分布式架构，支持弹性扩容，按需扩展吞吐，轻松应对高并发、海量数据
架构			

1.3.3 需求功能分类演化：HTAP一站式解决方案

需求功能决定数据库逻辑架构，逻辑架构决定访问性能。为实现海量数据存储、在线高并发，并同时支持OLTP和OLAP的功能，HTAP应运而生，按照具体架构可分为3类：

HTAP不同架构解析

	不同系统（OLTP+OLAP）	同一系统	
		不同存储结构	相同存储结构
特点	运用应用软件实现OLTP和OLAP的松耦合，底层共享存储来缩短数据同步时间	基于分布式，不同存储结构分别处理事务和分析，遵循Raft等协议在引擎间复制数据	使用单一存储引擎进行事务处理和分析，性能较差
架构	 通过ETL过程将OLTP结果库同步至OLAP数据库，OLTP和OLAP共享统一存储，在应用层实现混合处理，对外整体呈现HTAP能力	 在数据库层实现HTAP：行存储引擎负责OLTP、列存储引擎负责OLAP	 在最底层实现HTAP，仅通过单一存储引擎同时支持OLTP和OLAP
优劣	同步延迟、事务分析延迟	目前最主流的HTAP架构	不成熟、性能较差
代表	SAP、HANA Vora	HybridDB、TiDB、HANA、dashDB	Hive、Impala、Kudu

国产HTAP+分布式数据库热度持续高涨。根据墨天轮排名（截至2022/02），目前国内前六大数据库均为HTAP+分布式架构，包括PingCAP的TiDB、华为openGauss、蚂蚁OceanBase、达梦的达梦+、华为GaussDB、阿里PolarDB。其中排名第一的PingCAP基于Google的Spanner模型和Stanford的Raft算法研发TiDB数据库，借助底层数据同步及行列透明转换技术，将面向联机交易的行存引擎与面向实时分析的列存引擎融合改造为行列混合数据架构。在HTAP技术浪潮中，TiDB数据库凭借水平线性扩展、强一致分布式事务、故障自恢复的金融级高可用、真正跨数据中心多活等核心特性拔得头筹。

HTAP相较于单OLTP、OLAP数据库优势明显

	响应速度	吞吐量	并发访问	事务大小	数据访问	索引规模	实时性	物化视图
OLTP	强	弱	中	弱	弱	中	强	弱
OLAP	中	中	弱	中	中	强	弱	强
HTAP	强	强	强	强	强	强	中	强

1.3.4 部署方式分类演化：国内云部署实现弯道超车

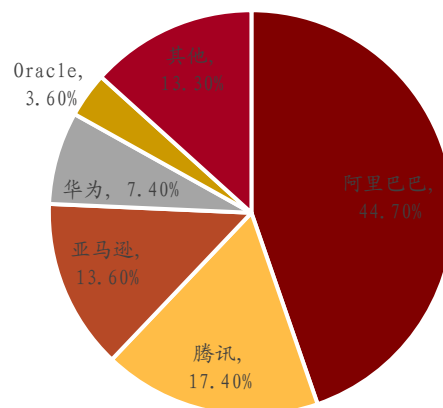
数据库上云成为重要发展趋势。随着云基础设施的成熟和企业解放DBA运维难度的需求日渐迫切，将传统数据库直接部署在云端，由云厂商提供安装和维护服务的云托管数据库逐渐受到业内追捧。据Gartner预测：2023年云DBMS收入将占DBMS市场总收入的50%，75%的数据库将部署在云端。在数据库上云过程中，数据迁移是最大的痛点。针对企业面临的完整性、安全性、一致性等难题，市场催生了安畅网络、数腾软件等提供数据上云迁移的第三方公司。

云托管到云原生演化。分析性需求的涌现和响应速度的要求挑战着传统数据库的IO性能瓶颈，读请求占比远超写请求成为业务常态，因此融合“Shared Storage”和“Shared Everything”以实现资源池化高效管理的云原生数据库应运而生，计算与存储解耦，弹性扩容，秒级备份，轻松应对海量数据存储；日志即数据，通过Paxos等协议通过存储层回放数据，避免直接写数据放大IO，保证数据一致性；一写多读。随着RDMA技术的发展，计算、内存、存储三解耦成为可能，因此通过共享写节点的内存buffer中的数据，从而实现“多写多读”成为云原生数据库演化的新方向。

	本地部署	云部署		
		云托管	云原生1.0（一写多读）	云原生2.0（多写多读）
优点	数据安全性更高	部署简单、开箱即用、自动运维	弹性扩容、数据一致、减少IO	- 云原生2.0版本尚在理论期：计算、内存、存储三者解耦，分层池化； - 目前面临的挑战众多：分布式共享内存技术；计算、内存、存储分层解耦的事务处理架构；网络和存储层算子下推等。
局限	管理维护成本高	传统数据库IO放大，数据读写效率低	写扩展性不足、读节点延迟	
代表	Oracle等	阿里云 RDS MySQL、Amazon RDS	AWS Aurora、华为云 GaussDB	

国内互联网巨头凭借云部署，有望打破关系型数据库市场海外厂商垄断局面。据IDC统计，2021年H1中国关系型数据库本地部署模式被Oracle、华为、Microsoft、IBM垄断，而同期的公有云部署市场由阿里巴巴、腾讯、AWS和华为领衔。2020年，阿里云陆续推出Lindorm、PolarDB-X、ADB等云原生分布式数据库，是首家被Gartner全球数据库魔力象限归入领导者象限的国内厂商。腾讯云研发TDSQL-C，积极拥抱开源，100%兼容MySQL和PostgreSQL，实现超百万级QPS高吞吐，助力中小企业，连续两年增速全球前三。华为云深耕金融行业信创，早在2015年就开始部署上线业界首个AI-

2021H1公有云部署关系型数据库份额



云原生数据库优势

计算存储分离

一致性

分布式

多云管理支持

高可用

高性能

“慧博资讯”专业的数据库GaussDB分享平台

1.3.5 商业模式分类演化：开源强化生态建设

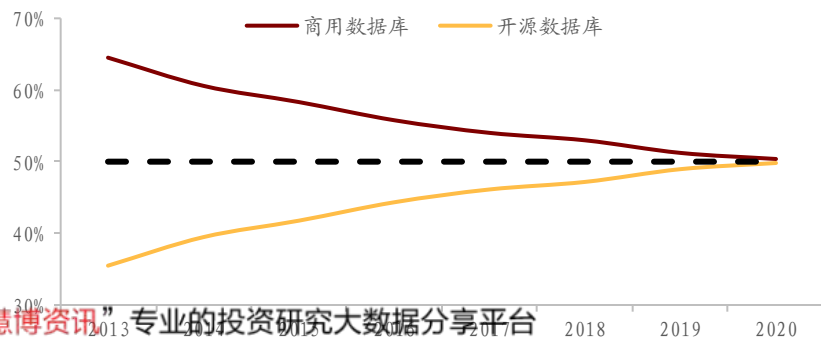
开源有助于强化数据库生态建设，已逐渐成为数据库产业的共识。2021年1月，根据DB-Engines，全球数据库开源许可证流行度首次超过商业许可证，开源数据库逐渐成为行业主流。截止2021年6月，开源与商业许可证数量分别为198和185个，流行度分别占比51.4%和48.6%。开源有助于扩大人才规模及上下游生态影响力，通过运营开源社区快速获得反馈并加快产品开发、提升产品质量，同时反哺社区开发者及独立软件开发商（ISV）等生态伙伴，能够达到多方共赢目的。据信通院调查：超八成的企业认可开源，开源主要应用在存储、大数据分析和数据库领域。

从2014年起，巨杉SequoiaDB、PingCAP TiDB、百度Doris和HugeGraph、涛思数据Tdengine、欧拉NebulaGraph、华为openGauss、阿里PolarDB、蚂蚁OceanBase等相继宣布开源，国内开源大军不断壮大。

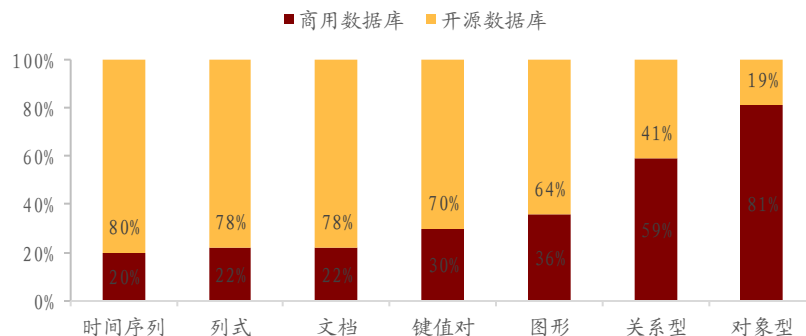
商用数据库VS开源数据库

	商用数据库	开源数据库					
产品特点	商用数据库由成熟的第三方机构提供完善的部署和后期运维，是金融、政企等对安全性要求较高的大型企业首选	开源数据库凭借其源代码透明性、部署灵活性和低成本成为自研能力较强的互联网巨头和预算有限、安全要求较低的初创企业首选					
定义	1) 由专业厂商提供全面解决方案；2) 通常不开放源代码	依据Copyleft License、Permissive License等开源协议开放底层源代码					
代表	Oracle、IBM、Teradata、SPA	Copyleft License协议			Permissive License协议		
		GPL	MPL	LGPL	Apache	BSD	MIT
		MySQL	H2 Database	MariaDB	Spark	PostgreSQL	jQuery
优点	1) 支持个性化定制；2) 完善的部署、培训、后续运维；	1) 费用低廉；2) 部署灵活；3) 可根据需求，灵活、自助修改源代码					
局限	高昂的产品费用（包括服务、license费用）	1) 高昂的选型成本、评测成本、迁移成本（产品成熟度有限，源代码、接口不兼容问题）；3) 后期运维要求高					

