



中信证券研究部

核心观点



杨泽原
计算机行业首席
分析师
S1010517080002



丁奇
云基础设施行业
首席分析师
S1010519120003



马庆刘
云应用分析师
S1010522090001

计算机行业

评级

强于大市（维持）

安全可控势在必行，数据库国产替代加速开展，以党政为代表的国产替代先行，并不断向金融、电信等领域拓展。通过对关键三个问题的深度解读，我们认为：1) 国产厂商已具备较强的 OLTP 数据库替代能力；2) OLAP 数据库未来仍将会持续快速发展，技术能力强者有望突围；3) 数据库行业朝着多元化方向演进，包括非关系型数据库、分布式数据库、HTAP 混合负载、云数据库等，与传统数据库类型形成良好的市场互补关系。数据库产业料将持续快速成长，带来诸多投资机遇，产品技术能力强、生态客户圈广、行业趋势把握准确者有望胜出。

■ 问题一：OLTP（联机事务处理）数据库的国产替代能力如何？

■ 回答一：关注产品性能、稳定性、安全性、兼容性等维度，国产厂商已具备较强的替代能力。20 世纪 70 年代，关系型数据库原型 System R 诞生，Oracle、IBM DB2、SQL Server 等主流关系型数据库崛起，并逐步进入中国市场，长期占据主要份额。海外巨头的产品具有优越的性能、独立的编程方言、广泛的应用生态等，如 Oracle 的 RAC 技术，大幅提升数据库性能、可用性、扩展性，提高了集群的事务处理能力。从替代能力来看，国产主流产品的性能指标已与海外龙头接近，外部扰动及政策支持下，信创需求加速释放，促使产品性能进一步提升，并正逐步从办公管理系统向核心业务系统拓展，加速打开市场空间。IDC 数据显示，2021 年我国本地部署关系型数据库市场份额 Top3 分别为 Oracle、华为、达梦，后两者市场份额超过了微软、IBM 等海外 IT 巨头，人大金仓、阿里巴巴等国产厂商亦有所斩获，国产厂商已具备较强的替代能力。

■ 问题二：OLAP（联机分析处理）数据库的发展到了什么阶段？

■ 回答二：聚焦分析处理，重视湖仓一体、智能湖仓等方向，技术能力强者有望胜出。OLAP（联机分析处理）的概念最早由关系模型之父埃德加·科德于 1993 年提出，主要支持面向分析场景的应用，用于业务反馈和辅助决策。OLAP 先后衍生出数据仓库、数据湖的发展路径，数据仓库聚焦于结构化数据处理能力，数据湖聚焦于数据完整性的实现，支持对于各类半结构化数据、非结构化数据的存储和分析。为了兼顾数据分析效率和数据完整性，近年来众多厂商开始进行湖仓一体、智能湖仓的探索。我们认为，分析型数据库正呈现出决策实时化、场景精细化、产品标准化等特点，需求的持续推动促使行业快速发展，带来诸多机遇，参与者如 Snowflake、Databricks、Cloudera、阿里巴巴、腾讯、星环科技等百花齐放，竞相成长，技术能力强者有望胜出。

■ 问题三：如何看待国产数据库的创新方向？

■ 1) 看架构：分布式&集中式——分布式成重要趋势，云数据库进一步打开市场。分布式数据库具有数据计算量大、运算速度快等特点，是数据库发展的重要趋势，但目前具有成本高、管理复杂等问题，与传统集中式数据库优势互补。云数据库进一步打开市场空间，IDC 数据显示，2021 年我国关系型数据库中，公有云市场规模增速已超过本地部署增速，IDC 预计未来 3 年云数据库的市场规模增速有望保持在~40%，2025 年市场规模有望达到 55 亿美元。

■ 2) 看模型：关系型&非关系型——相互补充，重视多模型能力构建。非关系型数据库核心解决半结构化、非结构化数据问题，智研咨询数据显示，2018 年我国关系型数据库市场规模占比达 85%，但呈现逐年下降趋势。中国信通院数据显示，截至 2021 年 6 月，我国关系型数据库产品数量占比约 60%。非关系型与关系型数据库相互补充，具备多模型兼容能力者具备更强竞争力。

- **3) 看生态：开源&闭源——并存发展，共促商业化生态繁荣。** 开源与商业化并不矛盾，商业化的数据库亦可以开源，如 MongoDB、Spark SQL、TiDB 等开源的商业化数据库，通过开源提升自身的影响力、促进代码技术的提升、构建更大的生态圈等，但开源也对安全可控带来一定考验。DB-engines 数据显示，目前开源数据库产品数量已超过 50%。我们认为，未来国内数据库厂商将呈现出开源和闭源共存的状态，共促商业化生态的繁荣。
- **风险因素：**国产数据库技术迭代不及预期；下游行业需求不及预期；数据库行业参与者众多，市场竞争加剧的风险；开源或免费软件的竞争风险；疫情及宏观环境扰动。
- **投资策略：**数据库是数字化时代的底层基石，是各厂商重点布局的关键战略领域。安全可控势在必行，数据库国产替代加速开展，以党政为代表的国产替代先行，并不断向金融、电信等领域拓展。通过对关键三个问题的深度解读，我们认为：
 - 1) 国产厂商已具备较强的 OLTP 数据库替代能力；
 - 2) OLAP 数据库未来仍将会持续快速发展，技术能力强者有望突围；
 - 3) 数据库行业朝着多元化方向演进，包括非关系型数据库、分布式数据库、HTAP 混合负载、云数据库等，与传统数据库类型形成良好的市场互补关系。数据库产业料将持续快速成长，带来诸多投资机遇，产品技术能力强、生态客户圈大、行业趋势把握准确者有望胜出。**建议关注：**
 - 1) 已上市及拟上市：**达梦数据（中国软件）、星环科技、人大金仓（太极股份）、海量数据、神舟通用（航天软件）、科蓝软件。**
 - 2) 未上市公司：PingCAP、巨杉数据库、南大通用、易鲸捷、云和恩墨等。

目录

报告亮点及创新之处	6
问题一：OLTP 数据库的国产替代能力如何？	7
核心聚焦：关系型 OLTP 数据库是国产替代的主要对象	7
替代挑战：海外巨头在产品性能、编程方言、应用生态等维度具备优势	9
替代能力：国产数据库已具备较强的替代能力	13
问题二：OLAP 数据库的发展到了什么阶段？	17
并驾齐驱：OLAP 成为继 OLTP 之后数据库的下一发展重心	17
技术复盘：把握数据处理效率、数据完整性两条发展主线	20
需求推演：决策实时化、场景精细化、产品标准化	24
问题三：如何看待国产数据库的创新方向？	27
看架构：分布式成重要趋势，云数据库打开更大市场	27
看模型：关系型与非关系型长期共存，重视多模型能力构建	30
看生态：开源闭源并存发展，共促商业化生态繁荣	33
风险因素	36
投资建议	37

插图目录

图 1：一张图看懂数据库发展路径	8
图 2：Oracle RAC 架构增强数据处理能力	10
图 3：Oracle Data Guard 用于容灾提升安全性	10
图 4：SQL 语言的特点	11
图 5：部分主流关系型数据库“方言”示例	11
图 6：TPC-C 基准测试全球数据库历史前 10 名（按照每分钟处理任务数排序）	13
图 7：达梦数据库具备高性能 OLTP 能力	14
图 8：人大金仓 KingbaseES 主要性能优势	14
图 9：信创核心品类有望进一步向区县层面下沉	15
图 10：行业信创推进顺序	15
图 11：达梦数据库客户案例	16
图 12：2021 年全球数据库市场份额	16
图 13：2021 年我国本地部署关系型数据库市场份额	16
图 14：数据治理赋能企业各个价值链环节	18
图 15：国内外分析型数据库市场代表厂商	19
图 16：2020、2021 年全球数据库魔力象限	19
图 17：数据仓库架构及优缺点	21
图 18：MPP（大规模并行处理）架构	21
图 19：数据湖架构及优缺点	22
图 20：Hadoop 技术生态体系	23
图 21：全球数据量预测（ZB）	23
图 22：我国企业潜在 AI/ML 应用场景数目调研	23
图 23：湖仓一体架构图	24
图 24：亚马逊智能湖仓架构图	24
图 25：TP/AP 分离数据库应用架构数据流	25
图 26：数据价值随实时性递减	25
图 27：数据分析需求在部分行业深化方向	26
图 28：数据中台产品技术 SaaS 化	26
图 29：硬件架构有中 4 种不同层次的数据共享	28
图 30：2020-2026 中国关系型数据库市场规模（按部署方式分类）	30
图 31：我国数据库产品分布情况（截至 2021 年 6 月）	32
图 32：各机构关于我国数据库及关系型数据库市场增速预测	32
图 33：常见开源许可证一览	34
图 34：全球开源数据库与商业数据库数量占比	35
图 35：开源数据库与商业数据库成本比较	35

表格目录

表 1：关系型数据库与非关系型数据库的对比	7
表 2：OLTP 和 OLAP 的对比	8
表 3：主流海外数据库品类分类一览	9

表 4: Oracle 发展历史.....	12
表 5: Oracle 产品线	12
表 6: 达梦数据主要上游生态合作伙伴	15
表 7: 2019-2021 年我国国产数据库管理软件市占率排名	17
表 8: 数据仓库与传统数据库对比	17
表 9: 不同技术架构分析型数据库关键指标	20
表 10: Hadoop 与 MPP 差异	22
表 11: 非 HTAP 架构与两种 HTAP 方案对比	25
表 12: 传统集中式数据库与新兴分布式数据库对比	27
表 13: 三种分布式架构的优劣势分析	28
表 14: 分布式数据库技术路线对比	29
表 15: 存算一体 vs 存算解耦	29
表 16: 云数据库的机遇与挑战	30
表 17: 四种非关系型数据库对比	31
表 18: 2022 年 10 月数据库管理系统流行程度排名及环比、同比变化情况	31
表 19: 星环大数据基础平台 (TDH) 与竞争对手支持数据模型对比	33
表 20: 不同多模型实现路径	33
表 21: 国产数据库开源情况	36

■ 报告亮点及创新之处

本报告以市场上核心关注的三个数据库行业问题为抓手，创新性的展开对数据库行业的讨论与分析，帮助读者重点理解当前数据库行业的核心矛盾，并梳理了对应的参与公司与建议关注的投资机遇。具体内容如下：

1) OLTP 数据库国产厂商替代能力探究

基于数据库产业发展历史的回顾，明确关系型 OLTP 数据库是目前国产替代的主要对象，从产品性能、编程方言、应用生态维度梳理海外巨头所具备的优势以及国产厂商面临的挑战，从性能、生态和市场的角度分别论证数据库国产替代的能力几何。

2) OLAP 数据库的技术复盘与格局推演

我们认为 OLAP 与 OLTP 并行发展是数据库行业重要趋势，并从供需角度分析 OLAP 增长动因。系统梳理 OLAP 领域从数据仓库、数据湖到湖仓一体的技术发展演进，总结各阶段技术架构与需求痛点。基于 OLAP 需求场景，我们认为 OLAP 数据库正在朝着决策实时化、场景精细化、产品标准化的方向发展。

3) 分析了国产数据库行业的创新发展方向

提出从架构、模型和生态三个维度看待国产数据库未来的创新发展方向。在架构维度，分布式与云将持续贡献增量；在模型维度，关系型与非关系型将长期共存，多模型兼容能力者有望胜出；在生态维度，预计国产厂商将积极拥抱开源建设，补充自身技术能力与生态，同时兼顾自主可控与商业化需求，打开更大市场。

■ 问题一：OLTP 数据库的国产替代能力如何？

核心聚焦：关系型 OLTP 数据库是国产替代的主要对象

产品分类：从需求的角度可将数据库分成以下两种——关系型数据库和非关系型数据库、OLTP 数据库和 OLAP 数据库。

1) 按数据模型分类：关系型数据库和非关系型数据库

关系型数据库是一种典型的数据库类型，采用关系模型，常用行和列等二维的形式来存储结构化数据，一系列的行和列被称为表，一组表组成了一个数据库。表的每一行称为一个元祖（Tuple），代表了一组值之间的联系；每一列称为一个属性（Attribute）或字段（Field），是对实体的具体描述，每一列的数据类型相同。关系模型凭借原子性、一致性、隔离性和持久性的 ACID 特性，取代层次、网状模型成为当代主流数据模型。

非关系型数据库是用非关系模型，存储非结构化的如图像、音视频等类型数据的数据库，分为列存数据库、键值数据库、文档数据库、图数据库等多种类别。随着 web2.0 的兴起海量半结构化、非结构化数据出现，非关系型数据库应运而生。

表 1：关系型数据库与非关系型数据库的对比

类型	关系型数据库	非关系型数据库
特性	关系型数据库，是指采用了关系模型来组织数据的数据库。关系模型指的就是二维表格模型，而一个关系型数据库就是由二维表及其之间的联系所组成的一个数据组织。	1、有键值存储数据库、列存储数据库、文档型数据库、图形数据库等多种分类；2、一般不支持 ACID 特性；
优点	1、容易理解：关系模型相对网状、层次等其他模型来说更容易理解；2、使用方便：关系型数据库使用 SQL 语言非常方便；3、易于维护：丰富的完整性减低了数据冗余和数据不一致的概率。	1、无需经过 SQL 层的解析，读写性能很高；2、数据容易扩展；3、存储数据的格式多样化：存储格式是 K-V 形式、文档形式、图片形式等等，而关系型数据库则只支持基础类型。
缺点	1、处理高并发读写性能较差；2、表结构较为固定，不便于更改扩展；	1.不提供 SQL 支持；2.一般来说无事务处理

资料来源：OSCHINA，中信证券研究部

2) 按应用类型分类：OLTP 和 OLAP

OLTP（On-Line Transaction Processing，操作型数据库，又称联机事务处理）主要关注一段时间内的实时数据，基本特征是接收的用户数据可以立即传送到计算中心进行处理，并在很短的时间内给出处理结果，是对用户操作快速响应的方式之一。OLTP 主要使用关系模型，用户多为一线业务人员，支持高并发、实时快速增删查改，典型应用场景包括金融交易、互联网电商等。

OLAP（On-Line Analysis Processing，分析型数据库，又称联机分析处理）主要是分析长期数据的规律走势，多应用于决策。OLAP 使用的数据对象不限于关系模型，用户多为分析师或管理层，支持对于历史数据的分析操作，典型应用场景包括风险预警、商业分析、辅助决策等。伴随企业信息系统大量业务数据的产生，从不同类型的数据中提取

出对企业决策分析有用的信息这一需求日渐显现。

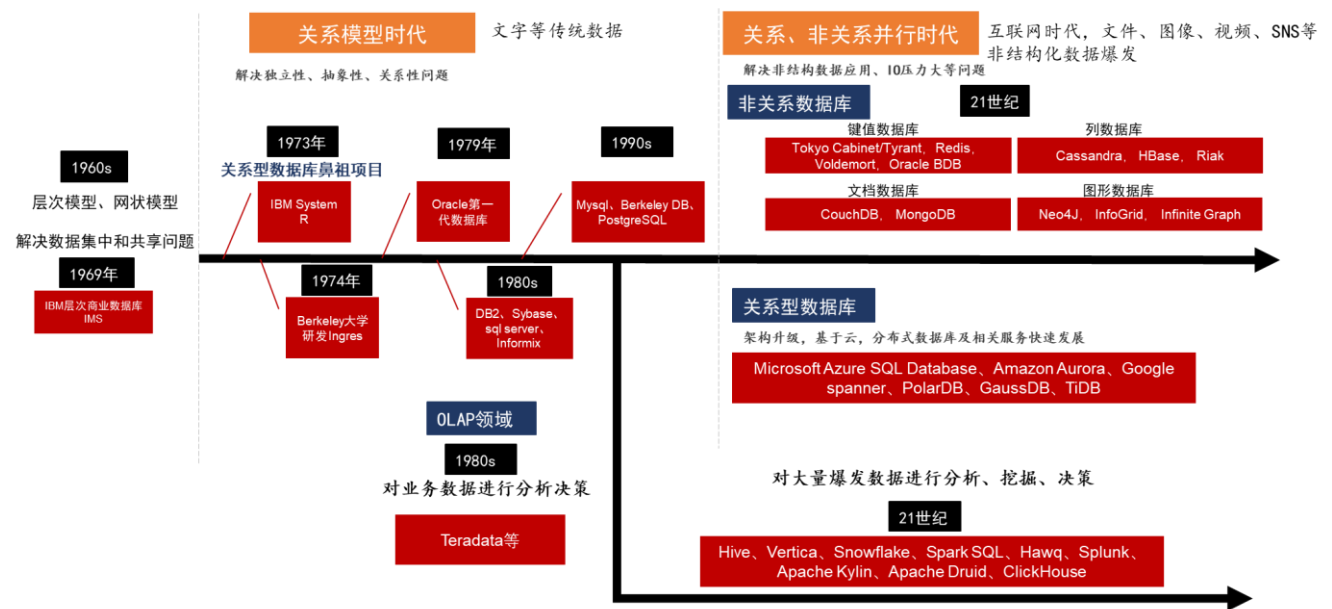
表 2: OLTP 和 OLAP 的对比

	OLTP	OLAP
本质	在线交易系统	在线数据检索和分析系统
数据处理操作	对数据库进行增、删、改、查	对数据记录进行选择
数据表	标准化	未标准化
数据源	OLTP 以及 OLTP 的处理记录	不同的 OLTP 是 OLAP 的数据源
数据量大小	实时数据, 数据量小	聚合历史数据, 数据量大
响应时间	毫秒级	秒、分钟或小时 (取决于处理的数据量)
存储方式	数据存储在传统的数据管理系统	数据存储和数据仓库
特点	方便创建数据表以及方便维护	引导用户在电子表格的帮助下创建视图进行分析
用途	关注一段时间内的实时数据	分析长期数据的规律走势, 为决策提供帮助
典型适用场景	快速处理高并发、小批量数据	使用复杂查询处理大量数据
主要用户	银行柜员、收银员、仓库管理员等	数据分析师、业务分析师、高管等

资料来源: Giant Tech, 艾瑞咨询, 中信证券研究部

发展历史: 国外数据库厂商相对于国内厂商早起步 20-30 年。国内厂商中, 如今占据国内市场份额较多的达梦数据成立于 2000 年, 南大通用成立于 2004 年, 而国外的 IT 巨头在上个世纪便已经在这一领域进行研究发展, 以 Oracle、IBM、微软为代表的海外 IT 巨头的相关产品于 20 世纪 80 年代末开始进入中国。先发优势带来的技术领先和客户粘性是如今国外厂商仍然占据国内数据库市场主要份额的重要原因。

图 1: 一张图看懂数据库发展路径



资料来源: 各公司官网, 中信证券研究部绘制

20 世纪 60-70 年代, 关系模型快速发展, 关系型数据库可解决数据存储的易用性、抽象性、独立性问题, 拉开了关系型数据库软件革命的序幕。1970 年, IBM 公司的研究员埃德加·科德在 Communications of ACM 上发表论文《A Relational Model of Data for Large Shared Data Banks》, 在层次模型和网状模型的数据库产品在市场上占主要位置

的时代，拉开了关系型数据库软件革命的序幕。

IBM 在 1973 年启动了 System R 项目来研究关系型数据库的实际可行性，各方关系型模型支持者吸取该项目经验，进行关系型数据库研发。1977 年，Oracle 创始人 Larry Ellison 与 Bob Miner 和 Ed Oates 在硅谷共同创办了一家名为软件开发实验室的计算机公司（Oracle 前身），开始进行关系型数据库的研发，同时期 Berkeley 大学也在进行关系数据库系统 Ingres 的开发。IBM 虽然 1973 年就启动了 System R 项目来研究关系型数据库的实际可行性，但是并没有及时推出这样的产品，因为当时 IBM 的 IMS（著名的层次型数据库）市场较好，公司当时认为，如果推出关系型数据库，会是对另一款产品的颠覆。

80-90 年代，大量数据库公司吸取关系模型经验，逐步推出自己的产品。1983 年，IBM 发布商业版数据库 DB2。1984 年，Sybase 公司成立，创始人之一 Bob Epstein 是 Ingres 大学版（与 System R 同时期的关系数据库模型产品）的主要设计人员。1988 年，微软推出 SQL Server，主要适配自身 Windows 生态，这个时期，Oracle 因为客户需求已经使用 C 语言开发出适用于多个系统版本的数据库产品。90 年代，MySQL、PostgreSQL 等开源版本数据库陆续发布。

表 3：主流海外数据库品类分类一览

	关系型	非关系型
OLTP	传统巨头：Oracle、 MySQL 、IBM DB2、SQL Server、SAP Adaptive Server、Microsoft Access 云计算巨头：Microsoft Azure SQL Database、Amazon Aurora、Google spanner 独立厂商： PostgreSQL 、 MariaDB 、 SQLite 、FileMaker、 Firebird	传统巨头：Oracle NoSQL、IBM Cloudant、 OrientDB 云计算巨头：Amazon DynamoDB、Microsoft Azure Cosmos DB、 RocksDB 、 LevelDB 独立厂商： MongDB 、 Redis 、 Cassandra 、 Neo4j 、 Hbase
OLAP	传统巨头：Teradata、Oracle Essbase、SAP IQ 云计算巨头：Google BigQuery、Amazon Redshift、Microsoft Azure Synapse Analytics 独立厂商： Hive 、Vertica、Snowflake、 Spark SQL 、 HAWQ	独立厂商： Elasticsearch 、 Splunk 、 Apache Kylin 、 Apache Druid 、 ClickHouse

资料来源：各公司官网，中信证券研究部 注：加粗为开源数据库

国产替代：重点关注海外 IT 巨头先入为主的关系型 OLTP 数据库的存量市场。外部确定因素扰动下，安全可控势在必行，数据库国产替代加速开展，以党政为代表的国产替代先行，并不断向金融、电信等领域拓展。纵观海内外数据库行业近 70 年的发展史，我国自上世纪 80 年代开始相关技术研发、21 世纪初开始逐步迈入成熟的商业化进程，整体进度落后于海外巨头 20 余年，导致在传统关系型 OLTP 数据库领域海外巨头占据主要市场份额。而后较为新兴的非关系型领域、OLAP 领域由于需求的碎片化以及云厂商和独立厂商的角力，加上国产数据库厂商紧紧跟随行业发展的步伐，海内外新兴数据库市场呈现出百花齐放的态势，海外厂商在新型数据库领域并不具备绝对的技术迭代优势和市场份额优势。因此数据库国产替代首先重点关注传统的关系型 OLTP 数据库的存量市场。

替代挑战：海外巨头在产品性能、编程方言、应用生态等维度具备优势

我们认为，海外 IT 巨头在数据库领域能够经久不衰的原因主要体现在优越的产品性能、独立的编程方言和广泛的应用生态等维度。这亦是数据库国产替代所面临的主要挑战，是探究国产数据库能否完成替代的重要关切。

1) **技术领先，性能加持**

数据库产品最重要的指标之一是性能，以海外数据库龙头 Oracle 为例，其产品在安全性、可伸缩性和并行性、兼容性、开放性等维度具备出众优势。

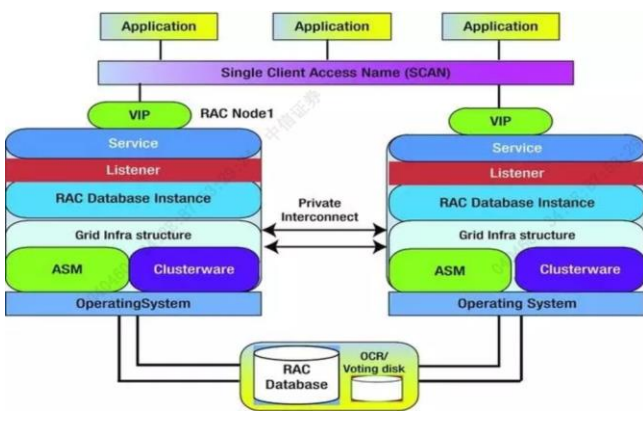
安全性方面，Oracle 的安全机制得到 17 家独立安全评估机构的认可，获得最高认证级别的 ISO 标准认证。Oracle Data Guard 是 Oracle 的高可用性数据库方案，主要功能是数据保护、数据容灾。Oracle Data Guard 在主节点和备用节点之间通过日志同步来保证主数据库与备用数据库之间数据的同步，实现数据库的快速切换和故障恢复，最大程度保护数据库的安全。

可伸缩性和并行性方面，Oracle 的服务器通过使一组结点共享同一簇中的工作来扩展，提供高可用性和高伸缩性的解决方案；Oracle 产品拥有 RAC 等数据库领域的硬核技术。Oracle RAC (Real Application Clusters)是 Oracle 的一项支持网格计算环境的关于应用集群的核心技术。在一个应用环境中，让多个服务器来管理同一个数据库，分散了每一台服务器的工作量。Oracle RAC 的技术大幅提升架构的可用性、性能、扩展性，即使某些实例宕机，也能维持系统正常工作；提高集群的事务处理能力，使得多个实例能够并发工作；能通过增加节点提高数据库的性能。

兼容性方面，Oracle Database 可以在 Windows、Unix、DOS 等多个系统上工作，没有 SQL Server 只能在 Windows 系统上运行的局限性，同时支持包括 TCP/IP、DECnet 在内的多种协议，可以与多种通讯网络连接。

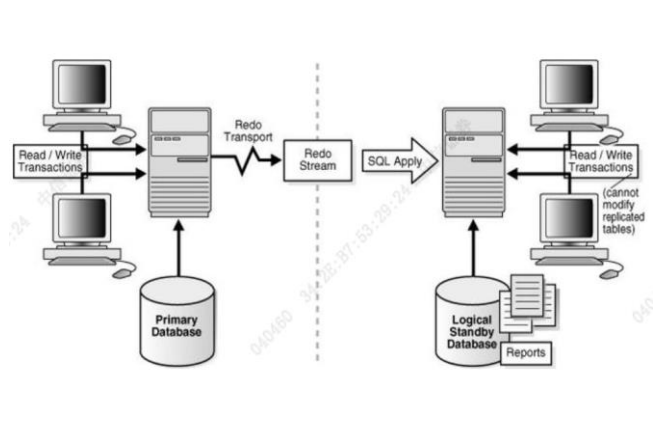
开放性方面，Oracle 的底层使用 C 语言开发而成，随着不断发展在开发中也加入 Java 语言和技术标准，并支持绝大多数编程语言，相比之下 SAP 等竞争对手均只支持几种编程语言，与其他技术与平台的兼容度低于 Oracle。

图 2：Oracle RAC 架构增强数据处理能力



资料来源：51CTO

图 3：Oracle Data Guard 用于容灾提升安全性



资料来源：Oracle 公司官网

2) **独立编程方言，提升用户粘性**

SQL 作为关系型数据库的标准语言，具备移植性强、简洁易用等优势。SQL 全称 Structured Query Language，是用于定义、查询、修改和管理关系型数据库的结构化查询

语言。1970 年 IBM 公司研究员埃德加·科德在其发表的论文《A Relational Model of Data for Large Shared Data Banks》中首次描述了关系模型，SQL 是对关系模型的第一个商业化语言实现，并于 1986 年成为美国国家标准学会（ANSI）的一项标准，在 1987 年成为国际标准化组织（ISO）标准。作为一种高度非过程化的编程语言，SQL 同时具备扩展性强和简洁易用的优势，它允许用户在不指定对数据的存放方法和不了解具体数据存放方式的情况下在高层数据结构上进行工作。

图 4：SQL 语言的特点

语言功能一体化	高度非过程化的编程语言	面向集合的操作	自含式+嵌入式语言	简洁易用
SQL 语言集数据定义、数据操纵、数据控制、事务控制功能于一体，在一次操作中可以使用任何语句，为数据库开发提供了良好的环境。	SQL 语言不涉及存取结构以及具体的执行过程，因而简化了编程的复杂性，提高了数据的独立性。	SQL 语言的操作对象和输出结果都是元组集合，可以实现对一组记录的增删查改。	SQL 作为自含式语言，可作为联机交互使用，每个 SQL 语句可以独立完成其操作；作为嵌入式语言，SQL 语句可以嵌入到高级程序语言中使用。	SQL 用简单的语句（核心命令：CREATE、DROP、ALTER、SELECT、DELETE、INSERT、UPDATE、GRANT、REVOKE）和接近英语口语的语法完成了数据库创建管理维护的复杂功能。

资料来源：艾瑞咨询，中信证券研究部

各家数据库产品在落地应用过程中逐渐形成 SQL “方言”，以解决标准语言无法解决的问题，提高了用户黏性，形成竞争壁垒。在商业实践中，由于各家数据库产品的数据源不同、应用场景不同、用户需求不同，众多数据库厂商均开始尝试在标准 SQL 基础上提供自己特有的功能，以提高用户的便捷性。不论是数据库龙头 Oracle、微软 SQL Server、IBM DB2 还是开源框架 MySQL、PostgreSQL，都逐渐形成了自己的 SQL “方言”，这大大提高了不同主流数据库产品之间的替换成本。同时，以 Oracle 为代表的全球数据库巨头不断完善自身产品生态，通过收购 MySQL 等途径提高自身在开源社区的影响力和话语权。持续提升的用户黏性帮助海外 IT 巨头实现对于传统数据库市场的垄断。

图 5：部分主流关系型数据库“方言”示例

	PostgreSQL	MySQL	Microsoft SQL Server	Google BigQuery Standard SQL
How to query fields?	SELECT field1, field2	SELECT field1, field2	SELECT [field1], [field2]	SELECT field1, field2
How to use aliases?	SELECT field1 AS f1	SELECT field1 AS f1	SELECT field1 = f1	SELECT field1 AS f1
Which quotation marks are accepted?	field = 'correct'	field = 'correct' OR field = "correct"	field = 'correct'	field = 'correct' OR field = "correct"
Is the language case sensitive?	YES: field = 'correct' is not the same as field = 'Correct'	NO: field = 'correct' is the same as field = 'Correct'	YES: field = 'correct' is not the same as field = 'Correct'	YES: field = 'correct' is not the same as field = 'Correct'
How to create date fields?	CURRENT_DATE() CURRENT_TIME() EXTRACT()	CURDATE() CURTIME() EXTRACT()	GETDATE() DATEPART()	CURRENT_DATE() CURRENT_TIME() EXTRACT()

资料来源：Towards Data Science@Marie Lefevre

3) 产品快速迭代, 完善应用生态

龙头数据库公司对于产品的更新换代较为积极, 能够产生较大的用户粘性, 使得市场份额优势持续。以 Oracle 为例, 在 Oracle9i 产品中引入网络(Internet)的特性, 在 Oracle10g 中加入网格计算(grid)的特性, 在 Oracle12c 中引入云(cloud)的概念, 不断让产品有新的突破。而通过每一次更新对于产品的漏洞进行及时修复、推出新的应用、优化产品的性能, 也都会吸引已有的用户持续使用这款产品。数据库的这些特征, 使其如同操作系统一样存在较强的用户粘性, 帮助行业龙头厂商迭代已建立的市场份额优势, 因此数据库行业是一个容易形成寡头的行业。

表 4: Oracle 发展历史

年份	事件
1977 年	埃里森与同事创立名为“软件开发实验室”(Software Development Labs)的公司
1979 年	推出 Oracle2, 计算机软件史上第一个由纯软件公司开发的商用关系型数据库管理系统
1986 年	Oracle 公司上市
1989 年	正式进入中国市场
1992 年	旗舰产品 Oracle7 面世
1997 年	推出面向网络计算的 Oracle8
2001 年	发布 Oracle9i, i 代表网络(Internet)
2003 年	发布 Oracle10g, 最大特点是加入网格计算的功能。版本号中的 g 代表网格(grid)。
2007 年	发布 Oracle11g, 实现信息生命周期管理(Information Lifecycle Management)等创新功能。
2010 年	收购了 MySQL 的所有者 Sun Microsystems
2013 年	发布 Oracle12c, c 代表云(cloud)
2018 年	发布 Oracle18c
2019 年	发布 Oracle19c