DB-Engines数据库流行程度：开源逐步与商用持平



商用VS开源在不同种类数据库中占比



1.3.6 存储介质分类演化：内存数据库提升读写性能

企业对数据读写性能的高要求是数据库存储介质演化的内驱动力：磁盘数据库DRDB->内存数据库MMDB。早期数据库将大部分数据保存在磁盘上，使用内存缓存磁盘内容作为临时存储，统称磁盘数据库；随着电商、直播、新闻查询等对读写性能要求高的场景出现，业界对响应速度的要求逐渐提升，将数据全部存储在内存中、数据更新过程不受磁盘I/O限制的内存数据库逐渐成为主流。据Research and Market统计，全球内存数据市场未来5年的CAGR为19%。

	磁盘数据库DRDB (Disk Historical Database)	内存数据库MMDB (Main Memory Database)
优点	1) 支持ACID事务、数据完整性较高； 2) 可用性高； 3) 产品成熟度高	1) 避开数据访问时磁盘的I/O瓶颈，存取速度快，系统性能高； 2) 直接采用内存地址查询，数据结构简单； 3) 并发控制
局限	1) 需要缓冲处理，占用大量系统资源； 2) 数据存取速度慢； 3) 数据存取时间不一致且难以预测	1) RAM介质掉电数据丢失，安全性较差； 2) 需要额外的日志和快速机制进行灾备； 3) 产品成熟度较低；4) 相对价格更高
代表	Oracle、Mysql、Mongodb、postgresql	SPARK、Redis、HANA、SQLite
逻辑架构		

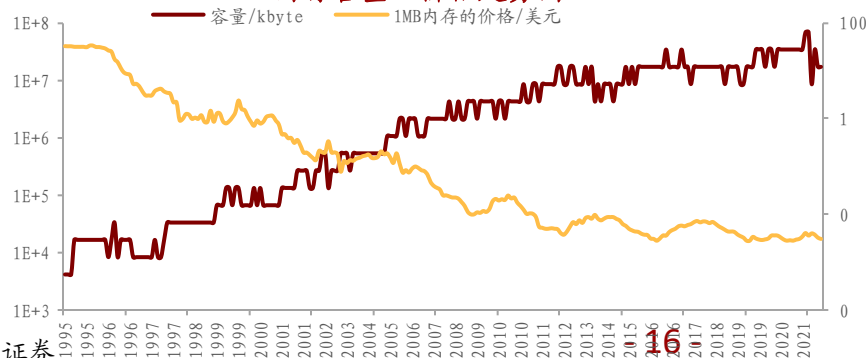
硬件设备价格持续下降是数据库存储介质演化的外驱动力。自1982年80286芯片推出以来，内存市场历经40年的发展，容量遵循摩尔定律翻倍，价格下降超过百万倍，数据大规模存储在内存中成为可能。国际市场上，Pivotal支持的Redis和SAP的HANA领衔键值对、关系型内存数据库；国内市场上，快立方的Qcubic和科蓝软件的Goldilocks已经应用在运营商、金融、交通等行业。

“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

资料来源：内存数据库白皮书、NewEgg、招商证券

内存容量VS价格走势图



1.3.7 治理模式分类演化：AI助力数据库自治

自治数据库提供自治驱动、自治修复、自治安全全生命周期服务。以Oracle的Autonomous Database为例，自治驱动自动针对特定负载配置调优、按需扩展提供高可用数据库；自治修复自动检测和防范系统故障、用户错误，在故障发生时自动切换备用数据库，确保数据零丢失；自治安全自动保护敏感和受监管数据，及时修补数据库安全漏洞，有效防范未经授权的访问。

智能自治成为保障数据库安全可控的重要发展趋势。数据安全是数据库建设的重要前提，自动检测、防范系统故障、运维保护和修补安全漏洞对数据库而言至关重要；随着信息化的发展，应用系统和数据规模井喷式增长，数据库的负载调优压力增大。因此，摆脱对DBA的经验决策依赖，基于人工智能自动执行安全、备份、打补丁和其他需要大量人工重复劳作、从根本上彻底消除人为差错、减轻DBA工作压力的自治数据库成为历史的必然选择。目前自治数据库仍停留在事后判断阶段，展望未来，提前预判数据规模增长趋势和数据库负载状况，事前主动调优是自治数据库的发展新方向。

	自适应数据库 (Self-Adaptive Databases) 1970-1990	自动调优数据库 (Self-Tuning Database) 1990-2000	云数据库 (Cloud Databases) 2000-2010	自治数据库 (Self-Driving Databases) 2018-
特点	使用 DBMS 来减轻应用程序开发人员的数据管理负担，开发人员只需编写一个查询来指定他们想要访问的数据，然后由 DBMS 找到执行该查询的最有效方法。开发人员不必担心使用什么算法、如何存储和检索数据，或者如何安全地交错更新数据的操作。	数据库提供半自动调优工具：DBA 提供示例工作负载，然后该工具建议一个或多个操作以提高性能，但仍然由 DBA 决定这些建议是否正确以及何时部署它们。主流的DB2，Oracle，mysql 都提供了自动参数调优，比如内存分配等，但仍然依赖于人的经验和决策	在云平台中实现操作自动化；利用云的弹性实现自动伸缩特性	自治流程：1) 自动选择优化某个特定功能的方法以及这个优化所需要的资源；2) 决定何时优化；3) 从优化中不断学习和进化。 具体优化的方法包括：1) 改变物理设计，如索引；2) 改变参数配置；3) 改变物理资源，增减节点

Oracle 自治数据库组件

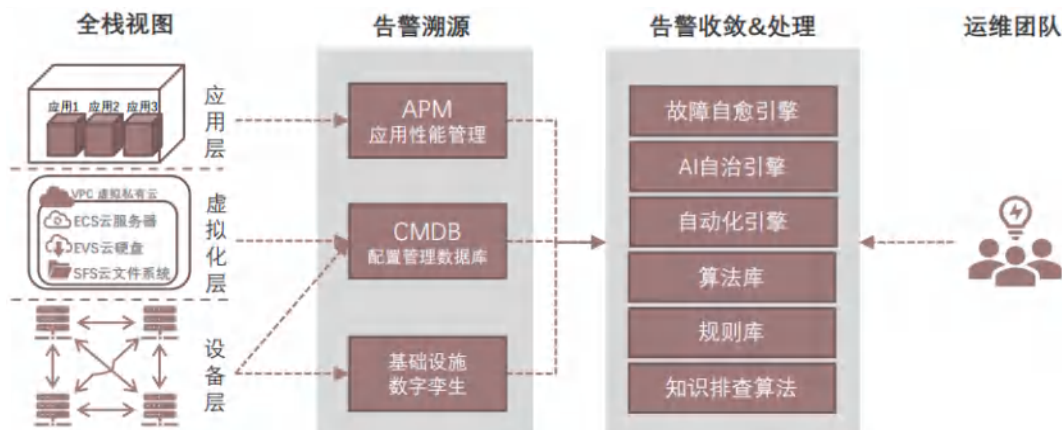
Oracle 数据库 (18c以上版本)
- 数据库企业版的多个自动化技术

Oracle 云基础设施
- 专注于自治的一个标准化平台

策略驱动的自动化

- 自动化每个工作负载的最佳配置，云运维+机器学习

智能运维架构



“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

招商证券

目录

□ 数据库：信息系统承上启下的关键环节

□ 国内数据库市场发展趋势分析

- ✓ 强化数据库建设在“十四五”规划中被多次提及
- ✓ 国内数据库市场三大主要增量需求：
 - 国产化进程逐步加速
 - 传统集中式数据库向分布式架构升级
 - 非关系型数据库应用场景不断丰富，快速发展