资料来源: Oracle 官网, 中信证券研究部

国外数据库公司注重技术创新和边界拓展, 不断获得用户黏性。以 Oracle 为例, Oracle 是第一个引入对象概念、多媒体等多种数据格式、并行技术、网格技术的数据库。作为数据库产品的标杆, Oracle 的 IT 布局十分完备, 开发的产品涵盖了行业管理软件、企业管理软件、中间件、数据库、操作系统、服务器、存储等多个领域。通过向上游基础设施和下游软件应用延伸产业链, 海外 IT 巨头得以进一步完善产品生态布局、提高基础技术实力, 从而持续稳固在数据库领域的龙头地位。

表 5: Oracle 产品线

数据库	中间件	应用	IT 基础设施
Database 19c 企业版/标准版	Application Testing Suite	E-Business Suite、PeopleSoft、JD Edwards、Siebel CRM	Grid Infrastructure
Database 18c 企业版/标准版	Data Integrator 和套件	AIA 流程集成	Oracle 闪存系统和 Pillar Axiom 插件
Database 12c 企业版/标准版	Enterprise Manager	HCM 云解决方案库	Oracle Linux 和 Oracle Enterprise Kernel
Database 11g 企业版/标准版	Content Management SDK	User Productivity Kit (UPK/UPK 专业版)	Solaris 10
MySQL	Enterprise Data Quality		Solaris 11
NoSQL 数据库	融合中间件		

资料来源: Oracle 官网, 中信证券研究部

替代能力：国产数据库已具备较强的替代能力

1) 性能为基：传统数据库领域，技术及性能可满足国产替代的要求

从 TPC-C 测试结果来看，OLTP 领域国产数据库在 TPC-C 等国际知名测试中性能已达到甚至赶超海外巨头水平。TPC 全称 Transaction Processing Performance Council，中文名称为事务处理性能委员会，是数据库性能测试的国际权威标准组织，目前拥有 20+ 成员公司，包括 Oracle、微软、IBM 等数据库领域 IT 巨头和华为、阿里、浪潮、柏睿数据等国产厂商。TPC-C 测试是衡量 OLTP 系统的工业标准，是行业中公认的权威和最为复杂的在线事务处理基准测试。它通过模拟仓库和订单管理系统测试 OLTP 数据库功能，包括查询、更新和队列式小批量事务，通过每分钟处理任务数 (tpmC) 衡量数据库性能。2020 年阿里云旗下 OLTP 数据库 OceanBase 以 7.07 亿 tpmC 的成绩登顶 TPC-C 测试历史榜首并延续至今，打破了 Oracle、IBM 等传统 IT 巨头对头部排名的垄断，反映了国产力量在 OLTP 领域已经达到较为领先的水平。

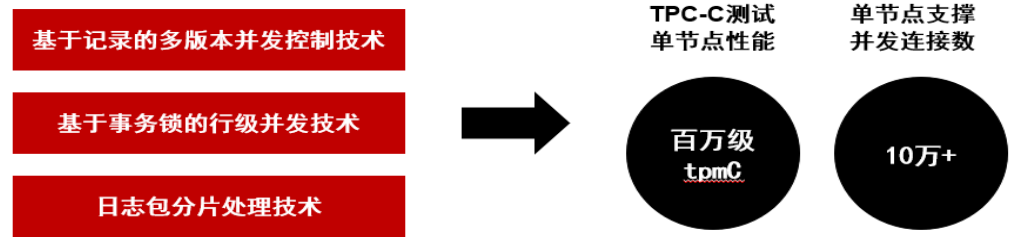
图 6：TPC-C 基准测试全球数据库历史前 10 名（按照每分钟处理任务数排序）

Hardware Vendor	System	v tpmC	Price/tpmC	Watts/KtpmC	System Availability	Database	Operating System	TP Monitor	Date Submitted
	Alibaba Cloud Elastic Compute Service Cluster	707,351,007	3.98 CNY	NR	06/08/20	OceanBase v2.2 Enterprise Edition with Partitioning, Horizontal Scalability and Advanced Compression	Alibaba Aliyun Linux 2	Nginx 1.15.8	05/18/20
	Alibaba Cloud Elastic Compute Service Cluster	60,880,800	6.25 CNY	NR	10/02/19	OceanBase v2.2 Enterprise Edition with Partitioning, Horizontal Scalability and Advanced Compression	Alibaba Aliyun Linux 2	Nginx 1.15.8	10/01/19
ORACLE	SPARC SuperCluster with T3-4 Servers	30,249,688	1.01 USD	NR	06/01/11	Oracle Database 11g R2 Enterprise Edition w/RAC w/Partitioning	Oracle Solaris 10 09/10	Oracle Tuxedo CFSR	12/02/10
	IBM Power 780 Server Model 9179-MHB	10,366,254	1.38 USD	NR	10/13/10	IBM DB2 9.7	IBM AIX Version 6.1	Microsoft COM+	08/17/10
ORACLE	SPARC T5-8 Server	8,552,523	.55 USD	NR	09/25/13	Oracle 11g Release 2 Enterprise Edition with Oracle Partitioning	Oracle Solaris 11.1	Oracle Tuxedo CFSR	03/26/13
ORACLE	Sun SPARC Enterprise T5440 Server Cluster	7,646,486	2.36 USD	NR	03/19/10	Oracle Database 11g Enterprise Edition w/RAC w/Partitioning	Sun Solaris 10 10/09	Oracle Tuxedo CFSR	11/03/09
	IBM Power 595 Server Model 9119-FHA	6,085,166	2.81 USD	NR	12/10/08	IBM DB2 9.5	IBM AIX 5L V5.3	Microsoft COM+	06/10/08
	Bull Escala PL6460R	6,085,166	2.81 USD	NR	12/15/08	IBM DB2 9.5	IBM AIX 5L V5.3	Microsoft COM+	06/15/08
ORACLE	Sun Server X2-8	5,055,888	.89 USD	NR	07/10/12	Oracle Database 11g R2 Enterprise Edition w/Partitioning	Oracle Linux w/Unbreakable Enterprise Kernel R2	Oracle Tuxedo CFSR	03/27/12
ORACLE	Sun Fire X4800 M2 Server	4,803,718	.98 USD	NR	06/26/12	Oracle Database 11g R2 Enterprise Edition	Oracle Linux w/Unbreakable Enterprise Kernel R2	Oracle Tuxedo CFSR	01/17/12

资料来源：TPC 官网 注：截止时间为 2022 年 10 月 14 日

国产 OLAP 数据库龙头厂商拳头产品性能已逐渐实现对于海外 IT 巨头的追赶。以达梦数据库为例，公司招股说明书显示，通过基于记录的多版本并发控制、基于事务锁的行级并发、日志包分片处理等大量先进性技术，公司产品具备优秀的并发事务处理性能。第三方软件测评实验室测试，单节点能够支撑数据库并发连接超过 10 万个；TPC-C 测试模型下，单节点性能可达百万级 tpmC，与海外主流 OLTP 产品 Oracle 11g、IBM DB2 9.5 性能达到同一数量级。

图 7：达梦数据库具备高性能 OLTP 能力



资料来源：达梦数据招股书，中信证券研究部

除了关注 OLTP 数据库基本的读写性能之外，国产厂商还高度重视产品可用性、稳定性、易用性、安全性等维度。人大金仓旗下拳头产品通用型关系型数据库 KingbaseES 实现对 97%以上 Oracle “方言”的兼容，便于用户实现低成本迁移；同时具备高稳定性、高可用性，标杆项目国家电网智能电网调度系统已实现 10 余年 7x24 稳定运行；易用性方面，KingbaseES 通过自研数据库辅助调优工具的应用，大大提高了性能诊断、辅助调优、故障修复等运维业务的效率；此外，KingbaseES 还通过了国家信息安全产品认证、Common Criteria EAL4+安全认证，达到主流产品 Oracle、SQL Server、IBM DB2 的安全级别。

图 8：人大金仓 KingbaseES 主要性能优势

<p>迁移难度低</p> <ul style="list-style-type: none"> 兼容97%+的Oracle语法，迁移平滑、成本更低 向导式智能数据迁移工具，无损、快速数据迁移 	<p>高稳定性，高可用性</p> <ul style="list-style-type: none"> 秒级RTO及全面容错体系，数据库7x24小时不间断运行 多层次高可用技术体系，系统可用性高达99.999% 	<p>高并发数，读写性能强劲</p> <ul style="list-style-type: none"> 读写分离集群，只读性能线性增长，承载“万”级用户并发数 多CPU并行处理数据 单机单实例，BenchmarkSQL tpmc 达175万+
<p>高易用性</p> <ul style="list-style-type: none"> 性能诊断信息自动收集和分析 辅助调优工具，1小时即可完成原来1天的调优工作 自动内存管理调整 集群故障节点自动恢复 	<p>高安全性</p> <ul style="list-style-type: none"> 国家信息安全产品认证，安全四级销售许可证 信息技术产品安全分级评估证书(EAL4+) 商用密码产品认证证书 	<p>高生态兼容性</p> <ul style="list-style-type: none"> 具备来自上下游1300多家公司4000+份兼容认证 全面适配国家专用项目相关产品 云联合方案覆盖主流云厂商

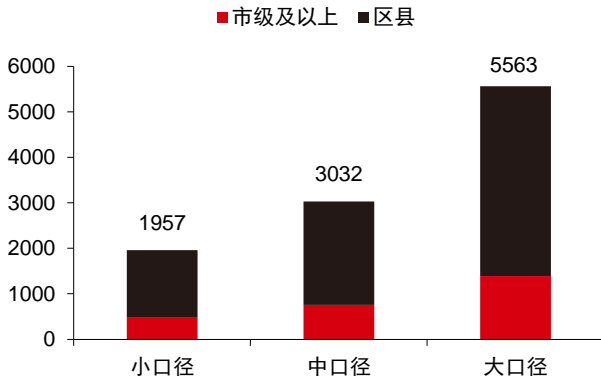
资料来源：人大金仓官网，中信证券研究部

2) 生态为纲：信创加速，国产数据库稳步推进上下游生态的适配

党政信创纵向下沉和行业信创横向拓宽持续利好国产数据库生态构建。信创需求是数据库国产替代的核心动力，外部环境不确定性提升国内信创产业发展的确定性，紧迫性、重要性持续获市场更深认知。作为对自主可控和数据安全要求最高的细分市场，党政信创开启最早，部委、省、市层面包括数据库在内的基础软硬件和 PC、应用推进顺利，信创产业后续有望从“纵向下沉”和“横向拓宽”两方面继续发展。**纵向方面**，信创核心品类预计将进一步向区县层面下沉，各条线工作落地节奏逐步清晰。我们认为当前阶段信创应用在部委、省、市层面已深入开展，未来有望进一步下沉，实现较第一轮信创三倍体量的扩展。同时在行业内部，国产化替代主要遵循“外围软件-管理支持-准核心系统-核心系统”

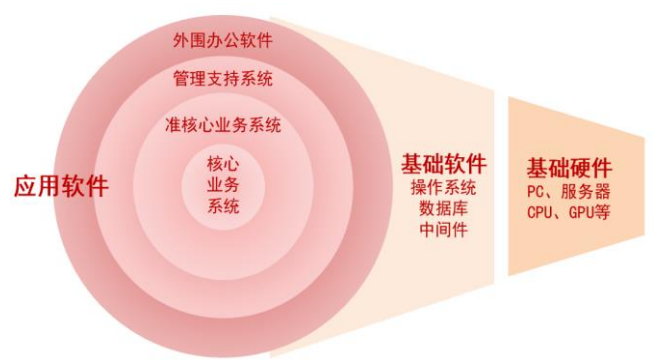
的顺序，按四个业务层级逐步深化。**横向方面**，信创核心品类有望从党政公文向电子政务、事业单位及其他行业加速渗透，按照“2+8”的自主可控体系由党政机关逐渐拓展至金融、电信、能源、教育、交通等八大行业，预计有望以 2022H2 为起点逐步进入高速发展期。信创产业的快速推进将持续利好国产数据库上下游行业生态的构建，不断加强对国产基础硬件、操作系统、中间件及各类应用的适配能力。

图 9：信创核心品类有望进一步向区县层面下沉



资料来源：国家统计局，中信证券研究部测算

图 10：行业信创推进顺序



资料来源：艾瑞咨询，中信证券研究部

上游生态：数据库软件作为基础软件，其上游主要是 CPU 芯片、服务器主机、存储设备、操作系统等基础软硬件行业。目前国内市场上除 IBM Power 小型机，以及 Intel、AMD 等主要国际 PC 服务器生态体系外，众多国产生态体系也走在快速发展的路上。其中 CPU 主要包括飞腾、龙芯、申威、鲲鹏、海光、兆芯等品牌，服务器主要包括浪潮、长城、曙光、联想等品牌，操作系统则有麒麟软件、统信软件等厂商。以达梦为例，达梦数据库与相关国内外上游计算生态企业有着良好合作关系，能够提供经过良好兼容优化的各类数据产品。此外，在上游存储设备领域，达梦也与宏杉、H3C、华为、浪潮、曙光、长城、联想、EMC 等主流厂商的存储产品具有良好的兼容适配性。

表 6：达梦数据主要上游生态合作伙伴

品类	合作伙伴
CPU	Intel、AMD、飞腾、龙芯、申威、鲲鹏、海光、兆芯等
服务器主机	IBM Power 小型机、浪潮、长城、曙光、联想等
存储设备	宏杉、H3C、华为、浪潮、曙光、长城、联想、EMC 等
操作系统	麒麟软件、统信软件等

资料来源：达梦数据招股书，中信证券研究部

下游生态：数据库软件的下游主要为应用软件开发行业，既包括传统信息化应用，如电子政务、电子商务、企业 ERP、财务管理、工业生产控制等，也包括新型的应用如大数据、人工智能、物联网等。数据库软件作为信息化系统中不可或缺的组成部分，广泛覆盖政府、金融、能源、教育、交通等大多数涉及国计民生的领域。目前我国应用软件产业整体发展较为成熟，在各行业领域拥有丰富的产品供给，形成了大量行业独立应用软件开发商（ISV）。ISV 是数据库与用户的重要桥梁，承担着数据库的应用和集成工作。与 ISV 的合作将是国产数据库公司下游生态建设的持续投入方向。

大型央企数字化转型和信创需求持续丰富国产数据库下游生态。由于海外数据库龙头较早进入国内市场打开市场，此前重点行业大型央企高度依赖海外数据库产品。随着“十四五”期间数字经济的不断加码和行业信创的稳步推进，大型央企数据库建设呈现出“升级改造”和“国产替代”的双重需求，为国产数据库厂商开拓下游用户生态创造了良好的发展机遇。目前国产数据库厂商在金融、电信、能源、交通等重点行业持续拓展大型用户，推进国产替代，通过打造标杆性行业用户不断积累行业 know-how，从而快速把握用户需求实现产品迭代。