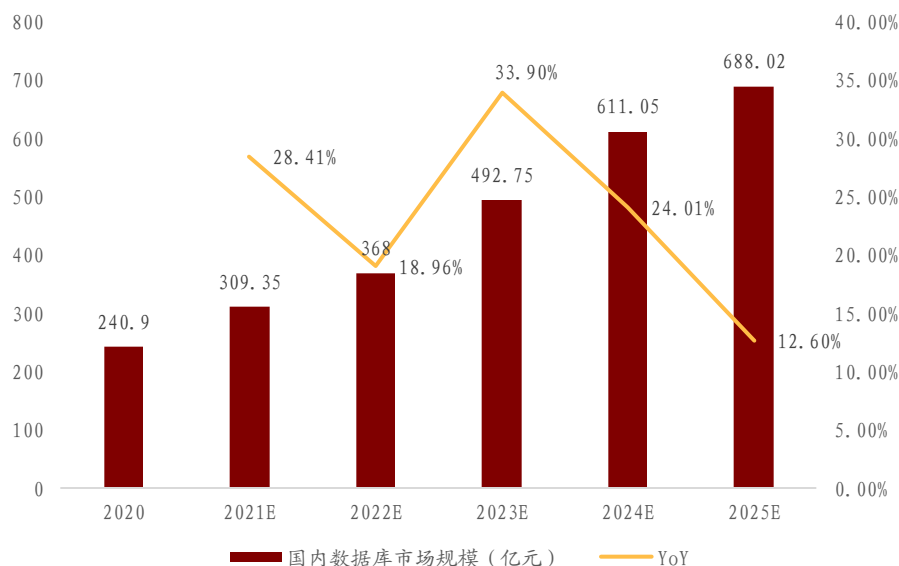
□ 国产数据库厂商详细梳理

2.1 行业规模与下游需求结构

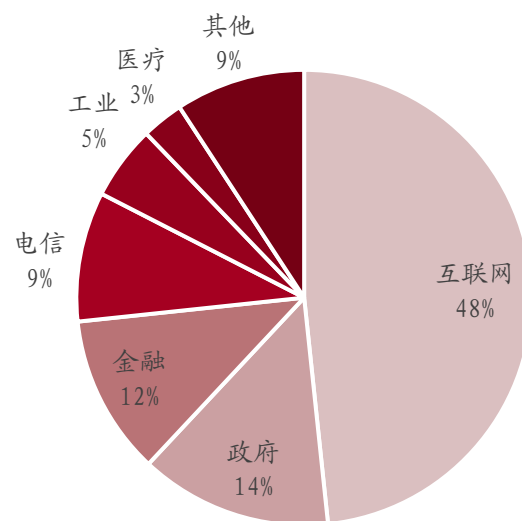
数据库作为信息化核心环节，在国内大力发展数字经济建设的大背景下，数据库行业规模将保持快速增长。根据信通院，2020年中国数据库市场规模为35亿美元（约合240.9亿元人民币），占全球总规模的5.2%。预计到2025年，全球数据库市场规模将达到798亿美元。中国的IT总支出将占全球12.3%。预计中国数据库市场在全球的占比将在2025年接近中国IT总支出在全球的占比，市场总规模将达到688亿元，CAGR为23.4%。

金融、电信、政务、制造、互联网五个行业为数据库行业下游主要需求领域，大数据应用快速发展为数据库提供增长动力。根据信通院统计，其合计市场份额高达80%以上，2020年整体采购额超过192亿元。此外，根据赛迪顾问，2020年国内大数据市场结构数据，其中互联网、政府、金融、电信、工业位列前五，占比分别为48%、14%、12%、9%、5%，合计约88%，大数据应用与数据库需求结构基本吻合，预计随着大数据应用领域拓展与规模持续增长，将为数据库行业规模增长提供充足动力。

国内数据库规模预计将快速增长



2020年国内大数据应用市场结构



2.2 强化数据库建设在“十四五”规划中被多次提及

去年各领域“十四五”规划陆续发布，数据库建设与应用在信息技术、金融、交通物流、农业等众多领域被多次提及，其中数据库标准化攻关的重要性首次被提升到前所未有的高度。我们认为，随着数据库顶层设计不断加码，十四五期间，国产数据库研发将步入快车道，国产数据库竞争实力将显著增强，行业有望进入加速发展期。

强化国产数据库建设在“十四五”规划中被重点提及

领域	政策文件	数据库相关内容
信息技术	《国家标准化发展纲要》	“开展数据库等方面标准攻关，提升标准设计水平，制定安全可靠、国际先进的通用技术标准。”
	《“十四五”软件和信息技术服务业发展规划》	“加速分布式数据库、混合事务分析处理数据库、共享内存数据库集群等产品研发和应用推广。” “突破全内存高速数据引擎、高可靠数据存储引擎、分布式数据处理与任务调度架构、大规模并行图数据处理等关键技术。推动高性能数据库在金融、电信、能源等重点行业关键业务系统应用。突破嵌入式操作系统、嵌入式数据库核心技术，加快相关产品研发与应用推广。”
金融	《银行业保险业数字化转型指导意见》	“加快数据库、中间件等通用软件技术服务能力建设，支持大规模企业级技术应用。”
交通物流	《“十四五”冷链物流发展规划》	“推动大数据、物联网、5G、区块链、人工智能等技术在冷链物流领域广泛应用。以冷链食品追溯为突破，形成以责任主体为核心的追溯闭环，对跨部门、跨地域的全链条追溯数据进行大数据分析，为信用评价提供数据支撑。”
	《“十四五”铁路科技创新规划》	“研发地理信息系统大数据（GIS）、云计算技术支持下的智能选线技术，实现多方案自动生成和多维度智能评价。大力推进北斗卫星导航、5G、人工智能、大数据、物联网、云计算、区块链等前沿技术与铁路技术装备、工程建设、运输服务等领域的深度融合，加强智能铁路关键核心技术研发应用，推进大数据协同共享，促进铁路领域数字经济发展，提升铁路智能化水平。”
	《数字交通“十四五”发展规划》	“推动交通运输数据资源在部、省两级有效汇聚整合，推动形成质量高、覆盖广、体系全的交通运输核心数据库。”
	《“十四五”现代流通体系建设规划》	“健全应急物流运转保障机制，引导建立应急物流大数据平台，推动与应急管理信息平台数据共享，完善信息采集、动态监测、数据分析、风险预警、信息发布等功能，重点加强对物流大面积中断风险的研究评估，提高应急物流组织能力。”
知识产权	《知识产权公共服务“十四五”规划》	“鼓励支持有条件的地区围绕知识产权密集型产业、战略性新兴产业、数字经济核心产业等重点产业和关键核心技术，建设多种类知识产权专题数据库，并向全国推广应用。”
	《版权工作“十四五”规划》	“组织开展软件兼容性测试，以制度建设、软件采购、使用管理、日常检查、年度报告和考核评议为基础，以基础软件、工业软件、平台软件为重点，探索建立软件正版化工作大数据中心和正版软件防伪溯源平台。”
农业	《“十四五”全国农业农村科技发展规划》	1、“开展农业种质资源精准鉴定与基因挖掘，构建种质资源DNA分子指纹图谱库、特征库和种质资源数据库。” 2、“围绕农作物产品、畜禽产品、水产品、特色农产品农兽药残留等质量安全指标开展常态化监测，优化监测参数和品种，完善我国农产品质量安全监测数据库。” 3、“构建适用于农业领域的非结构化数据库系统、农业数据仓库、农业知识计算引擎、农业可视交互服务引擎等核心技术。建立乡村建设与可持续发展数据库，构建智慧乡村大数据云平台。” “围绕种植业、养殖业、渔业的时空格局、生产制度及演变规律等环境要素开展动态监测，系统收集典型生产区水、土、气、物候、投入品等数据信息，构建环境演变数据库。”

2.3 数据库市场增量需求

基于对数据库发展趋势与国家政策导向的分析，我们认为，未来数据库市场增量需求主要来自于以下三方面：

1. 国产化进程逐步加速

2. 传统集中式数据库向分布式架构升级

3. 非关系型数据库应用场景不断丰富，得到快速发展

首先，在国产化进展方面，根据CSIA与海比研究院联合研究，政府、教育、金融、电信等已经进入规模化阶段，且根据对央企CIO调研统计，预计央企在2024年将进入数据库国产化建设高峰，在2030年基本完成，国产数据库未来规模增长有望进一步提速。

各行业数据库国产化进程

行业	替换难易	国产化进度
政府	难度较低，数据结构相对简单，新建系统可直接采用国产化数据库；同时，存量数据库替换，已有众多成功案例，技术成熟。	新建数据库系统已基本实现国产化，存量Oracle数据库国产化正全面展开
教育		在线教育进展较快，是目前Oracle数据库替换重点领域，多数在线教育公司直接采用国产化数据库构建自身系统；传统教育领域国产化暂未大规模展开
金融	难度较高，金融、电信行业对于数据的实时性、安全性要求较高，且业务数据量巨大，数据库功能较为复杂。替换过程中需要保证业务运行的连续性、数据传输实时性与安全性，整体难度较高，根据统计，单系统替换至少2-3年。	目前，银行、保险企业的互联网业务、二三线城市商业银行、农村商业银行都在不同程度地进行数据库国产化
电信		运营商业务互联网化转型下，对数据库国产化需求较为强烈；目前，二、三线分公司整逐步进行替换试点
制造业	难度高，由于工业软件领域长期被海外厂商主导，且前期缺乏相关政策指引，企业CIO或数智化对系统替换动力不足，导致国产化进展较为缓慢。随着21年“十四五”相关规划出台，制造业、能源等行业国产化有望显著加快。	暂未开展规模化替代，目前在部分边缘小型应用领域进行替换尝试
能源		暂未开展规模化替代，只在流程相对简单的业务系统进行替换试点

“慧博资讯”专业的投资研究大数据分享平台



招商证券

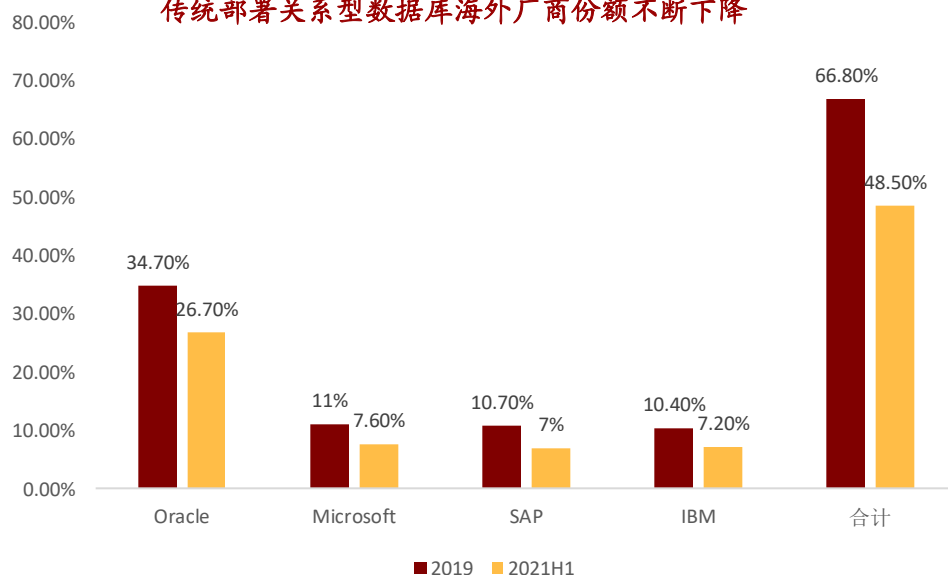
资料来源：信通院、中国软件行业协会CSIA、海比研究院、招商证券

2.4.1 数据库国产化浪潮不可阻挡

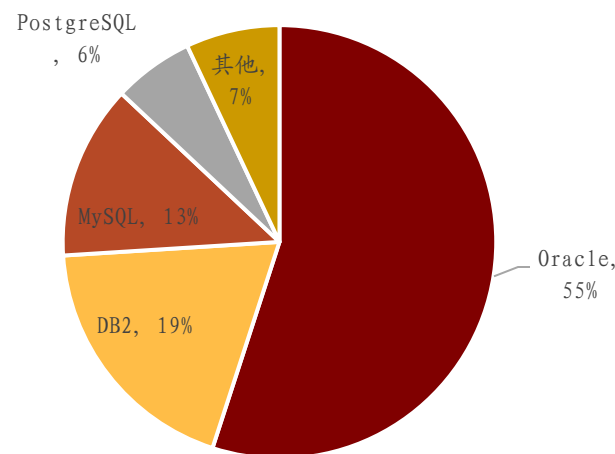
国产数据库能力不断提升，逐步抢占海外厂商优势领域份额。以Oracle、Microsoft、SAP、IBM为代表的海外厂商，一直以来，因其产品具有更好的性能、稳定性、安全性、与海外ERP、CRM等更好地兼容性等优势，在企业核心OLTP业务场景一直占据主导地位。近些年，随着国产数据库产品功能、性能水平不断增强，其产品水平已经逐步趋近于海外厂商，根据IDC，在传统关系型数据库领域，Oracle、Microsoft、SAP、IBM合计份额从2019年66.8%下降至2021H1的48.5%，首次降至半数以下，显示出国产数据库技术正趋于成熟，在海外厂商为主导的核心OLTP应用场景，已经逐步具备运营能力，国产化进程持续加速。

根据信通院统计，目前我国金融行业海外Oracle与IBM DB2占比约74%，占据主导，基于MySQL与PostgreSQL二次开发的国产数据库合计占比约为19%，国产化仍有广阔空间。其中，我们通过对金融行业国产化数据库建设案例梳理发现，**国产数据库已经从边缘系统逐步实现核心系统国产化**，显示出国产数据库产品功能、性能水平不断增强。预计随着行业信创逐步落地，对国产数据库需求将进一步提速。

传统部署关系型数据库海外厂商份额不断下降



金融行业各数据库占比（截至2021H1）



2.4.1 数据库国产化浪潮不可阻挡

以金融行业为例，国产化正呈现出由股份制银行、城商行向国有大行快速渗透，逐步实现支撑其核心业务系统，同时近两年替换进程显著加快。金融行业对数据的实时性、安全性、高并发性能等具有较高的要求，我们认为，随着国产化在金融业的规模化展开，国产厂商凭借其在金融业积累的经验，未来针对其他行业的国产化进程将显著加快。

金融行业数据库国产化进程近两年持续加速

时间	企业	国产化案例
2021.8	友邦人寿	友邦人寿宣布加速 重要业务系统 上云进程，包括保险出单、保单契约等在内的数十套业务系统开始采用阿里云PolarDB云原生数据库
2021.4	邮储银行	基于底层数据库采用国产开源数据库openGauss，新一代 个人业务核心系统 的技术平台、智能运维平台和个人国际汇款业务相继上线
2020.9	工商银行	采用蚂蚁自研数据库OceanBase，其对 公理财系统 已完成从大型主机到OceanBase分布式架构的改造
2020.7	中信银行	中兴通讯和中信银行联合宣布，GoldenDB分布式数据库顺利完成在中信银行 核心系统“凌云” 系统的投产
2020.4	贵阳银行	基于易鲸捷新一代国产融合型分布式数据库搭建的贵阳银行核心交易“沙箱”系统联机实时运行，成功实现 核心交易“沙箱”系统 与贵阳银行现有核心系统并行处理
2019.9	张家港农商行	基于分布式数据库腾讯云TDSQL打造的张家港农商银行新一代 核心业务系统 成功上线投产
2019.5	湖北银行	其新核心系统建设采用长亮科技分布式核心系统搭载国产大型通用关系型数据库-达梦SQL数据库，初步实现Oracle与国产数据库在 核心关键业务 中异构融合，验证了核心业务系统替换国外商用数据库平台的方案可行性
2015.3	微众银行	采用基于开源系统的腾讯自研TDSQL数据库，支持海量数据的分布式技术架构，彻底从技术层面保证了系统运行的高效和稳定，成为国内首个去IOE的分布式核心系统

2.4.2 集中式数据库向分布式升级

随着数据类型、规模的指数级增长，传统的终端计算场景已经难以应对数据存储处理的工作量与复杂度。云计算通过存算分离、资源弹性动态分配、边缘节点计算打破了传统计算场景的瓶颈，实现了当前数据处理的需求，云计算场景的需求推动了分布式数据库的应用发展：

1. 通过资源池化管理实现物理或逻辑层的相互隔离与资源的自由伸缩，具备弹性扩张、HTAP事务能力、多租户管理能力、高可用性，与云计算场景需求相匹配。
2. 对多种访问接口和数据类型的兼容，可以实现对存储与不同物理服务器、不同格式的数据进行结构与算法的优化，突破服务器类型的限制，为上层不同类型的应用提供多模式的数据服务。

分布式数据库在扩容成本与并发能力优势明显

比较	集中式	分布式
可扩展性	纵向扩展	横向扩展
扩容弹性	大变更需停机	快速迭代、小时级投产、分钟级在线扩容
成本	扩产需要更高配置硬件，成本较高：单账户成本几十元，单笔交易成本几角钱	低配置廉价PC Server组件集群，成本较低：但账户成本1元以内，单笔交易成本几分钱
并发能力	支持表单千万级数据量存储，难以支撑高强度的读写并发，容量与性能存在瓶颈	多节点关联支撑，可减少单节点数据量；可实现并行计算，支持PB级数据访问，与百完级高并发读写

分布式架构凭借在经济性、安全自主、灵活性、可伸缩性等方面的优势，逐步实现对集中式架构替代。目前，互联网、金融、电信等行业分布式升级进展较快，以金融和电信行业为例，其核心业务数据类型均为关系型数据库，此类集中式数据库所面临的业务体量将呈现爆发性增长。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，业界已有分布式数据库在股份制银行、城商行、二三线城市运营商等实现了核心业务系统的改造落地，同时，去年12月，腾讯分布式数据库TDSQL落地东吴证券核心交易系统，也标志着分布式数据库在券商核心系统领域取得突破。

分布式数据库落地案例

时间	企业	分布式案例
2021	东吴证券	腾讯云TDSQL落地东吴证券国产数据库首次落地证券核心交易系统
2021	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
2019	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
2018	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
2013	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

“慧博资讯” 专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

资料来源：沙利文、头豹研究院、信通院、墨天轮、招商证券

2.4.2 国产分布式数据库热度持续高涨

墨天轮国产数据流行度排行（截至2022/02），前十大热门数据库均为分布式架构，其应用场景涵盖各行业，显示出国产分布式数据库已经趋于成熟，正成为各行业数据库建设重要选择。

墨天轮前十大热门数据库均为分布式数据库

数据库	厂商	数据模型	技术架构	需求功能	部署方式	商业模式	特点	场景
TiDB	PingCAP	NewSQL	分布式	HTAP	云原生	开源	实时HTAP、高并发OLTP、兼容MySQL生态	金融、电信、政府、能源、公共事业、高端制造、高科技、物流
openGauss	华为	关系型	分布式	HTAP	云原生	开源	具备AI4DB、DB4AI能力，参数智能调优，支持XGBoost等高级算法	交易型为主的应用（电商、金融、O2O、电信CRM/等）和物联网
OceanBase	蚂蚁集团	关系型	分布式	HTAP	云原生	开源	高度兼容MySQL、Oracle，提供一键数据迁移方案，支持国产操作系统和国产芯片	金融、运营商、政府公共服务
达梦+	达梦	关系型	分布式	HTAP	云原生	商业	支持web应用、对称多处理机系统	国家电网、航空航天、国家安全、国防军工、金融银行、电子政务、公检法司
GaussDB	华为	关系型、非关系型	分布式	HTAP	本地、云原生	开放 openGauss 生态	AI-Native数据库系列	金融政企
PolarDB	阿里巴巴	关系型	分布式	HTAP	云原生	开源	100%兼容开源MySQL、RDS MySQL和PostgreSQL；“一主多从”架构保障数据安全	电商
TDSQL	腾讯云	关系型	分布式	OLAP	云原生	开源	智能运维，兼容主流数据库	游戏、电商、移动互联网、云开发
GBase	南大通用	关系型、非关系型	分布式	OLAP	本地、云部署	商业	高性能、高性价比、高易用性	金融、电信、政务、国防、企事业
KingBase	人大金仓	关系型	分布式	OLTP	本地、云部署	商业	高度容错，可支撑6级灾难恢复能力等级要求	电子政务、国防军工、电力、金融
AnalyticDB	阿里云	关系型	分布式	OLAP	云原生	商业	超大规模、高并发写入、自动索引、智能优化	经典实时数仓、实时计算清洗回流场景