图 11：达梦数据库客户案例

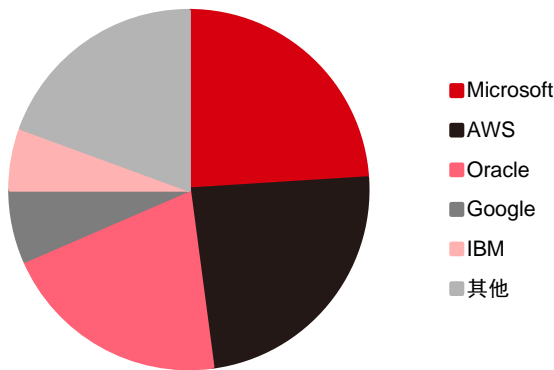


资料来源：达梦数据官网，中信证券研究部

3) 市场为证：传统数据库领域国产份额稳步提升，竞争格局逐渐清晰

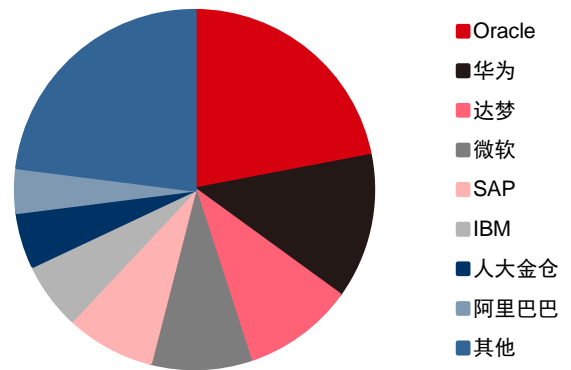
市场份额：国产厂商在国内传统数据库市场已逐渐与海外龙头分庭抗礼。Gartner 数据显示，2021 年全球数据库主要市场份额仍被微软、AWS、Oracle 等海外龙头占据。反观国内传统数据库领域，国产替代已经初具成效。IDC 数据显示，2021 年我国本地部署关系型数据库市场份额 Top3 分别为 Oracle、华为、达梦，后两者的市场份额超过了微软、IBM 等海外 IT 巨头，人大金仓、阿里巴巴等国产厂商亦在国内市场有所斩获。

图 12：2021 年全球数据库市场份额



资料来源：Gartner，中信证券研究部

图 13：2021 年我国本地部署关系型数据库市场份额



资料来源：IDC，中信证券研究部

国产企业：传统领域国产数据库竞争格局逐渐清晰。赛迪顾问数据显示，传统领域国

产数据库市场 Top5 在过去三年基本没有发生变化，市场格局逐渐趋于稳定。达梦数据库、人大金仓、优炫软件、南大通用和神州通用作为国内老牌商业数据库厂商，已成为 OLTP 领域国产替代的中坚力量。

表 7：2019-2021 年我国国产数据库管理软件市占率排名

国产数据库厂商	2021	2020	2019
达梦数据 	1	1	1
人大金仓 	2	2	5
优炫软件 	4	3	4
通用数据 	3	4	2
神州通用 	5	5	3

资料来源：赛迪顾问，达梦数据招股说明书，中信证券研究部

注：因公有云厂商（如阿里云、华为、腾讯等数据库产品）的经营模式差异较大，因此未做排名统计。

问题二：OLAP 数据库的发展到了什么阶段？

并驾齐驱：OLAP 成为继 OLTP 之后数据库的下一发展重心

20 世纪 90 年代以前，早期 OLAP 需求场景尚不成熟，OLAP 和 OLTP 在同一个数据库产品中实现，主要应用于简单的历史数据查询分析。前文中我们提到，1990s 之前，以增、删、查、改为核心的 OLTP 需求是数据库领域发展的重心。随着企业数据管理系统应用的深化，数据量的高速积累、数据应用场景的不断丰富和数据模型的不断完善，分析师和企业管理层逐渐看到数据分析的价值。

20 世纪末，分析型数据库开始崭露头角，OLAP 技术路线独立，成为继 OLTP 之后数据库领域的另一发展重心。OLAP(联机分析处理)的概念最早由关系模型之父埃德加·科德于 1993 年提出。他认为 OLTP 已不能满足终端用户对数据库查询分析的要求，用户需要对关系型数据库进行大量的计算才能辅助决策分析。OLAP 的技术路线由此独立并得到蓬勃发展，在传统数据库的基础上逐渐发展出数据仓库的产品形态，主要支持面向分析场景的应用，提供结构化的、主题化的数据用于业务反馈和辅助决策。

表 8：数据仓库与传统数据库对比

比较指标	传统数据库	数据仓库
数据内容	当前值	历史的、归档的、集成的、计算过的数据（处理过）
数据目标	面向业务操作程序、重复操作	面向主题、分析应用
数据特性	动态变化、更新	静态、不能直接更新、只能定时添加和更新
数据结构	高度结构化、复杂、适合操作计算	简单、适合分析
使用频率	高	低
数据标准化	高度标准化的静态 Schema	非标准化 Schema，例如星型 Schema 或雪花型 Schema

比较指标	传统数据库	数据仓库
数据存储	往往采用“行存储”，对在单行型物理块中执行高吞吐量写入操作进行了优化	使用“列存储”进行了优化，可实现轻松访问和高速查询列性能
数据访问	大量小型读取操作	为最小化 I/O 并最大化数据吞吐量进行了优化
数据访问量	每个事物一般之访问少量记录	每个事务一般访问大量的记录
响应时间要求	计时单位小，如秒甚至毫秒	计时单位相对较大、如分钟、小时等

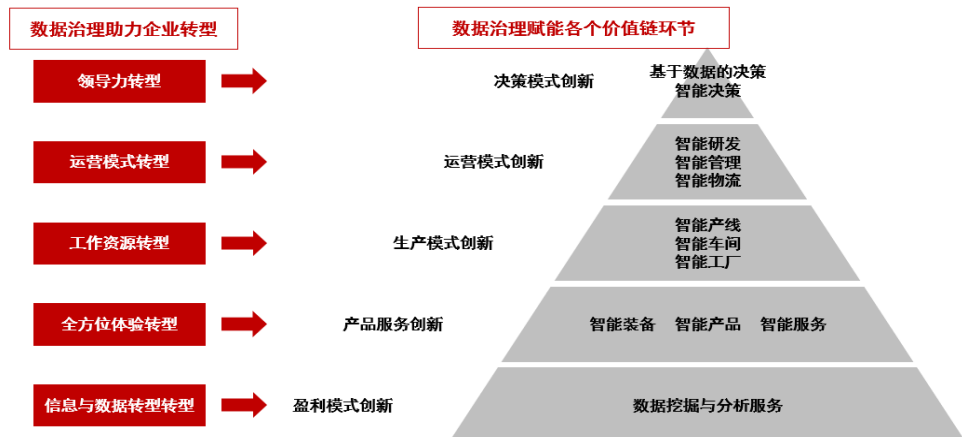
资料来源：华为云官网，中信证券研究部

我们认为，OLAP 需求的独立和分析型数据库的爆发是数据库行业发展的必然趋势，其驱动因素主要包括需求侧和供给侧两个维度：

1) 需求侧：数据量的积累带来数据赋能的潜力，分析处理的应用场景不断丰富

数据治理能够实现对企业各个价值链环节的赋能，提升企业的运营与决策效率。数据量的积累使得基于历史数据的分析决策成为可能，企业的顶层决策、生产运营、后台研发等一系列环节将逐步由数字化迈向智能化。我们认为，信息密集型、劳动密集型行业的数据治理赋能成果更易显现，在业务运营过程中容易产生体量巨大、数据结构不统一的数据。基于对历史数据的分析可以充分赋能产品研发、营销销售、售后服务等诸多环节。以金融行业为例，基于个人消费行为数据、征信数据、储户信用报告数据、交易数据的分析可以帮助企业更高效地开展风险评估以及理财产品的定制化推介营销活动。

图 14：数据治理赋能企业各个价值链环节



资料来源：帆软官网，中信证券研究部

2) 供给侧：海内外传统数据库巨头、云厂商、独立厂商百家争鸣，各有千秋

分析型数据库领域由于场景需求碎片化、技术路径多样化，海内外各类厂商呈现出百花齐放的态势。传统 IT 巨头多在关系模型领域深耕，凭借在 OLTP 领域的先发优势率先进行探索，整体占据主导地位。但云计算、大数据的快速发展带来了需求的进一步爆发，各类数据模型、各种应用场景的需求逐渐分化。同时随着开源生态的不断丰富，以 Apache 软件基金会为代表的开源体系也为巨头之外的数据库厂商的发展提供了一片沃土。云计算巨头、独立数据库公司的数据库产品快速崛起。

图 15：国内外分析型数据库市场代表厂商



资料来源：各公司官网，中信证券研究部

传统巨头：在 OLTP 领域起步相对较早具有先发优势，产品具备高稳定性、高安全性的优势，且具备良好的客户基础，市场份额较高。但技术架构相对传统，需要承担较高的运维成本和改造成本，在新技术的适配性上存在短板。典型代表包括海外 Oracle、IBM、Microsoft、SAP 和国内人大金仓、达梦数据库等。

云厂商：对于应用场景（特别是互联网领域）的理解更加深刻，产品矩阵类型丰富。但相对缺少中立性，销售绑定云服务，降低企业可选择性，同时对于私有化部署相对缺乏服务能力。典型代表包括海外亚马逊、谷歌和国内阿里、腾讯、华为等。

独立厂商：技术架构先进，能够满足更加多元化的分析需求，各自在自身的强势领域深耕细作。但商业化验证维度存在欠缺，客户消费意愿、消费能力以及市场空间均有待验证，财务表现相对较弱。典型代表包括海外 Databricks、Snowflake、MongoDB 和国内环数科技、PingCAP、偶数科技等公司。

图 16：2020、2021 年全球数据库魔力象限



资料来源：Gartner

技术复盘：把握数据处理效率、数据完整性两条发展主线

路径回顾：OLAP 先后衍生出数据仓库、数据湖的发展路径，现在正在进行湖仓一体、智能湖仓的实践尝试。数据仓库聚焦于结构化数据处理能力的问题，由传统 OLTP 数据库提供底层数据，主流采用 MPP（大规模并行计算）的无共享架构，相较于早期分析型数据库显著提升了扩展性和对于结构化数据的处理性能，但不支持非结构化、半结构化数据的存储和分析；数据湖聚焦于数据完整性的实现，支持对于各类半结构化数据（CSV、XML、日志等）、非结构化数据（文档、图片、音频、视频等）的存储和分析，大大拓展了数据分析的使用场景和功能，但在结构化数据处理、ACID 特性支持、数据的实时性与可靠性等维度存在短板。为了兼顾数据分析效率和数据完整性，同时在分析过程中与 AI/ML 更紧密结合，近年来众多分析型数据库厂商开始进行湖仓一体、智能湖仓的尝试。

表 9：不同技术架构分析型数据库关键指标

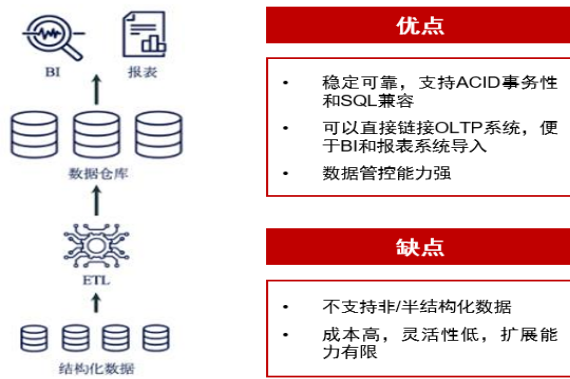
	数据仓库	数据湖	湖仓一体
数据类型	结构化数据	结构化、半结构化/非结构化数据	结构化、半结构化/非结构化数据、实时数据
数据量级	GB 级、TB 级	TB 级、PB 级	TB 级、PB 级
Schema	在写入数仓前设计	在取用时写入	统一元数据管理
ACID 事务支持	良好	较弱	良好
性价比	较高的成本获得、很快的查询结果	较低的成本获得、较快的查询结果	较低的成本获得、最快的查询结果
指标要求	性能、SQL 标准、ACID 特性、可扩展性	可扩展性、在线扩容、弹性扩容	性能、SQL 标准、ACID 特性、可扩展性、在线扩容、弹性扩容、敏捷开发
用户	业务分析师	业务分析师、数据开发人员、数据科学家	业务分析师、数据开发人员、数据科学家
适用场景	大规模结构化数据查询和分析	大规模数据的离线批量计算、高级预测性分析	超大规模结构化数据查询和分析、实时数据处理、AI/ML 应用
业务价值	对单个业务环节实现描述性分析、诊断性分析	实现诊断性、预测性分析，对现有及未来客户价值具有前瞻性	实时产品与服务、决策与行动最优化

资料来源：IDC，艾瑞咨询，AWS 官网，CSDN@ Data+Science+Insight，中信证券研究部

1) 数据仓库：基于 MPP 架构实现较大规模结构化数据计算效率优化，但在可用性、可扩展性和数据模型灵活性上仍存在短板

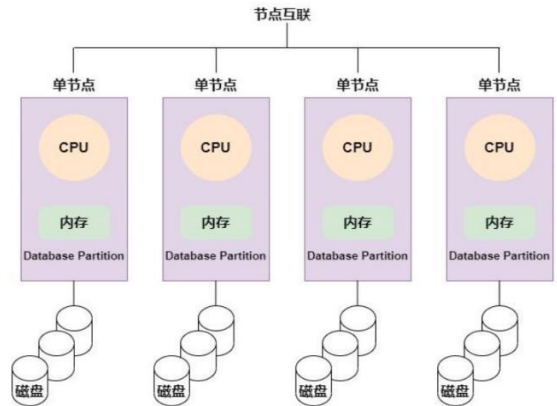
技术架构：数据仓库的分析对象主要来自 OLTP 数据库的结构化数据，通过预先定义 Schema 的方式，运用 ETL（抽去、转换、加载）操作将数据导入数据仓库后，用户可以较为便捷地链接 BI 系统和报表系统。由于与 OLTP 数据库高度结合，数据仓库对于元数据的要求十分严格，很多数据仓库同样满足 ACID 事务能力。早期数据仓库主流采用 MPP（大规模并行处理）架构，通过一定的节点互联网连接多台 SMP 服务器，每个节点之间采用完全无共享（Shared Nothing）结构，具有独立的 CPU、内存和磁盘资源。在实务过程中，来自 OLTP 数据库的数据将根据来源场景、应用特点分配到不同的节点上，在每个处理单元上并行地进行计算分析，最终每个节点计算完成后再统一汇总得到最终结果。

图 17：数据仓库架构及优缺点



资料来源：阿里云官网，Databricks，中信证券研究部

图 18：MPP（大规模并行处理）架构



资料来源：墨天轮

性能分析：基于 MPP 架构的数据仓库在 ACID 事务性支持和中等规模数据分析效率上具备优势。由于数据源来自高度结构化的 OLTP 系统，数据仓库具有稳定可靠、支持 ACID 事务性和 SQL 兼容的优势，同时多个节点的并行计算也提高了数据仓库所能处理的数据量水平。但是，基于 MPP 架构的数据仓库在数据模型的灵活性、可用性和扩展性的维度上仍存在短板。Web2.0 时代的来临使得企业在日常运营过程中积累了大量非结构化、半结构化数据（如日志、图片、文档、音视频等），需要提前设计 Schema 的数据仓库无法应对非/半结构化数据的处理需求。此外，由于 MPP 的各个节点并行处理任务，一旦某个节点出现性能短板或性能故障，将会降低整个系统的处理性能。因此 MPP 架构的可用性（部分节点发生故障时继续运行的能力）、并发度（单位时间内所能够处理的任務数）仍然存在缺陷，这也进一步造成了 MPP 数据仓库可扩展性以及扩展成本上的短板，使得数据仓库无法应对大数据时代 PB 级甚至更高的数据处理需求。

2) 数据湖：以 Hadoop 架构为代表的数据库提高了可扩展性和数据模型的灵活性，但牺牲了一定程度数据的实时性和可靠性

技术架构：数据湖的核心是存储业务数据的完整副本（原始数据），包括结构化数据、非结构化数据以及半结构化数据。Hadoop 是企业数据湖建设的典型架构，以分布式文件系统 HDFS、分布式计算引擎 MapReduce 为核心组件，将所有机器的存储资源与计算资源进行分层抽象设计。2003 年前后，Google 连续发表三篇论文，奠定了大数据的框架基础。此后基于理论又形成了 Hadoop 原始的“3+1”式软件栈：即分布式文件系统 HDFS、分布式计算引擎 MapReduce、Hbase NoSQL 数据库，以及 YARN 资源调度。Hadoop 定义了最基础的分布式大数据批处理架构，打破了传统数据库一体化的模式，将计算与存储分离，聚焦于解决海量数据的低成本存储与规模化处理。Hadoop 在面对上百 PB 数量级的大数据查询分析时能够极大地提升效率，同时通过使用廉价硬件集群搭建的分布式系统实现成本效益。