2.4.3 非关系型数据库快速发展

我国非关系型数据库快速发展，正逐步进入世界前列。我国非关系型数据库主要以基于开源数据库如Redis、InfluxDB、CouchDB等产品进行二次开发为主，根据DB-Engines 2022年2月排名，国内浙江智史和阿里云TSDB，分别位列时序数据库第11、21位；图数据库方面，欧若数网Nebula Graph、百度智能云开源产品HugeGraph、华为云GraphBase分别位列15、26、32位，显示出国产NoSQL数据库正逐步趋近于世界前列。

NoSQL数据库在制造、政府、金融等领域逐步落地

随着工业互联网、互联网创新型业务、车联网等应用的快速发展，非关系型数据库应用将加速落地。根据DB-Engines热度排名，图数据库与时序数据库排名第一、二，是近些年热度增长最快的细分数据库领域：

- 图数据库：能够支撑社交网络、金融反欺诈等互联网新型应用场景的关联分析业务，根据Gartner预测，图技术将广泛应用于金融安全、互联网、工业、生物医药、公共安全、智慧城市等多个领域，到2025年将应用于80%的数据和分析创新，远高于2021年的10%，图技术是图数据库的核心技术，从而将推动图数据库市场规模持续增长；
- 时序数据库：在工业领域具有广阔的应用潜力，其可以高效存储和快速处理海量时序大数据，在工业互联网快速发展下，传感器与设备产生的大量实时数据，将推动时序数据库应用快速发展。

行业	时间	企业	分布式案例	数据库类型
制造业	2021	睿信世达	涛思数据 TDengine 进行睿信物联网平台改造	时序数据库
	2020	三一重工	腾讯时序数据作为设备数据分析引擎	时序数据库
政府	2019	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	图数据库
	2020	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	时空数据库
金融	2021	华瑞银行	阿里GDB图数据库辅助智能风险管理	图数据库
	2019	民生银行	巨杉SequoiaDB数据库用于分布式影像平台	NoSQL

图、时序、键值等NoSQL数据库热度持续上升（截至2022/02）



目录

- 数据库：信息系统承上启下的关键环节
- 国内数据库市场发展趋势分析
- 国产数据库厂商详细梳理
 - ✓ 达梦：融合新兴技术，独占鳌头十余年
 - ✓ 南大通用：扩大生态领域，全面开花
 - ✓ 人大金仓：深耕关系型数据库领域
 - ✓ 巨杉：无惧差异，融为一体
 - ✓ 华为GaussDB：将AI能力植入数据库内核
 - ✓ 蚂蚁OceanBase：分布式数据库的极致性能
 - ✓ PingCAP：向云而生，拥抱开源

3.1.1 国内主流数据库厂商

流行度	数据库	厂商	数据模型	技术架构	需求功能	部署方式	商业模式	特点	场景
4	达梦+	达梦	关系型	分布式	HTAP	云原生	商业	支持web应用、对称多处理机系统	国家电网、航空航天、国家安全、国防军工、金融银行、电子政务、公检法司
8	GBase	南大通用	关系型、非关系型	分布式	OLAP	本地、云部署	商业	高性能、高性价比、高易用性	金融、电信、政务、国防、企事业
9	KingBase	人大金仓	关系型	分布式	OLTP	本地、云部署	商业	高度容错，可支撑6级灾难恢复能力等级要求	电子政务、国防军工、电力、金融
13	Sequoia DB	巨杉	NewSQL、多模	分布式	HTAP	云原生	商业	多租户能力，支持多种级别的物理与逻辑隔离	证券、保险、电信、政府、互联网、交通
15	ShenTong	神舟通用	关系型	分布式	HTAP	本地部署	商业	获公安部三级安全认证，安全性高	航天、航空、国防、船舶、政府、金融、电信、能源
19	GreatDB	万里开源	关系型	分布式	HTAP	云原生	商业	集群高可用，数据副本多种存储策略，可存储在不同磁盘、机柜、数据中心	IT、金融、电商、运营商
22	UXDB	优炫	NewSQL	分布式	HTAP	云部署	商业	提供多层次的数据库自适应、自配置、自管理、自调优、自修复能力，以及丰富且直观的功能操作	政府、军工、金融、能源、制造、医疗
51	K-DB	浪潮	关系型	分布式	OLTP	云部署	商业	全面兼容Oracle数据库，实现自动化迁移工具的一键式操作	金融、能源、政府、交通
68	HighGoDB	瀚高	关系型	分布式	OLTP	本地、云部署	商业	中文全文检索、状态监视器实时监控运行状态	医疗卫生、大型企业、教育、政府

注：流行度来自墨天轮对国产数据库的2022年2月排行榜

慧博资讯 专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

资料来源：墨天轮、招商证券

3.1.2 互联网厂商

流行度	数据库	厂商	数据模型	技术架构	需求功能	部署方式	商业模式	特点	场景
2	openGauss	华为	关系型	分布式	HTAP	云原生	开源	具备AI4DB、DB4AI能力，参数智能调优，支持XGBoost等高级算法	交易型为主的应用（电商、金融、O2O、电信CRM/等）和物联网
3	OceanBase	蚂蚁集团	关系型	分布式	HTAP	云原生	开源	高度兼容MySQL、Oracle，提供一键数据迁移方案，支持国产操作系统、芯片	金融、运营商、政府公共服务
5	GaussDB	华为	关系型、非关系型	分布式	HTAP	本地、云原生	开放openGauss生态	AI-Native数据库系列	金融政企
6	PolarDB	阿里巴巴	关系型	分布式	HTAP	云原生	开源	100%兼容开源MySQL、RDS MySQL和PostgreSQL，数据安全	电商
7	TDSQL	腾讯云	关系型	分布式	OLAP	云原生	开源	智能运维，兼容主流数据库	游戏、电商、移动互联网、云开发
10	AnalyticDB	阿里云	关系型	分布式	OLAP	云原生	商业	超大规模、高并发写入、自动索引、智能优化	经典实时数仓、实时计算清洗回流场景
11	GoldenDB	中兴通讯	关系型	分布式	OLTP	本地、云部署	商业	深耕金融行业，彻底解决事务一致性、数据实时一致性问题	银行、互联网、政务
100	ByteGraph	字节跳动	图	分布式	OLTP	本地部署	商业	支持有向属性图数据模型、Gremlin 查询语言、灵活的写入和查询接口	字节产品线（头条、抖音、TikTok、西瓜、火山等）
108	HugeGraph	百度	图	分布式	OLTP	本地、云部署	开源	支持从多数数据源批量导入数据	网络安全、电信诈骗、金融风控、广告推荐、社交网络和智能机器人

“慧博资讯”专业的投资研究大数据分享平台

注：流行度来自墨天轮对国产数据库的2022年2月排行榜

点击进入  <http://www.hibor.com.cn>

资料来源：墨天轮、招商证券

3.1.3 部分初创厂商

流行度	数据库	厂商	数据模型	技术架构	需求功能	部署方式	商业模式	特点	场景
1	TiDB	PingCAP	NewSQL	分布式	HTAP	云原生	开源	实时HTAP、高并发OLTP、兼容MYSQL生态	金融、电信、政府、能源、公共事业、高端制造、高科技、物流
14	EsgynDB	易鲸捷	NewSQL	分布式	HTAP	云部署	商业	自主研发核心底层架构，安全可控	金融、智能制造、电信
16	TDEngine	涛思数据	时序	分布式	HTAP	本地、云部署	商业	全栈时序数据处理引擎	物联网、车联网、工业互联网、IT运维
24	DolphinDB	智臾科技	时序	分布式	OLAP	本地、云部署	商业	为海量结构化数据的快速存储、检索、分析及计算提供一站式解决方案	量化金融、工业物联网
33	ArgoDB	星环科技	关系型	分布式	HTAP	本地、云部署	商业	支持内存/闪存/磁盘三级混合存储	金融、政府、能源、交通、运营商
50	PandaDB	中科知道	图	分布式	OLAP	本地部署	商业	通过模型和资源管理,实现 AI 模型的灵活部署、高效按需运行、有效屏蔽	工业互联、科技管理、智能制造、精准营销
57	GDM	蜀天梦图	图	分布式	HTAP	本地部署	开源	图遍历同时支持广度优先遍历和深度优先遍历	政府、互联网、企事业单位
66	Nebula Graph	欧诺数网	图	分布式	HTAP	云原生	开源	类 SQL 查询语言 nGQL, 减少了程序员迁移的成本, 同时表达能力强	社交、风控、知识图谱、智能推荐
71	SinoDB	星瑞格	NoSQL	分布式	HTAP	云部署	商业	支持嵌入式设备, 广泛应用于物联网	环保、农业、科技、金融、政府、医疗、水利、电力
99	xigemaDB	华盛信泰	NewSQL	分布式	HTAP	云部署	商业	具有高度智能引擎, 无需人工干预即可自主调整资源需求和利用, 以适应实时的应用程序请求	国家重要自然资源的数据资产管理方面: 如农、林、牧、副、渔、水利、测绘、气象等领域

“慧博资讯”专业的投资研究大数据分享平台

注：流行度来自墨天轮对国产数据库的2022年2月排行榜
 资料来源：墨天轮、招商证券

3.2 达梦：融合新兴技术，独占鳌头十余年



科技创新，夯实信息化安全基石，蝉联国产数据库市占率冠军十余年。武汉达梦数据库股份有限公司（以下简称达梦）成立于2000年，是中国电子信息产业集团（CEC）旗下基础软件企业，由中国软件与技术服务股份有限公司控股25%。达梦致力于数据库管理系统的研发、销售和服务，为国内外数十万用户提供全栈数据产品和解决方案，为打破欧美等国的技术封锁，达梦全自研底层逻辑代码，立足国内客户，深挖用户需求，先后完成了近60项国家及省市级的科研开发项目，取得了近400项研究成果，获得国家、省部级奖励超过30项。据赛迪顾问统计：2000年以来，达梦在数据库市场的市占率已经蝉联冠军十余年。

达梦自研高性能数据库管理系统DM，不断融合分布式、云计算等新兴概念升级系统架构。DM系列自上世纪80年代开始研发实验室原型，2000年初商业化，已经更新迭代8个版本，对灵活性、易用性、可靠性、高安全性等方面进行了大规模改进：分布式架构实现动态分配计算资源、精细化资源利用；行列融合，支持超大规模并发HTAP；技术生态升级，支持多云环境部署。

研发大规模数据处理集群DMMPP，支持TB到PB级别的数据存储与分析。基于完全对等无共享架构，最多可将1024个DM8节点组织为一个并行计算网络，支持HASH、范围、随机多种数据分布和水平、垂直、多级混合分区，提供高可用性和动态扩展能力。

DM1-DM3 (1984-2003)

1. 实验室原型；
2. 技术积累阶段；
3. 制定并实现各类标准；
4. 1988年开发出我国第一个自主版权数据库CRDS；

DM4-DM6 (2004-2007)

1. 国家863数据库重大专项；
2. 支持定制化开发；
3. 支持大规模数据存储管理；
4. 引入物理操作符、虚拟机；
5. 实现企业级高级特性；
6. 新增数据复制与外部连接；

DM7-DM8 (2011-)

1. 多维融合，满足多样性：读写分离架构、基于行列融合2.0的混合事务分析处理技术；
2. 分布式架构，提升用户体验：融合分布式、弹性计算与云计算的优势，对灵活性、易用性、可靠性、高安全性进行了大规模改进；
3. 技术生态升级：支持更广泛的技术选型，支持多种云计算基础设施环境、软硬件平台；
4. 升级运维管理工具，增强系统安全性。

产品化、企业级、兼容其他数据库

HTAP融合、分布式、云部署

代表客户&项目

深圳证券交易所OA系统；武汉公积金核心业务系统

国家电网一体化调度；南方电网信息化建设

联合中国航信打造民航客票

中铁建财务共享服务平台

全国检察机关统一业务系统；海南社保、医保一体化

金融

能源

航空

交通

政务

“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

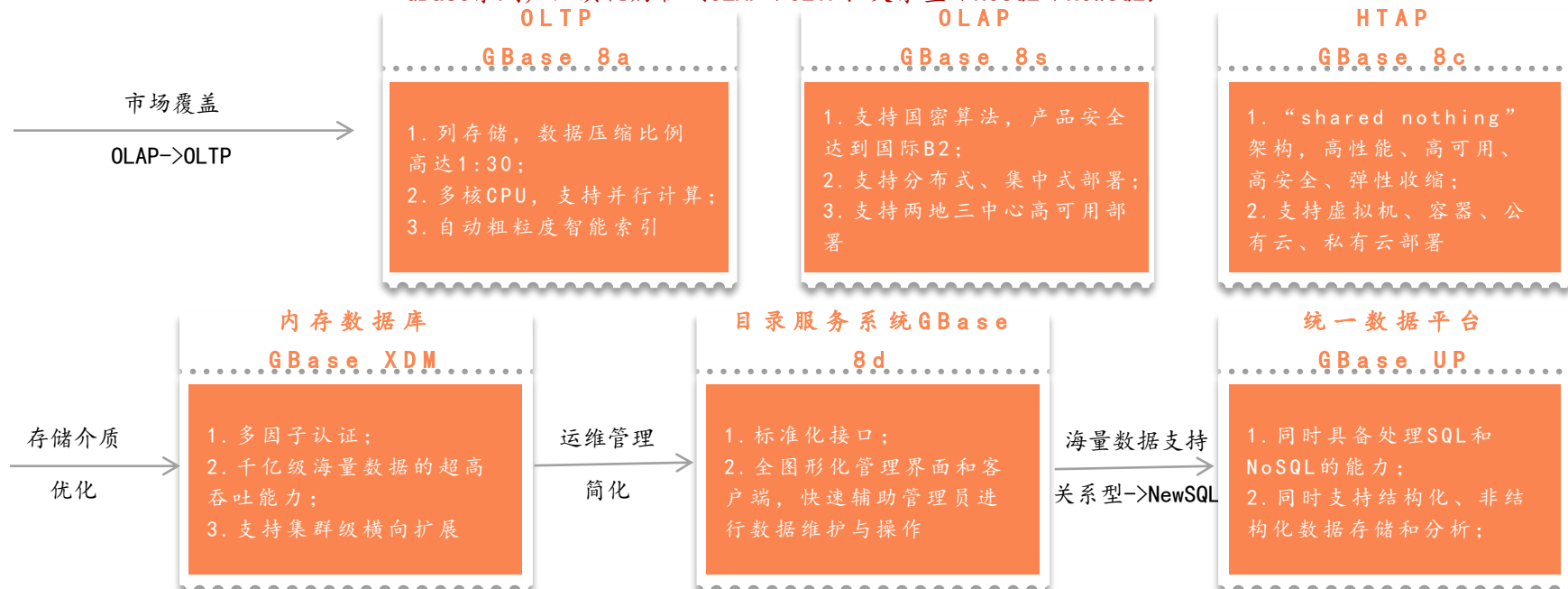
资料来源：达梦公司官网、达梦技术白皮书、赛迪顾问、招商证券

3.3 南大通用：扩大生态领域，全面开花 GBASE®

南大通用入选专精特新“小巨人”企业。天津南大通用数据技术股份有限公司（以下简称南大通用）成立于2004年，注册资金1.4亿元，近7年连续被赛迪顾问评为国产数据库龙头企业，并在2021年成功入选首批国家级重点专精特新“小巨人”企业。

GBase系列基于Informix不断创新发展，从OLAP到OLTP，关系型→NoSQL→NewSQL，逐渐扩大生态，提高市占率。南大通用深耕金融行业，自主研发国内首个基于列存的新型分析型数据库GBase 8a；2014年与IBM签署战略合作，获得Informix的源代码和技术授权，站在巨人的肩膀上研发复杂事务处理数据库产品GBASE 8s，而后又融合分布式架构研发8c，提高系统可用性。

GBase系列产品演化历程（OLAP→OLTP、关系型→NoSQL→NewSQL）



代表客户&项目

中国银行总行大数据平台；中国农业银行数据仓库项目

国家海洋大数据中心；通信管理局大数据分析平台

中国移动总部集中经营分析系统；北京移动分析中心

海关总署金关工程；金税管理决策系统；海南社保云平台

重庆地铁AFC项目

金融

党政

电信

政企

交通

“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

资料来源：南大通用公司官网、南大通用技术白皮书、招商证券

3.4 人大金仓：深耕关系型数据库领域

KING BASE™

人大金仓产学研一体，快速发展，下沉市场，布局未来。北京人大金仓信息技术股份有限公司（以下简称人大金仓），由中国人民大学的数据库专家于1999年发起创立，先后承担了国家“863”、“核高基”等重大专项，是中国电子科技集团有限公司（CETC）的成员企业，在党政军及各级企业级市场应用广泛，目前已覆盖全国近3000个县市，完成装机部署近100万套，入围工信部“2021年数字技术融合创新应用典型解决方案”。

集中关系型数据库领域发力，市占率持续领跑。据IDC统计，2019年下半年国产数据库市场的市占率中，人大金仓仅次于阿里和南大通用，位列第三。相比前两名开疆扩土式的全面发展，人大金仓聚焦关系型数据库领域深耕，分别针对OLTP、OLAP、HTAP场景研发Kingbase ES (KES)、Kingbase AnalyticsDB (KADB)、KSone系列数据库。结合信创改革浪潮，人大金仓积极适配国产上下游芯片、操作系统、中间件等，与生态伙伴完成近万个产品兼容互认证，力图为用户持续提供稳定服务。