图 19：数据湖架构及优缺点



资料来源：阿里云官网，Databricks，中信证券研究部

性能分析：基于 Hadoop 架构的数据湖解决了半/非结构化数据的存储问题，同时通过存算分离的架构设计提高了可扩展性。数据湖中各种类型的数据均按原样存储，在分析时采用 Schema-on-read 模式，能够满足互联网场景下多种数据类型存储和分析的需求。但也以牺牲 ACID 事务性作为代价。如果要基于 Hadoop 实现 BI、报表等功能，需要将数据库的数据经过 ETL 进入数据仓库、在版本控制、数据索引等维度存在短板。

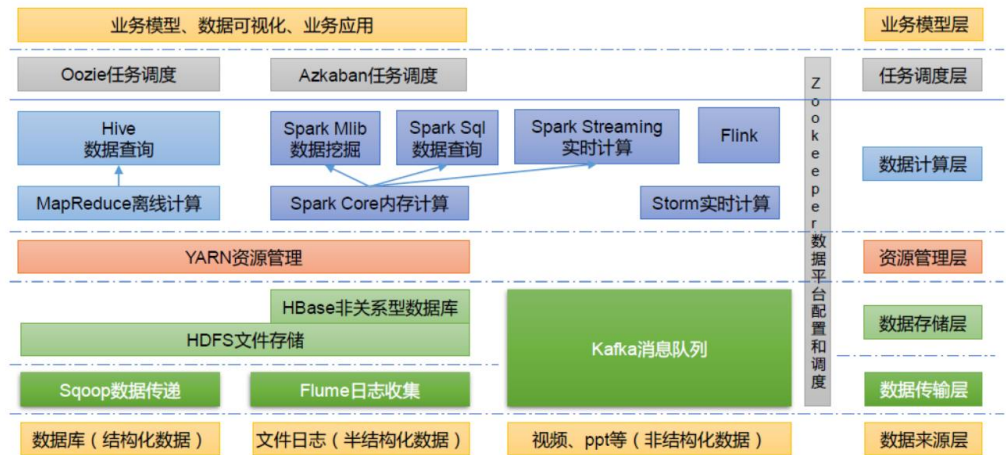
表 10：Hadoop 与 MPP 差异

特性	Hadoop	MPP
平台开放性	完全开源	专有
节点数	平均数百个节点，最大数千个	平均数十个节点，最大数百个
数据量	支持大于 10PB	一般不大于 10PB
查询延迟	10-20 秒	10-20 毫秒
技术价格	每个节点免费或数千美元	每个节点数十到十万美元
计算性能	对非关系型操作效率高	对关系型操作效率高
数据结构	结构化、半结构化和非机构化数据	结构化数据
可扩展性	完整的编程接口，扩展性强	有限扩展能力，主要通过 UDF 支持
事务支持	有限	完整

资料来源：CSDN@Shockang，中信证券研究部

生态演化：核心组件基础之上衍生出庞杂的开源 Hadoop 生态圈。仅有 HDFS、MapReduce 组件并不能支撑企业级的大数据分析应用，在此基础上衍生出丰富的生态组件，包括资源管理系统、各类计算引擎、ETL 工具、安装部署工具、数据库/数据仓库产品等。同时，还出现了 Hadoop 发行版商业公司，通过提供整合、加强后的打包产品和服务，解决繁杂组件带来的版本管理混乱、部署过程繁琐、升级过程复杂等问题。

图 20: Hadoop 技术生态体系

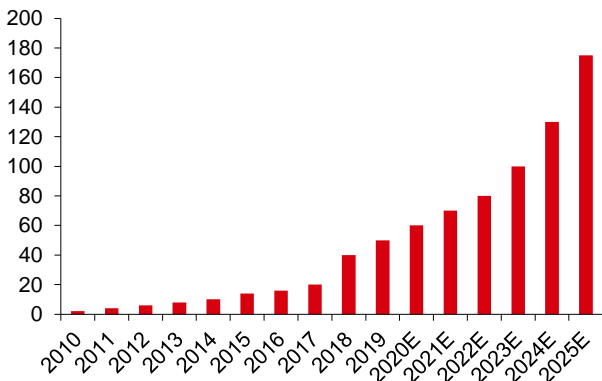


资料来源: CSDN@从一到无穷大

3) 湖仓一体和智能湖仓：“仓”“湖”结合，兼顾事务性、扩展性和灵活性，并逐渐向数据全生命周期管理发展

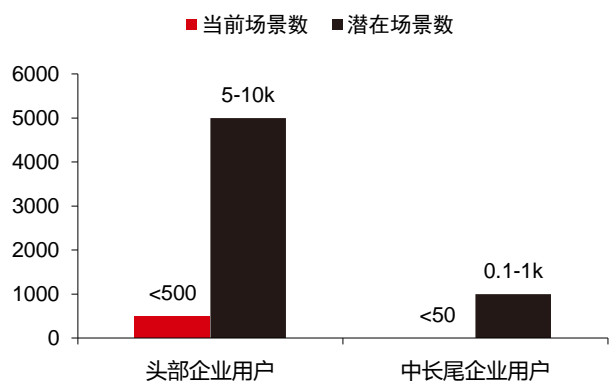
数据量的爆发增长和应用场景的不断丰富为企业分析型数据库提出了更高的要求。随着云、5G 基础设施的成熟带动互联网的深入发展，各种应用程序、移动设备、边缘设备、传感器所产生的数据总量正在以前所未有的速率爆发式增长。IDC 预计 2025 年全球数据总量将达到 175ZB，其中超过 25%为实时数据。数据仓库在扩展性和数据模型的局限性亟待突破。此外，企业数据分析的应用场景不断丰富，智能化分析水平不断提高，爱分析调研显示未来企业存在广阔的 AI/ML 应用空间，企业数据分析智能化需求有望爆发。

图 21: 全球数据量预测 (ZB)



资料来源: IDC (含预测), 中信证券研究部

图 22: 我国企业潜在 AI/ML 应用场景数目调研



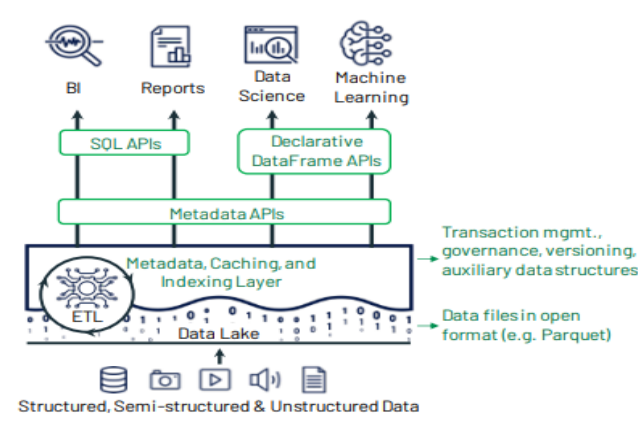
资料来源: 爱分析, 中信证券研究部

基于此，兼具数据仓库和数据湖优势的湖仓一体应运而生。2021 年创新数据系统研究会议 (CIDR) 上 Databricks, UC Berkeley 和 Stanford University 联合发布的论文《Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics》，系统描绘了新一代湖仓一体架构。数据湖仓的核心是将“湖内”和“仓内”的数据和元数据进行打通，实现自由流动。各类结构化、非/半结构化数据使用标

准文件格式（如 Parquet），通过对象存储的方式依然存储在底层的数据湖当中。在数据湖之上建立的元数据层实现 ACID 事务性、版本控制等数据管理功能。元数据层作为存储层和计算层之间的中间层，通过缓存、索引、辅助数据和数据布局优化等多种优化手段减少计算和存储层之间的 I/O 流量，优化 OLAP 工作负载的性能。元数据层之上的各类计算引擎（包括面向 BI/报表的 SQL 类工作负载和面向数据挖掘的机器学习工作负载）共享统一的数据存储，可以按需摄取热数据、回注冷数据。

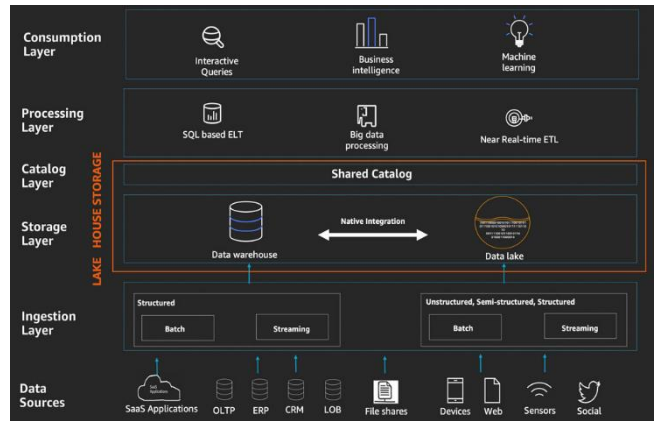
未来“智能湖仓”架构将把湖、仓以及所有其他数据处理服务组成统一且连续的整体。AWS 提出的智能湖仓架构旨在以数据为中心构建“数据服务环”。数据湖作为数据中央存储库，围绕数据湖建立包括数据仓库、机器学习、大数据处理、日志分析等一系列专用服务，各项服务共享同一的数据存储，按需对湖内数据进行摄取和回注，同时彼此之间可以以低成本、高效率地进行数据交换，最终实现企业数据全生命周期管理。

图 23：湖仓一体架构图



资料来源：Databricks

图 24：亚马逊智能湖仓架构图



资料来源：AWS 官网

需求推演：决策实时化、场景精细化、产品标准化

数据库的发展历史是用户数据治理需求的变迁史，需求的演变方向决定技术路线的演进方向。纵观数据库近 70 年的发展历程，从 1960s 增删查改的事务性需求的出现带来了 OLTP 数据库的兴起，到 1990s 针对历史数据的分析和辅助决策需求推动了 OLAP 数据库的发展，用户需求的变迁决定了数据库技术的发展重心。分析型数据库的发展脉络，数据处理效率的更高要求催生了数据仓库，半/非结构化数据的治理需求催生了数据湖，而用户对于数据分析事务性、扩展性和灵活性的统一追求催生了湖仓一体和智能湖仓。

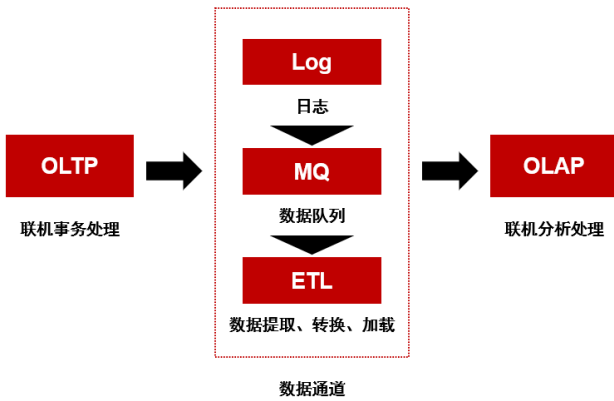
我们认为，当下分析型数据库正呈现出决策实时化、场景精细化、产品标准化的需求，这亦是未来 OLAP 数据库的演进方向。

1) 决策实时化：打通 TP/AP 消除 ETL 延时，HTAP 助力实现实时决策

OLAP 与 OLTP 之间的数据传输延时导致在处理实时性极高的分析业务时存在短板。不论是数据仓库还是数据湖，在进行分析处理时都需要基于事务处理所产生和积累的数据，必须经过数据提取、转换、加载的 ETL 过程，在此过程中为了保证系统的高可用将会产生

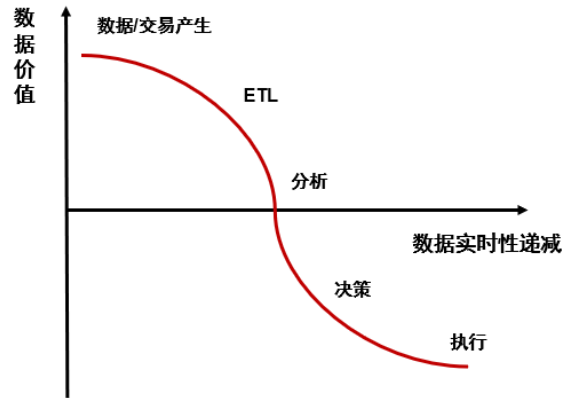
大量且分散的副本数据造成数据冗余，最终导致较高的同步难度和运维成本。同时，当用户面临实时性要求极高的分析业务场景时，OLAP 与 OLTP 之间分钟级甚至小时级的数据传输延时将难以满足分析需求，数据实时性所蕴含的数据价值也会随着 ETL 的延时而逐渐消弭。此外，当用户需要调用不同系统之间的数据进行聚合分析时，实时性方面的短板将被进一步放大。

图 25：TP/AP 分离数据库应用架构数据流



资料来源：PingCAP，中信证券研究部

图 26：数据价值随实时性递减



资料来源：DTCC，腾讯云，中信证券研究部

HTAP 混合事务和分析处理消除了 OLAP 和 OLTP 之间的间隔，可以更好满足实时分析和决策需求。目前市场上的 HTAP 实现路径主要由三种：第一种在上层应用层实现混合处理，通过 OLAP/OLTP 的松耦合和底层共享存储缩短数据同步时间，只能在数据库和应用的整体层面呈现 HTAP 能力；第二种分别运用行存储引擎和列存储引擎进行 OLTP 和 OLAP，存储引擎在物理上进行隔离，通过分布式协议进行实时复制和同步；第三种采用单一存储引擎，在最底层实现 HTAP，但目前仍处于技术探索阶段。**第二种分离存储架构、同一系统的 HTAP 是目前的主流解决方案。**

表 11：非 HTAP 架构与两种 HTAP 方案对比

	不同系统的 HTAP	分离存储架构 HTAP	统一存储架构 HTAP
存储引擎	/	行存+列存分离存储引擎	单一存储引擎
架构			
定义	OLTP 和 OLAP 松耦合，通过 ETL 将 OLTP 结果同步到 OLAP，底层共享存储缩短数据同步时间，在应用层实现混合处理，整体呈现 HTAP 能力。	基于分布式架构，行存储引擎处理事务 OLTP、列存储引擎进行分析 OLAP，遵循共识协议在引擎间复制数据，在数据库层实现 HTAP。	使用单一存储引擎同时支持 OLTP 事务处理和 OLAP 分析，在最底层实现 HTAP。
产品	SAP	TiDB、PolarDB、Oceanbase、GaussDB、TDSQL、F1	Hive、Impala、Kudu、Hyper、MemSQL
优势	初步融合 TP、AP	支持弹性扩容，按需扩展 资源隔离技术成熟，高性能	彻底融合 TP 和 AP，延迟低 数据可见度高
劣势	运维成本高 同步延时、事务分析延时	存在数据同步延迟 数据可见度差	技术不成熟 行列隔离性差

资料来源：DTCC，头豹研究院，艾瑞咨询，中信证券研究部

2) 场景精细化：深耕细分领域积累行业 know-how，应用场景愈加精细化

数据分析与数据管理的应用场景在未来将持续拓展和深化。数据分析的应用将继续向各行业领域的核心业务渗透，数据的采集、流通、分析、应用的价值闭环将持续完善。由数据分析需求逐渐衍生出的大数据管理将逐渐改变各个行业的各个价值链环节。一方面，基于历史数据分析的销售预测、趋势分析、营销策略设计、客群画像匹配的优化建议将提高用户的运营效率和决策效率；另一方面，基于数据分析的如 AI、大数据的应用有望带来新商业模式、新产品形态、新应用场景的开拓，如无人驾驶、智能安防、智慧物流等。