完善的数据迁移工具平滑、高效地解决异构数据库数据迁移难题，助力数据库推广。

Kingbase系列产品：分布式、新硬件、自治贯穿发展历程



“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

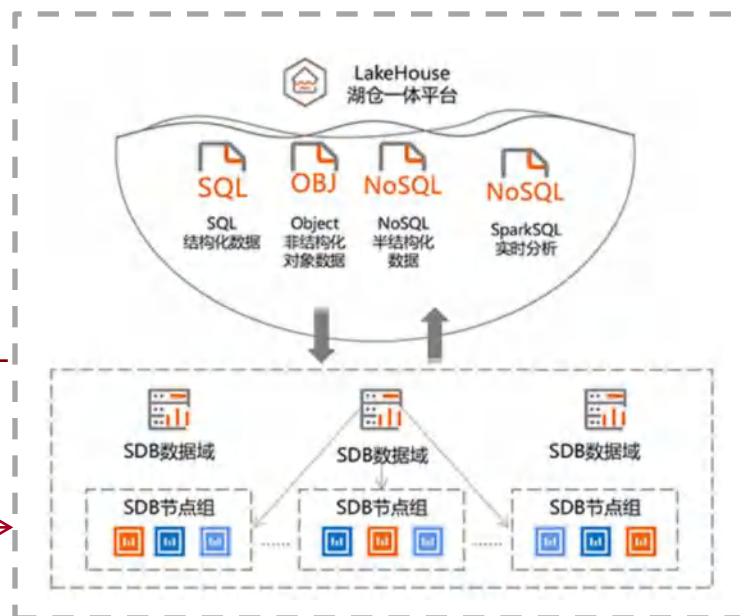
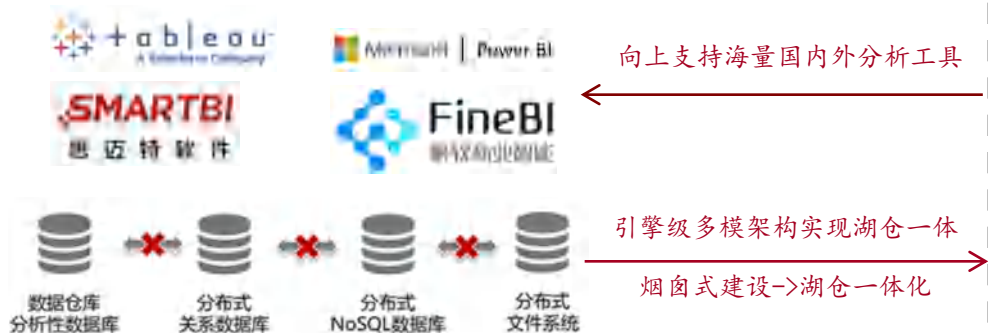
资料来源：人大金仓公司官网、招商证券

3.5 巨杉：无惧差异，融为一体

2017年巨杉数据库与阿里云成为首家入选Gartner报告的国产独立数据库厂商，连续三年入榜Gartner全球权威报告。深圳巨杉数据库软件有限公司（以下简称巨杉）2011年成立，专注金融行业数据库产品研发，目前已在超过100家大型银行及金融机构的生产业务条线上应用。巨杉的核心团队成员主要来自于前IBM DB2北美研发团队和华为分布式储存团队，天生具备企业级基因，同时引入包括阿里在内的国内创新型企业人才，是业界最顶尖的企业级分布式数据库团队。

引擎级多模，丰富上下游生态。巨杉数据库兼容多种SQL结构化数据、NoSQL半结构化数据和Object非结构化对象数据引擎，自研分布式数据库内核，研发四款核心产品：数据融合平台SequoiaDB-DP、事务型数据库SequoiaDB-TP、内容管理数据库SequoiaDB-CM、文档型数据库SequoiaDB-DOC。核心产品兼容Apache SparkSQL引擎，助力实时数据分析，避免传统数据仓库架构导致的ETL延迟及数据多份冗余。

打破数据湖与数据仓库割裂的体系，打造湖仓一体化，构建全新的融合数据平台，轻松实现跨引擎数据中台。主打产品SequoiaDB-DP在湖仓一体化的框架下，可以轻松打造跨引擎数据中台，实现一次写入、多引擎可读。助力客户实现提升数据管理水平、降低成本、提升运营效率、提升用户体验等目标。



代表客户&项目

广发银行；上海银行；中国人民银行

嘉实基金

广发证券

广州市人民政府；国家税务总局；国家外汇管理局

中国移动；中国电信

银行

基金

证券

政务

运营商

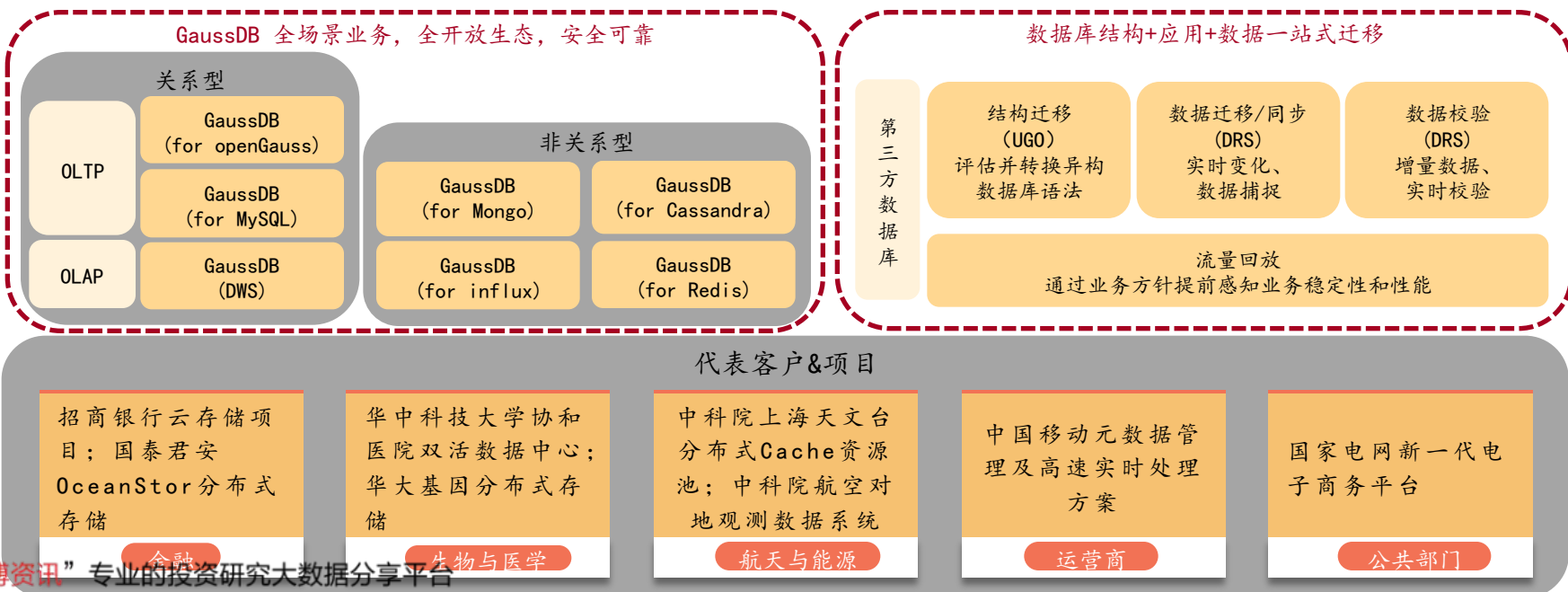
“慧博资讯”专业的投资研究大数据分享平台

3.6 华为GaussDB：将AI能力植入数据库内核 GaussDB

华为GaussDB是一个企业级AI-Native分布式数据库，将AI能力植入到数据库内核的架构和算法中，智能诊断、索引推荐，为用户提供更高性能、更高可用、更多算力支持的分布式数据库。GaussDB系列基于统一架构，底层选用分布式存储，中间层是每个DB特有的数据结构，上层是各个生态接口，支持关系型与非关系型数据库引擎，开启数据库极速与多模融合时代。

GaussDB全栈产品和解决方案，让客户选型安心、迁移放心、超高可用、自由扩展、开放生态、自主研发、安全可靠。

- 选型安心：自主生态+开源生态，SQL+NoSQL，分布式统一的架构，性能卓越，安全可靠；
- 迁移放心：从数据的在线迁移到应用的结构转换，迁移无忧；
- 超高可用：支持同城跨机房、异地、异地多活高可用，支持分布式强一致，数据0丢失；
- 自由扩展：GTM-Lite技术打破传统分布式性能瓶颈，计算与存储可自由水平扩展，支持1000+扩展能力；
- 超高性能：分布式优化器+Numa-Aware技术加持，支持PB级海量存储，在TPC-C、TPC-H等多种企业级负载下性能卓越；
- 开放生态：打造openGauss开放生态，基于openGauss生态的客户可以避免被闭源数据库锁定的风险，最大化保护客户资产；
- 自主研发：基于数据库领域10年+积累，核心代码100%知识产权，产品所有模块都具备代码级的问题修复及架构优化能力；
- 安全可靠：透明加密，防篡改，WDR诊断报告；智能诊断、索引推荐，有效提升客户开发运维效率。



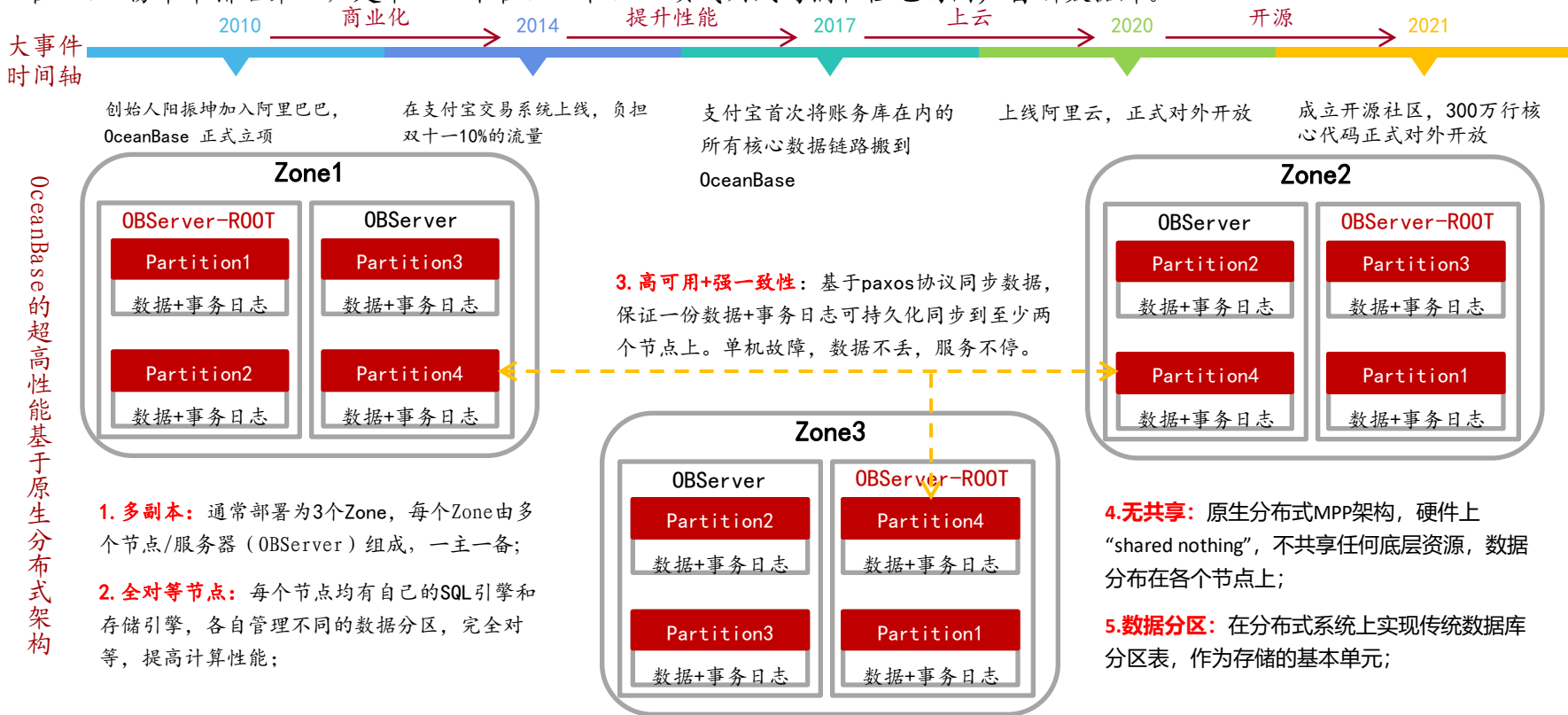
“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

资料来源：GaussDB公司官网、墨天轮、招商证券

3.7 蚂蚁OceanBase：分布式数据库的极致性OCEANBASE

OceanBase是由蚂蚁集团完全自主研发的企业级分布式关系数据库。OceanBase始创于2010年，具有数据强一致、高可用、高性能、在线扩展、高度兼容SQL标准和主流关系数据库、低成本等特点。据国际事务处理性能委员会 (TPC) 测评，OceanBase在TPC-H榜单中排名第一，是唯一一个在OLTP和OLAP领域测试均摘取桂冠的国产自研数据库。



代表客户&项目

中国工商银行两地三中心城市级容灾；南京银行互联网金融核心账户系统

银行

中国人民保险核心运算处理DB架构；中华保险全新分布式架构

保险

招商证券历史收益系统

证券

中国移动政企网格智慧运营系统

运营商

中国石化加油卡业务内部支持交易流水系统

公共部门

“慧博资讯”专业的投资研究大数据分享平台

3.8 PingCAP: 向云而生, 拥抱开源



PingCAP成立6年内已完成D轮融资, 投后规模超过30亿美元, 发展迅速。平凯星辰(北京)科技有限公司(以下简称PingCAP)成立于2015年, 是一家企业级开源分布式数据库厂商, 致力于为全球用户提供稳定高效、安全可靠、开放兼容的新型数据服务平台、解决方案与咨询、技术支持与培训认证服务, 解放企业生产力, 加速企业数字化转型升级。PingCAP在帮助企业释放增长空间的同时, 也提供了一份具有高度可参考性的开源建设实践样本。

整合K8S, 转向云原生, 降低用户成本。 PingCAP在2020年发布TiDB Cloud产品, 提供开箱即用的云托管服务; 同时其分布式架构天然支持云原生, 无缝衔接K8S, 可以轻松部署在任何公有云、私有云和混合云之上, 计算和存储分离, 按需使用、规模化扩展, 降低用户成本, 提升资源利用率。为提升系统安全程度, PingCAP通过专用的云主机、云上网络企业级加密。

独行快, 众行远, 贯彻执行开源, 兼容MySQL协议和生态。 PingCAP自2015年于GitHub上创建, 开放全部源代码、开发体系和运营体系, 目前标星人数已经超过4万人, 坚持走开源路线使其收获了美团、小米等众多实力雄厚的生态伙伴, 提高海外知名度和接受度; 高度兼容MySQL的设计也使得数据迁移更加简单。针对社区反馈, PingCAP在6年间连续升级5个版本, 满足用户多样化需求。开源社区同时也是公司最好的面试官, 为企业甄别了优秀的员工候选人。

初创期: TiDB1.0 (2015-2017)

1. 2015年在GitHub上创建, 2017年发布GA版本, 源代码全部开源;
2. 基于Google的spanner模型研发, 分布式强一致性事务、在线弹性水平扩展、故障自恢复的高可用、跨数据中心多活;
3. 兼容MySQL, 支持NO_SQL_CHANCE语法, 控制存储引擎对缓存的使用;

发展期: TiDB2.0-4.0 (2018-2020)

1. 一键安装部署, 方便操作;
2. 支持日志搜索和性能分析;
3. 增加TiFlash节点, 基于ClickHouse实现高效的列式计算, 支持OLAP场景应用;
4. 云原生, 支持容器化部署;

成熟期: TiDB5.0 (2021-)

1. 在TiFlash节点引入了MPP架构, 支持实时HTAP;
2. 引入聚簇索引功能, 提升数据库的性能;
3. 开启异步提交事务功能, 降低写入延迟;
4. 通过提升优化器的稳定性及限制系统任务对I/O、网络、CPU、内存等资源的占用, 降低系统的抖动。

结合新硬件、整合K8S、云原生

MPP架构、聚簇索引、高可用、安全性

代表客户&项目

中国银行新一代运维监控平台; 光大银行; 微众银行

中通快递、快成物流

美团、58集团、汽车之家、易车

小红书、多点、小米

爱奇艺、虎牙直播、BIGO LIVE

金融

物流

生活

零售

视频

“慧博资讯”专业的投资研究大数据分享平台

点击进入 <http://www.hibor.com.cn>

资料来源: PingCAP公司官网、招商证券

分析师承诺

负责本研究报告的每一位证券分析师，在此申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与，未来也将不会与本报告中的具体推荐或观点直接或间接相关。

刘玉萍：计算机行业首席分析师，北京大学汇丰商学院金融学硕士。优势领域云计算，2020年入围水晶球最佳分析师。

周翔宇：招商证券计算机行业分析师，三年中小盘研究经历，获得2016/17年新财富中小市值团队第五、第二名。

孟 林：招商证券计算机行业分析师，中科院信息工程研究所硕士，两年四大行技术部工作经验，两年一级市场投资经验，2020年加入招商证券研究所。

感谢实习生**康宇璇**对本文的贡献！

投资评级定义

公司短期评级

以报告日起6个月内，公司股价相对同期市场基准（沪深300指数）的表现为标准：

强烈推荐：公司股价涨幅超基准指数20%以上

审慎推荐：公司股价涨幅超基准指数5-20%之间

中性：公司股价变动幅度相对基准指数介于±5%之间

回避：公司股价表现弱于基准指数5%以上

公司长期评级

A：公司长期竞争力高于行业平均水平

B：公司长期竞争力与行业平均水平一致

C：公司长期竞争力低于行业平均水平

行业投资评级

以报告日起6个月内，行业指数相对于同期市场基准（沪深300指数）的表现为标准：

推荐：行业基本面向好，行业指数将跑赢基准指数

中性：行业基本面稳定，行业指数跟随基准指数

回避：行业基本面向淡，行业指数将跑输基准指数

重要声明

本报告由招商证券股份有限公司（以下简称“本公司”）编制。本公司具有中国证监会许可的证券投资咨询业务资格。本报告基于合法取得的信息，但本公司对这些信息的准确性和完整性不作任何保证。本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价，在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。除法律或规则规定必须承担的责任外，本公司及其雇员不对使用本报告及其内容所引发的任何直接或间接损失负任何责任。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突。

本报告版权归本公司所有。本公司保留所有权利。未经本公司事先书面许可，任何机构和个人均不得以任何形式翻版、复制、引用或转载，否则，本公司将保留随时追究其法律责任的权利。





感谢您宝贵的时间
Thank You

“慧博资讯”专业的投资研究大数据分享平台

点击进入  <http://www.hibor.com.cn>