图 27：数据分析需求在部分行业深化方向

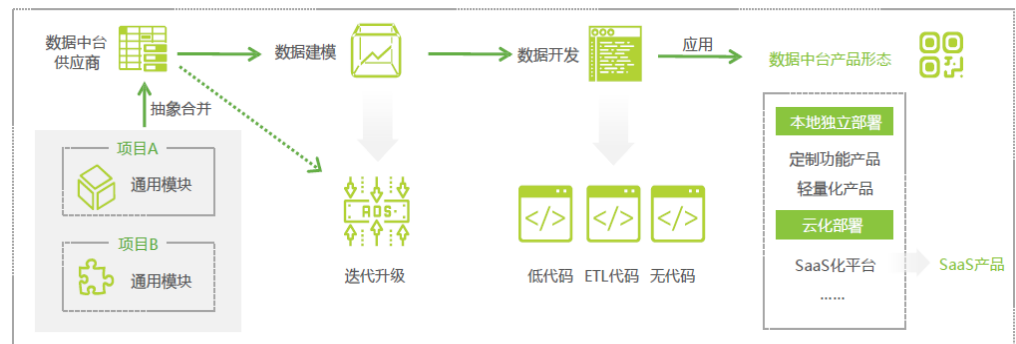


资料来源：前瞻产业研究院，中信证券研究部

3) 产品标准化：技术 SaaS 化、解决方案标准化打开长尾下沉市场

分析型数据库产品将逐渐实现标准化，进入下沉市场提高中小企业渗透率。受限于数据治理需求碎片化、场景理解不够深入等因素，现阶段分析型数据库产品主要集中在大型企业客户市场，且定制化程度相对较高。未来伴随更多业务场景能力的沉淀，分析型数据库厂商将不断丰富产品矩阵，完善数据治理服务的深度和广度，通过产品标准化的途径降低成本，从而提高在长尾下沉市场的渗透率。

图 28：数据中台产品技术 SaaS 化



资料来源：艾瑞咨询

■ 问题三：如何看待国产数据库的创新方向？

看架构：分布式成重要趋势，云数据库打开更大市场

按照架构模式进行分类，数据库可以分为分布式数据库和集中式数据库。这种分类方式的诞生，一方面是由于传统集中式数据库缺乏扩展性，为了实现扩展而出现了分布式数据库，另一方面，是缘于云技术和网络技术快速发展，推动分布式技术升级，形成新型分布式数据库。**集中式数据库**由一个处理器、与它相关联的数据存储设备以及其他外围设备组成，将数据集中在一台机器上进行处理，被物理地定义到单个位置。典型代表有 Oracle、DB2、人大金仓、武汉达梦等；**分布式数据库**采用分布式架构，将数据在网络上分开存储于多个机器中进行处理。分布式数据库是一个数据集合，这些数据在逻辑上属于同一个系统，但物理上却分散在计算机网络的若干站点上，并且要求网络的每个站点具有自治的处理能力，能执行本地的应用。分布式数据库典型代表如谷歌的 Google Spanner、阿里巴巴的 OceanBase、华为的 GaussDB 等。

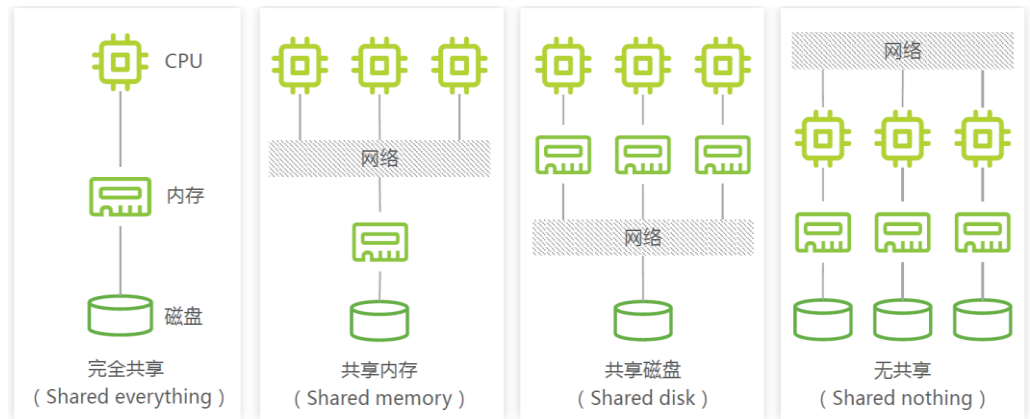
表 12：传统集中式数据库与新兴分布式数据库对比

	优势	劣势
传统集中式数据库	<p>1、较高的安全性和稳定性：数据存储集中在中央计算机上，只要对中央计算机做好保护措施以及数据备份工作，便能保证较高的安全性和稳定性，适合存储重要的核心数据；</p> <p>2、运算开销和通信开销小：服务器结构部署简单，存取数据的运算开销小，通信的开销也较小；</p> <p>3、产品商品化程度高：现有的集中式数据库产品大都经过了时间的考验，稳定可靠，商品化程度高，数据库公司服务能力成熟，对于客户来说便捷安全；</p>	<p>1、存储空间有限：集中式数据库把数据都存储在主机中，存储空间有限，只依靠主机计算和存储数据将难以解决这一问题。</p> <p>2、响应速度较慢：中央计算机需要执行所有的运算，当终端很多时，会导致响应速度变慢。</p> <p>3、灵活性差：如果终端用户有不同的需要，要对每个用户的程序和资源做单独的配置，在集中式系统上做起来比较困难，而且效率不高。</p>
新型分布式数据库	<p>1、满足数据量、计算量大的需求：只要布置足够多的计算机，便能储存足够多的数据，相比传统集中式数据库可以储存更多的数据，做更大量的数据计算；</p> <p>2、读取数据和运算的速度更快：可并发读取数据，可并行运算，因此读取数据和运算速度都会比集中式数据库更快；</p> <p>3、灵活、可扩展性好：可以方便地添加新机器对系统进行扩充，而集中式数据库就没有这种便捷的可扩展性；</p>	<p>1、稳定性、安全性、保密性不如传统集中式数据库：分布式系统通过通信网络来控制局部，而通信网络的保密性和安全性存在缺陷，例如容易受到黑客攻击；</p> <p>2、分布式事务的代价较高：两阶段的提交造成过多消息传输，可能的锁争用变大，保证数据同步的代价较高，分布式事务在未来将还是个很难消除的瓶颈</p> <p>3、管理复杂：分布式数据库有更多的节点和组件，更难于管理，故障排除和管理方面要求更高。</p>

资料来源：CSDN@Q.E.D.，中信证券研究部

硬件架构：数据库硬件架构主要有完全共享、共享内存、共享磁盘和无共享四种。完全共享（Shared Everything）模式拥有完全透明共享的 CPU、内存和磁盘，属于集中式数据库的范畴，天然具有较好的 AICD 事务性，但扩展性和并行性较差；**共享磁盘（Shared Disk）和共享内存（Shared Memory）**模式允许增加内存节点和磁盘节点以提高并行处理能力，但是随着数据体量的爆发式增长，共享磁盘的接口数量容易达到上限，共享内存的内存访问和网络带宽之间冲突增强，系统处理速度将会遭遇瓶颈。**无共享（Shared Nothing）**模式下每个节点具备独立的 CPU、内存、磁盘，每个处理单元独立运行，各单元之间通过协议通信。无共享架构具备良好的扩展能力和并行处理能力，从 MPP 数据仓库时代起逐渐得到广泛应用。随着硬件成本的下降，无共享模式已逐渐成为分布式硬件架构的主流。

图 29：硬件架构有中 4 种不同层次的数据共享



资料来源：艾瑞咨询

主流应用：通过无共享架构实现的分布式架构已成为大数据管理的主流解决方案。数据量的爆发式增长以及应用负载的快速增加使得传统单一服务器架构的集中式数据库出现瓶颈，包括传统集中式数据库厂商、新兴厂商在内的各类玩家均开始探索数据功能的分布式实现。三种分布式架构中，无共享架构凭借高可用性、高扩展性、低带宽要求等优势已成为分布式架构的主流解决方案。

表 13：三种分布式架构的优劣势分析

	共享磁盘架构	共享内存架构	无共享架构
定义	每个处理器自带私有内存区域 通过内部通讯机制访问所有磁盘	多个处理器共享一片内存区域 通过内部通讯机制访问所有内存	每个处理器都私有内存和磁盘空间 处理器之间通讯通过网络连接
优势	快速适应变化的负载 高可用 数据不需要分区	故障恢复能力高 单个系统镜像，管理简单 主存转换效率高，消息数量积累少	无限的水平扩展 无单点，可用性强 业务侵入性小
劣势	写入瓶颈并没有解决，单机写入依靠 文件系统的多副本复制及高可用	内存访问干涉与网络带宽竞争问题 扩展能力有限	实现较为复杂 在某些场景下需要重新思考业务设计
并行控制能力	劣	中	中
灾难恢复能力	劣	优	中
数据库设计易度	中	中	劣
负载均衡能力	中	优	劣
高可用性	中	劣	优
消息数量处理	中	优	劣
带宽要求	中	劣	优
横向扩展能力	中	劣	优
远程部署能力	中	劣	优
关键回话敏感性	中	劣	优
系统镜像数量	劣	优	劣
热点敏感度	劣	劣	劣

资料来源：Oracle, ScaleDB, 头豹研究院, 韩国工业科学院 RIST, 中信证券研究部

技术实现：分布式架构的实现方式将逐渐从借助中间件向原生分布式过渡。分布式架构的实现路径包括借助中间和原生分布式两类，其中原生分布式包括共享存储分布式数据库、去中心化的分布式数据库，不同技术路线产品各有千秋。分库分表+中间件的模式相对成熟，但整体依然基于单机数据库的存算性能，依托中间件进行数据分配和任务管理，在并发性和扩展性上仍有局限。原生分布式实现了存储层、计算层的全面分布式改造，但目前技术成熟度相对较低。

表 14：分布式数据库技术路线对比

	分库分表+中间件	共享存储分布式数据库	去中心化的分布式数据库
方案	下层的单机数据库提供存储和执行能力，在多个单机数据库上封装一层中间层补充分布式能力，以统一的数据分片规则管理分布在不同数据库节点的数据，并提供 SQL 解析，请求转发和结果合并的能力。	计算节点独立并且共享一个不带计算功能的存储集群(Shared storage)，数据存储的底层是可动态扩容的分布式高性能存储，以存算分离架构，计算层和存储层都可以动态扩容，并且这些分布式数据库都会对网络以及存储层的优化来保证高可用和高性能。	每个节点有独立的计算和存储功能并且节点之间不共享数据 (Shared-nothing)，为了平滑的扩缩容也采用了存算分离的架构，分布式集群的每个节点都是独立节点，通过 multi-paxos 或者 multi-raft 等共识算法来保证多副本的可用性。
优势	可以利用现有开源数据库成熟稳定的产品功能，具备高性能、低成本、稳定性、用户门槛低，（能力上限低但下限高）	事务性能优、读写响应最快、最大程度提升写入容量限制	架构解耦性高、高兼容性、高可移植部署性、强一致高可用
劣势	Sharding（分片）拆分成本高、底层架构不具备分布式能力，中间件通讯及单体数据库功能受限存在扩展性瓶颈	架构可改造性低、依赖共享存储系统，移植性低	具备较高的硬件要求、分布式事务锁机制，多写性能低
案例	GoldenDB、TDSQL MySQL 版、GreatDB、HotDB、MogDB、GaiaDB-X、openGauss	AWS Aurora、PolarDB、TDSQL-C、SequoiaDB-MySQL、GaussDB for MySQL、ArkDB	TiDB、Oceanbase、Google Spanner、Cockroach、Hubble

资料来源：头豹研究院，中信证券研究部

技术内核：从存算一体到存算解耦，硬件成本的降低和网络带宽的提高保障分布式架构的实现。 20 世纪 80 年代，Oracle 推出了首款数据库产品。彼时服务器硬件成本高昂，硬件算力、存储、网络带宽都十分有限。因此数据库产品在优化过程中难以依托服务器之间的信息交换，而是聚焦于在单服务器的 CPU、内存、磁盘固定配置下进行极致优化。因此在软件架构的设计中，存储与计算高度耦合，其核心思想是通过存算一体实现性能的极致优化。随着硬件成本的大幅降低和网络带宽的大幅提高，通过集群服务器的硬件设计，联合多个节点进行协议通信以实现分布式计算成为可能。软件算法的设计无需再基于存储和计算的深度绑定，存算解耦的思想为分布式的实现提供了更多想象力。

表 15：存算一体 vs 存算解耦

	存算一体	存算解耦
优势	针对固定硬件配置做极致优化，单机性能较好	使用通用型服务器，可配置范围广，硬件的通用性和灵活性高； 集群水平扩展、垂直扩展灵活，可以按需单独扩展计算或存储； 集群扩展后自动管理数据均匀分布，无需数据重分布操作，连续提供对外服务；
劣势	使用专有服务器或硬件配置较为固定，硬件的通用性和灵活性不足 集群水平扩展、垂直扩展不灵活，单独扩展计算或存储较困难； 集群扩展后需要做数据重分布，期间影响甚至停止对外服务；	为保持硬件的通用性和灵活性，单机性能优化较一般。

资料来源：CSDN@星环科技，中信证券研究部

分布式数据库的“资源池化”思想与云计算的“按需服务”理念具有异曲同工之处，天然满足云原生的需求。 分布式数据库迁移到云计算平台后可以轻松实现数据与业务的分离、存储与计算的分离。云数据库可以相对不受限制地实现基础设施资源的调动，以满足上层对于高扩展性、高并发、高吞吐量、灵活配置的需求。因此，云数据库在成本、可用性、易用性、扩展性和并行处理方面较传统数据库有绝对优势。但同时，由于现阶段云数

数据库产品仍处于相对不成熟阶段，且市场的普遍存在公有云和私有云的混合部署需求，云数据库在数据迁移、数据质量、性能优化和规范标准方面仍有局限。

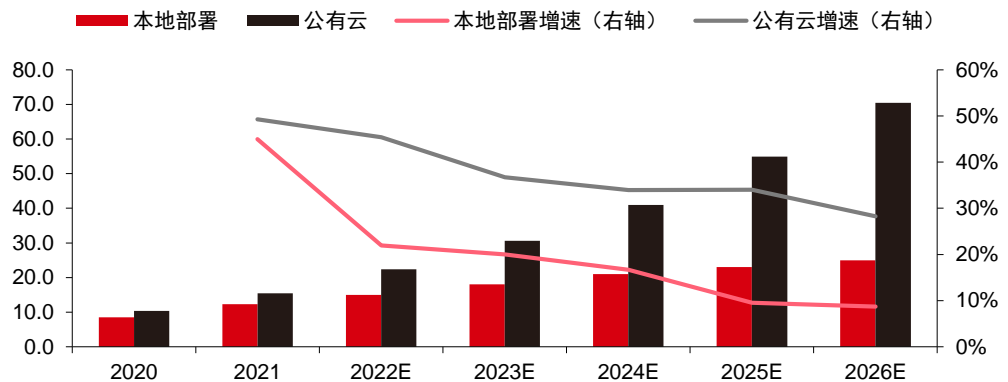
表 16：云数据库的机遇与挑战

机遇	挑战
低成本：多租户模式，用户之间共享资源且只用按需付费，节省了成本	数据质量：云数据库在大数据环境下，容易产生脏数据，影响事务一致性
高可用：高水平的容错能力，一个节点崩溃，其他节点也可以继续工作	数据迁移：将大量、复杂的企业内部数据库数据迁移上云存在一定困难
易用性：不需要关心底层服务器、系统等部署和运维，开箱即用	数据融合：本地数据与云数据长期并存，需要有效的融合机制，统一管理
动态可扩展：具有无限可扩展性，可以满足不断增加的数据存储需求	性能优化：云环境为动态负载均衡、资源分配管理提出了新的要求
大规模并行处理：并行处理能力强，面对海量数据，几乎可以做到实时的响应	规范标准：各大厂商独立发展云数据库，在查询语言、语言模型和安全等方面缺乏统一的规范标准

资料来源：艾瑞咨询，中信证券研究部

在未来，上云需求将持续为数据库市场带来增量。IDC 数据显示，2021 年我国关系型数据库中，公有云部署的市场规模增速已经超过本地部署的增速，预计从 2022 年开始二者的增速差将进一步拉大。IDC 预测，未来三年关系型数据库中云数据库的市场规模增速有望保持在 40%左右，而本地部署模式的规模增速仅为 20%，云数据库的市场份额有望进一步提高。

图 30：2020-2026 中国关系型数据库市场规模（按部署方式分类）



资料来源：IDC（含预测），中信证券研究部

看模型：关系型与非关系型长期共存，重视多模型能力构建

数据模型先后经过了层次模型、网状模型和关系模型的变迁，互联网的兴起推动非关系模型和 NoSQL 数据库登上历史舞台。20 世纪 80 年代以来，结构化的关系模型始终占据市场主流，随着 Web2.0 的繁荣非结构化和半结构化数据（如日志、图片、文档、音视频等）出现爆发式增长，面向非关系型数据的 NoSQL 数据库开始走向市场，区别于关系数据库，它们往往不保证关系数据的 ACID 特性，对于超大规模和高并发数据具有较好的处理能力。NoSQL 数据库种类繁多，数据之间无关系，容易扩展。NoSQL 数据库具有非常高的读写性能，尤其在大数据量下，主要在于它的无关系性，数据库的结构简单。目前对于非关系型数据库主要有四种数据存储类型：键值对存储（key-value），文档存储

(document store), 基于列的数据库(column-oriented), 图形数据库(graph database)。

表 17: 四种非关系型数据库对比

类型	产品代表	应用场景	数据模型	优点	缺点
键值对存储	Redis, Voldemort, Oracle BDB	内容缓存, 主要用于处理大量数据的高访问负载, 用于一些日志系统、统计等。	Key 指向 Value 的键值对, 通常用 hash table 来实现	查找速度快	数据无结构化, 通常只被当作字符串或者二进制数据
文档型数据库	CouchDB, MongoDB	网站数据; 缓存; 大尺寸, 低价值的; 高伸缩性的数据; 用于对象及 JSON 数据的存储。	Key-Value 对应的键值对, Value 为结构化数据	数据结构要求不严格, 表结构可变, 不需要像关系型数据库一样需要预先定义表结构	功能相对局限
列存储数据库	Cassandra, HBase, Riak	分布式的文件系统	以列簇式存储, 将同一列数据存在一起	查找速度快, 可扩展性强, 更容易进行分布式扩展	查询性能不高, 而且缺乏统一的查询语法。
图形数据库	Neo4J, InfoGrid, Infinite Graph	社交网络, 推荐系统等。专注于构建关系图谱	图结构	利用图结构相关算法。比如最短路径寻址, N 度关系查找等	很多时候需要对整个图做计算才能得到需要的信息, 而且这种结构不太好做分布式的集群方案。

资料来源: 各公司官网, CSDN@Coder-CT, 中信证券研究部

放眼全球: 从市场反馈来看, 多模型数据库更受企业青睐, 企业用户关注平台的兼容性与可扩展性。 DB-Engines 发布的 2022 年 10 月数据库管理系统流行程度排名显示, 排名前 8 的数据库管理系统均为多模型数据库, 支持文档模型, 键值模型, 图模型等多种数据模型。而随着排名逐渐靠后, 多模型数据库的比重也逐渐下降, 排名 11-20 的数据库管理系统中仅有 5 个多模型数据库。由此可见多模型数据库受企业欢迎的程度更高。国内公司凭借对于主流数据模型更高的兼容性, 有望在非关系型数据库领域与国际厂商同台竞争, 凭借大数据基础平台等核心产品实现国产替代。

表 18: 2022 年 10 月数据库管理系统流行程度排名及环比、同比变化情况

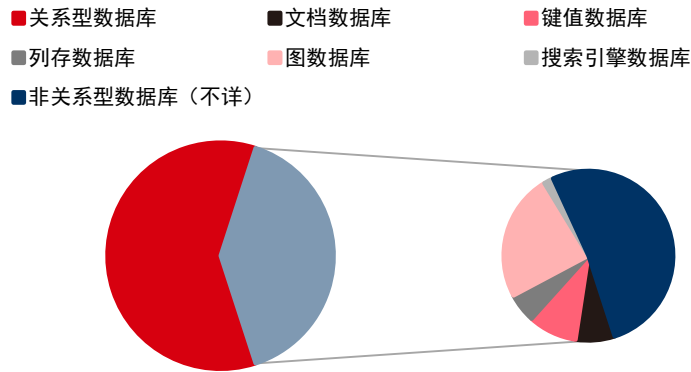
2022 年 10 月	2022 年 9 月	2021 年 10 月	数据库管理系统	数据模型
1	1	1	Oracle	关系型, 多模型
2	2	2	MySQL	关系型, 多模型
3	3	3	Microsoft SQL Server	关系型, 多模型
4	4	4	PostgreSQL	关系型, 多模型
5	5	5	MongoDB	文档型, 多模型
6	6	6	Redis	键值型, 多模型
7	7	8	Elasticsearch	搜索引擎, 多模型
8	8	7	IBM DB2	关系型, 多模型
9	9	11	Microsoft Access	关系型
10	10	9	SQLite	关系型

资料来源: DB-engines, 中信证券研究部 注: 截止时间为 2022 年 10 月 17 日

聚焦国内: 非关系型数据库占比呈现上升趋势, 关系型数据库在市场规模和产品数量上仍占据主流。 智研咨询数据显示, 2018 年我国关系型数据库市场规模占比高达 85%, 但呈现逐年下降趋势。综合多方关于我国数据库市场规模的数据 (中国信通院、IDC、艾瑞

咨询), 2021 年我国关系型数据库市场规模占比约为 60%。中国信通院数据显示, 截至 2021 年 6 月, 我国关系型数据库的产品数量占比约为 60%。

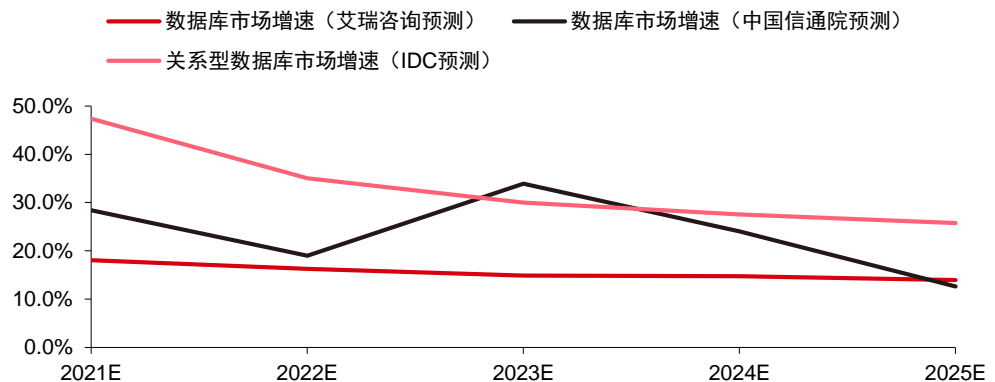
图 31: 我国数据库产品分布情况 (截至 2021 年 6 月)



资料来源: 中国信通院, 中信证券研究部

我们认为, 非关系型数据库与关系型数据库长期共存, 具备多模型兼容能力者有望胜出。根据 IDC、艾瑞咨询、中国信通院对于未来我国数据库市场规模的增速预测, 未来关系型数据库仍将占据主流市场, 但非关系型数据库也将成为行业生态中不可或缺的一部分, 二者将长期共存。处理半结构化、非结构化数据的治理水平或将成为未来衡量数据库厂商能力的重要指标之一, 具备多模型兼容能力者有望胜出。

图 32: 各机构关于我国数据库及关系型数据库市场增速预测



资料来源: 艾瑞咨询, 中国信通院, IDC, 中信证券研究部

以国产大数据厂商星环科技为例, 旗下核心大数据基础平台 (TDH) 中包含 9 种独立的存储引擎, 支持业界主流的 10 种存储模型。相关核心子产品主要包括关系型分析引擎 Inceptor、宽表数据库 Hyperbase、图数据库 StellarDB、搜索引擎 Scope、时空数据库 Spacture、时序数据库 TimeLyre、键值数据库 KeyByte、事件存储库 Event Store、文档数据库 DocStore, 对于多模型的兼容能力相较于海外主流厂商存在优势。

表 19: 星环大数据基础平台 (TDH) 与竞争对手支持数据模型对比

DBMS	Relational	Key-Value	Graph	Search Engines	Wide Column	Text, XML, Json	Object	Geospatial	Time Series	Event
TDH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IBM DB2	✓		✓			✓	✓		✓	✓
PostgreSQL	✓	✓				✓	✓	✓		
ArangoDB		✓	✓	✓		✓				
SQL Server	✓		✓			✓	✓			
MarkLogic			✓	✓		✓	✓			
Oracle NoSQL DB	✓	✓	✓						✓	
OrientDB		✓	✓			✓				
Oracle MySQL	✓	✓								✓
Cassabdra			✓		✓					✓
MongoDB		✓				✓	✓			
Oracle DB	✓					✓	✓			
InfluxDB										✓
Hbase					✓					

资料来源: 星环科技官网, 中信证券研究部

TDH 的多模型实现路径相较于其他主流产品具备优势。传统的多模型实现路径包括为每一种新数据模型开发独立完整的存算策略、用单一存储引擎支撑多个存储模型、在多种独立数据库之上提供统一的用户界面等, 这些策略暴露出存算资源消耗过高、存储引擎与存储策略不匹配、语言不一致提高开发难度等问题。星环 TDH 通过提供统一的 SQL 编译器层, 统一的分布式计算引擎层, 统一的分布式数据管理系统层以及统一的资源调度层, 将不同的数据库架构在统一多模型数据平台中, 跨库的关联分析不需要额外的数据导出导入过程, 避免了数据冗余。同时 TDH 提供 9 种独立的存储引擎子产品, 用户可以根据业务的需要, 随时增减不同的存储引擎, 做到资源按需分配。

表 20: 不同多模型实现路径

代表产品	IBM DB2	Oracle DB	MongoDB	MySQL	Couchbase	Marklogic	TDH8.0 (星环科技)
实现时间	2007	2013	2016	2011	2010	2008	2020
支持模型种类	6	3	3	3	2	4	10
实现路径	为每一种新数据模型开发独立完整的存算策略	用单一存储引擎支撑多个存储模型	在多种独立数据库之上提供统一的用户界面, 对底层多个数据库进行转发。	统一计算框架, 统一数据管理, 按需增减异构数据存储模型			
局限性/改进	存算耦合, 支持的模型越多, 系统的开发量和复杂度就越高, 消耗存算资源也较多。	由于不同计算数据模型对于存储的要求不同, 单一存储引擎无法随之匹配适合的存储策略, 从而限制了多模型数据库的性能。	由于底层多个数据库开发语言不一致, 导致了实际开发时的高难度, 排除故障的成本也较高。	用一套统一的架构来同时支持多模型、高可用与高性能, 并解决上述问题			

资料来源: CSDN@星环科技, 中信证券研究部

看生态: 开源闭源并存发展, 共促商业化生态繁荣

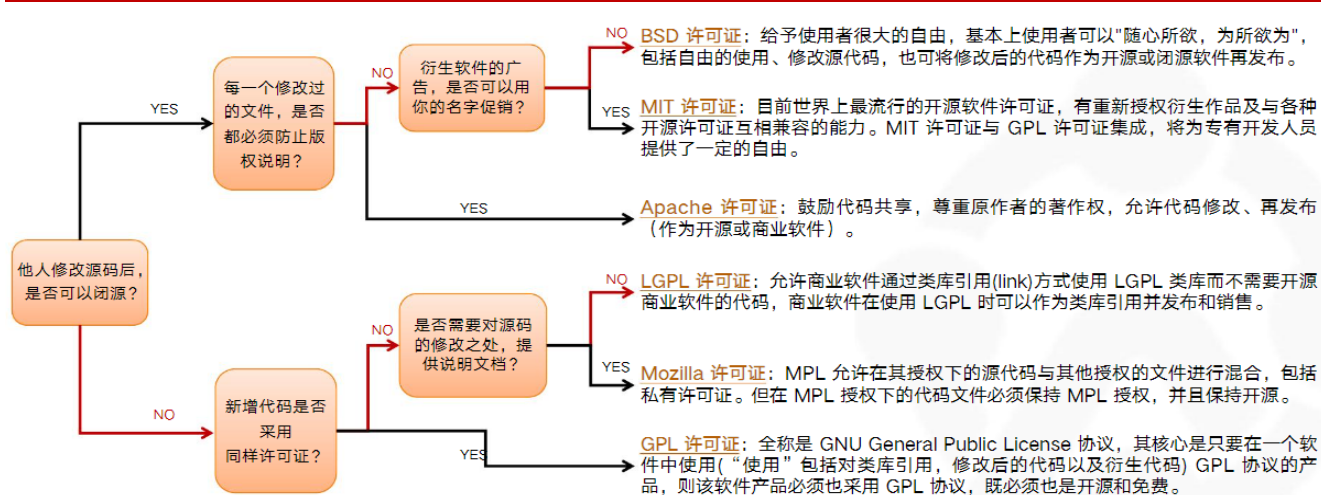
开源即开放源代码, 用户拥有基于源代码进行修改的权利。虽然源代码一般均免费提供使用者, 但开源系统的版权依然受到法律保护。开源软件标准权威发布机构 OSI (Open Source Initiative) 发布的对于开源的定义及要求主要包括如下三个方面:

内容方面：开放的源软件必须包含源代码，且必须确保源代码可被理解和可被运用；不得故意混淆源代码；开源代码需以源码或编辑后文件的形式传播。允许用户对开源项目及其他衍生分支进行修改，且必须允许其按照与初始软件相同的许可证发行。

传播规范方面：开源许可证不能限制开源软件的再传播，不得利用此条件进行收费。必须允许更改后的源代码所建立的程序发行许可证。当且仅当开源软件配合补丁文件一起发布时，开源许可证才可以限制源代码以修改后的形式发行。开源许可证不得限制其他铜许可软件一起发行的其他软件，不得限制特定软件的项目内容。

公平性、中立性准则：开源项目不得歧视任何研究领域、个人或团体。所有获得该项目的主体拥有所有附加到开源项目上的内容的使用权，无需当事方执行额外许可。开源许可必须独立于技术，不应指定任何特定的技术或接口。

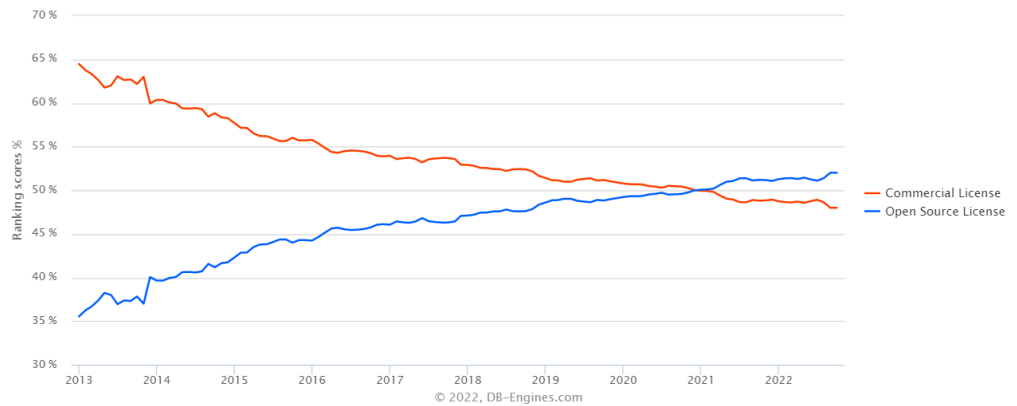
图 33：常见开源许可证一览



资料来源：墨天轮

从数据库厂商的视角来看，积极开源有助于构建服务生态，提高产品迭代速度和适配能力，及时捕捉用户需求的同时降低开发成本。通过构建开源生态社区，数据库厂商一方面可以依托广泛的开发者群体提高产品创新效率和迭代速度，节省自身开发成本和下游客户的 IT 成本，另一方面可以更加敏锐地捕捉新兴需求，并基于此迅速迭代产品抢占市场，亦可通过开源社区提高品牌影响力和行业话语权。

图 34：全球开源数据库与商业数据库数量占比



资料来源：DB-engines 注：截止时间为 2022 年 10 月 17 日

从用户视角来看，开源不同于免费，选型采购阶段的成本将转移到后续的开发部署和运维使用阶段。对用户来说，采用开源数据库可以一定程度节约选型采购阶段的 license 费用，但同时对于自身二次开发的能力提出了较高要求，数据库的部署、运维、迁移、配套升级等环节需要开源厂商提供数据库服务，也需要额外的人力投入和资金投入。此外，用户由于缺乏相关领域的重复实践经验，在应用场景和性能的扩展能力上可能不及直接采购商业数据库。因此政务、金融等对于数据安全性、一致性要求更高的场景倾向于使用商业数据库，越来越多厂商开始尝试“开源+商业”的混合策略。

图 35：开源数据库与商业数据库成本比较

	前期：选型采购	中期：开发部署	后期：运维使用
开源	<p>选型成本 开源产品/社区繁多造成选型障碍</p> <p>评测成本 需要结合需求做大量的POC测试</p>	<p>开发成本 自研案例缺失</p> <p>部署成本 问题不确定性</p> <p>迁移成本 源码、接口不兼容问题</p>	<p>扩展限制 企业级性能差</p> <p>Bug优化 部分Bug较多</p> <p>配套升级 缺乏配套产品</p> <p>人力投入 大量人力运维</p>
商业	<p>产品费用 商业license价格高昂</p> <p>咨询费用 前期咨询产生的费用</p>	<p>服务费用 包括个性化定制、部署、培训、后续运维的费用</p>	

资料来源：艾瑞咨询

我们认为，开源与商业并不冲突，未来国内数据库厂商将呈现出开源和商业共同繁荣的格局，数据库厂商将在积极拥抱开源生态的同时，兼顾自主可控及商业化需求。一方面，在传统数据库领域，我国相较于海外龙头企业仍有差距，开源生态能够帮助国内厂商更加快速实现追赶；另一方面，近年来数据库领域持续迸发出新技术、新应用、新模式，参与开源项目能够帮助企业更快把握技术革新与市场机遇，实现生态构建的正向循环。此外在信创大背景下，开源生态能够促进国产数据库上下游的适配能力，加快自主可控的步伐。

表 21：国产数据库开源情况

名称	贡献者	开源时间	已处理 issue	已合并 Pull 请求	Watch	Star	Fork	Contributors
TiDB	PingCAP	2015 年	9043	22110	1300	31200	5100	741
Apache Doris	百度	2017 年	2733	5454	227	4300	1400	314
TDengine	涛思数据	2019 年	2426	9417	679	18300	4300	181
Nebula Graph	悦数科技	2020 年	1399	2439	170	7400	827	79
StarRocks	鼎石科技	2021 年	1376	3787	82	2500	529	104
云树@Shard	爱可生	2016 年	1109	1988	76	925	310	29
Apache HAWQ	偶数科技	2015 年	-	1505	95	648	320	76
openGauss	华为	2019 年	1056	1167	277	817	655	165
OceanBase	OceanBase	2021 年	484	234	91	4300	963	126
RadonDB	青云	2018 年	294	444	86	1600	214	9
gStore	北京大学	2014 年	65	46	38	515	187	26
AliSQL	阿里巴巴	2016 年	61	1	497	4500	856	3
TDSQL	腾讯云	2019 年	58	17	74	1200	237	18
PolarDB	阿里云	2021 年	45	100	56	2300	350	55
IvorySQL	瀚高	2021 年	28	56	10	83	47	63
SequoiaDB	巨杉科技	2015 年	3	68	22	272	111	-

资料来源：Github, Gitee, 墨天轮, 中信证券研究部 注：截止时间为 2022 年 5 月

■ 风险因素

国产数据库技术迭代不及预期。数据库行业中，海外巨头 Oracle、微软、Google 等企业技术领先，客户覆盖广泛，亦有诸如 MongoDB、Snowflake 等独立数据库公司，技术能力全球领先。国产数据库公司与海外巨头存在一定技术差距，其生存和发展很大程度上取决于是否能够及时、高效地进行技术更新与产品升级，以满足客户不断升级的需求。若国产数据库未能及时把握技术发展趋势、技术迭代不及预期，不排除国内外竞争对手或潜在竞争对手率先在新技术领域取得重大突破，而推出更先进、更具竞争力的技术和产品，将对国产数据库产生不利影响。

下游行业需求不及预期。近年来，国家产业政策支持对基础软件产业发展有较大程度的影响，信创产业正加速发展。同时，各行业快速成长的过程中对数据管理的需求不断增长，带来数据库行业快速发展。若未来出现国家信息化趋势放缓，金融、电信、政务等领域核心业务系统信息化进程和相关政策支持力度不及预期等不利变化，国产数据库行业将面临收入增长放缓的风险。

数据库行业参与者众多，市场竞争加剧的风险。根据墨天轮统计，2022 年 6 月，中国数据库厂商已超过 200 家，参与者包括大型云厂商、独立软件开发商、运营商、基础设施类厂商等。同时，随着用户对数据存储和分析服务的需求不断增长，各参与者加强争夺市场份额，可能导致行业竞争加剧。如果相关公司在市场竞争中不能有效保持技术先进水平，不能充分利用现有的市场影响力和差异化竞争策略，无法在当前市场高速发展的态势下迅速扩大自身规模并增强资金实力，可能导致其市场地位出现下滑。

开源或免费软件的竞争风险。开源软件和免费软件具有较低的应用成本，仍然在当前软件行业中有较为广泛的应用基础，对商用软件的推广构成一定的竞争。商用软件通常针对客户需求进行开发，有较高的技术门槛，在对安全、性能等方面有刚性需求场景中，商

用软件仍然占据主流。目前传统关系型数据库仍以商用产品为主，NoSQL 数据库作为新兴数据库产品，在应用中，免费及开源产品使用较为广泛。若未来免费或开源产品的技术快速迭代，应用范围更广，可能加剧市场竞争。

疫情及宏观环境扰动。近年来，地缘政治、新冠疫情等因素使得全球经济增长具有更大的不确定性，数据库面向政府、金融、运营商、教育、医疗等各行各业，受宏观经济影响较大，产业链中各合作伙伴也无可避免地受到影响。若未来地缘政治冲突持续存在、新冠疫情不能得到抑制，宏观经济出现波动，则将对行业发展造成不利影响。

■ 投资建议

数据库是数字化时代的底层基石，是各厂商重点布局的关键战略领域。安全可控势在必行，数据库国产替代加速开展，以党政为代表的国产替代先行，并不断向金融、电信等领域拓展。通过对关键的三个问题的深度解读，我们认为：1) 国产数据库厂商已具备较强的 OLTP 数据库替代能力；2) OLAP 数据库未来仍将会持续快速发展，技术能力强者有望胜出；3) 数据库行业朝着多元化方向演进，包括非关系型数据库、分布式数据库、HTAP 混合负载、云数据库等，与传统数据库类型形成良好的市场互补关系。数据库产业料将持续快速成长，带来诸多投资机遇。**建议关注：**

1) 已上市及拟上市：**达梦数据（中国软件）、星环科技、人大金仓（太极股份）、海量数据、神舟通用（航天软件）、科蓝软件。**

2) 未上市公司：PingCAP、巨杉数据库、南大通用、易鲸捷、云和恩墨等。

分析师声明

主要负责撰写本研究报告全部或部分内容的分析师在此声明：(i) 本研究报告所表述的任何观点均精准地反映了上述每位分析师个人对标的证券和发行人的看法；(ii) 该分析师所得报酬的任何组成部分无论是在过去、现在及将来均不会直接或间接地与研究报告所表述的具体建议或观点相联系。

一般性声明

本研究报告由中信证券股份有限公司或其附属机构制作。中信证券股份有限公司及其全球的附属机构、分支机构及联营机构（仅就本研究报告免责条款而言，不含 CLSA group of companies），统称为“中信证券”。

本研究报告对于收件人而言属高度机密，只有收件人才能使用。本研究报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。本研究报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。中信证券并不因收件人收到本报告而视其为中信证券的客户。本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断并自行承担投资风险。

本报告所载资料的来源被认为是可靠的，但中信证券不保证其准确性或完整性。中信证券并不对使用本报告或其所包含的内容产生的任何直接或间接损失或与此有关的其他损失承担任何责任。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可跌可升。过往的业绩并不能代表未来的表现。

本报告所载的资料、观点及预测均反映了中信证券在最初发布该报告日期当日分析师的判断，可以在不发出通知的情况下做出更改，亦可因使用不同假设和标准、采用不同观点和分析方法而与中信证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。中信证券并不承担提示本报告的收件人注意该等材料的责任。中信证券通过信息隔离墙控制中信证券内部一个或多个领域的信息向中信证券其他领域、单位、集团及其他附属机构的流动。负责撰写本报告的分析师的薪酬由研究部门管理层和中信证券高级管理层全权决定。分析师的薪酬不是基于中信证券投资银行收入而定，但是，分析师的薪酬可能与投行整体收入有关，其中包括投资银行、销售与交易业务。

若中信证券以外的金融机构发送本报告，则由该金融机构为此发送行为承担全部责任。该机构的客户应联系该机构以交易本报告中提及的证券或要求获悉更详细信息。本报告不构成中信证券向发送本报告金融机构之客户提供的投资建议，中信证券以及中信证券的各个高级职员、董事和员工亦不为（前述金融机构之客户）因使用本报告或报告载明的内容产生的直接或间接损失承担任何责任。

评级说明

投资建议的评级标准		评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即：以报告发布日后的 6 到 12 个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准；美国市场以纳斯达克综合指数或标普 500 指数为基准；韩国市场以科斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅 20%以上
		增持	相对同期相关证券市场代表性指数涨幅介于 5%~20%之间
		持有	相对同期相关证券市场代表性指数涨幅介于-10%~5%之间
		卖出	相对同期相关证券市场代表性指数跌幅 10%以上
	行业评级	强于大市	相对同期相关证券市场代表性指数涨幅 10%以上
		中性	相对同期相关证券市场代表性指数涨幅介于-10%~10%之间
		弱于大市	相对同期相关证券市场代表性指数跌幅 10%以上

特别声明

在法律许可的情况下，中信证券可能（1）与本研究报告所提到的公司建立或保持顾问、投资银行或证券服务关系，（2）参与或投资本报告所提到的公司的金融交易，及/或持有其证券或其衍生品或进行证券或其衍生品交易。本研究报告涉及具体公司的披露信息，请访问 <https://research.citicsinfo.com/disclosure>。

法律主体声明

本研究报告在中华人民共和国（香港、澳门、台湾除外）由中信证券股份有限公司（受中国证券监督管理委员会监管，经营证券业务许可证编号：Z20374000）分发。本研究报告由下列机构代表中信证券在相应地区分发：在中国香港由 CLSA Limited（于中国香港注册成立的有限公司）分发；在中国台湾由 CL Securities Taiwan Co., Ltd. 分发；在澳大利亚由 CLSA Australia Pty Ltd.（商业编号：53 139 992 331/金融服务牌照编号：350159）分发；在美国由 CLSA（CLSA Americas, LLC 除外）分发；在新加坡由 CLSA Singapore Pte Ltd.（公司注册编号：198703750W）分发；在欧洲经济区由 CLSA Europe BV 分发；在英国由 CLSA（UK）分发；在印度由 CLSA India Private Limited 分发（地址：8/F, Dalamal House, Nariman Point, Mumbai 400021；电话：+91-22-66505050；传真：+91-22-22840271；公司识别号：U67120MH1994PLC083118）；在印度尼西亚由 PT CLSA Sekuritas Indonesia 分发；在日本由 CLSA Securities Japan Co., Ltd. 分发；在韩国由 CLSA Securities Korea Ltd. 分发；在马来西亚由 CLSA Securities Malaysia Sdn Bhd 分发；在菲律宾由 CLSA Philippines Inc.（菲律宾证券交易所及证券投资者保护基金会）分发；在泰国由 CLSA Securities (Thailand) Limited 分发。

针对不同司法管辖区的声明

中国大陆：根据中国证券监督管理委员会核发的经营证券业务许可，中信证券股份有限公司的经营经营范围包括证券投资咨询业务。

中国香港：本研究报告由 CLSA Limited 分发。本研究报告在香港仅分发给专业投资者（《证券及期货条例》（香港法例第 571 章）及其下颁布的任何规则界定的），不得分发给零售投资者。就分析或报告引起的或与分析或报告有关的任何事宜，CLSA 客户应联系 CLSA Limited 的罗鼎，电话：+852 2600 7233。

美国：本研究报告由中信证券制作。本研究报告在美国由 CLSA（CLSA Americas, LLC 除外）仅向符合美国《1934 年证券交易法》下 15a-6 规则界定且 CLSA Americas, LLC 提供服务的“主要美国机构投资者”分发。对身在美国的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所述任何观点的背书。任何从中信证券与 CLSA 获得本研究报告的接收者如果希望在美国交易本报告中提及的任何证券应当联系 CLSA Americas, LLC（在美国证券交易委员会注册的经纪交易商），以及 CLSA 的附属公司。

新加坡：本研究报告在新加坡由 CLSA Singapore Pte Ltd.，仅向（新加坡《财务顾问规例》界定的）“机构投资者、认可投资者及专业投资者”分发。就分析或报告引起的或与分析或报告有关的任何事宜，新加坡的报告收件人应联系 CLSA Singapore Pte Ltd，地址：80 Raffles Place, #18-01, UOB Plaza 1, Singapore 048624，电话：+65 6416 7888。因您作为机构投资者、认可投资者或专业投资者的身份，就 CLSA Singapore Pte Ltd. 可能向您提供的任何财务顾问服务，CLSA Singapore Pte Ltd 豁免遵守《财务顾问法》（第 110 章）、《财务顾问规例》以及其下的相关通知和指引（CLSA 业务条款的新加坡附件中证券交易服务 C 部分所披露）的某些要求。MCI（P）085/11/2021。

加拿大：本研究报告由中信证券制作。对身在加拿大的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所载任何观点的背书。

英国：本研究报告归属于营销文件，其不是按照旨在提升研究报告独立性的法律要件而撰写，亦不受任何禁止在投资研究报告发布前进行交易的限制。本研究报告在英国由 CLSA（UK）分发，且针对由相应本地监管规定所界定的在投资方面具有专业经验的人士。涉及到的任何投资活动仅针对此类人士。若您不具备投资的专业经验，请勿依赖本研究报告。

欧洲经济区：本研究报告由荷兰金融市场管理局授权并管理的 CLSA Europe BV 分发。

澳大利亚：CLSA Australia Pty Ltd（“CAPL”）（商业编号：53 139 992 331/金融服务牌照编号：350159）受澳大利亚证券与投资委员会监管，且为澳大利亚证券交易所及 CHI-X 的市场参与主体。本研究报告在澳大利亚由 CAPL 仅向“批发客户”发布及分发。本研究报告未考虑收件人的具体投资目标、财务状况或特定需求。未经 CAPL 事先书面同意，本研究报告的收件人不得将其分发给任何第三方。本段所称的“批发客户”适用于《公司法（2001）》第 761G 条的规定。CAPL 研究覆盖范围包括研究部门管理层不时认为与投资者相关的 ASX All Ordinaries 指数成分股、离岸市场上市证券、未上市发行人及投资产品。CAPL 寻求覆盖各个行业中与其国内及国际投资者相关的公司。

印度：CLSA India Private Limited，成立于 1994 年 11 月，为全球机构投资者、养老基金和企业提供股票经纪服务（印度证券交易委员会注册编号：INZ000001735）、研究服务（印度证券交易委员会注册编号：INH000001113）和商人银行服务（印度证券交易委员会注册编号：INM000010619）。CLSA 及其关联方可能持有标的公司的债务。此外，CLSA 及其关联方在过去 12 个月内可能已从标的公司收取了非投资银行服务和/或非证券相关服务的报酬。如需了解 CLSA India “关联方”的更多详情，请联系 Compliance-India@clsa.com。

未经中信证券事先书面授权，任何人不得以任何目的复制、发送或销售本报告。

中信证券 2022 版权所有。保留一切权利。