

大数据系列专题（2）： 国产数据库百花齐放，崛起正当时

行业研究 · 深度报告
投资评级：超配（维持评级）

证券分析师：熊莉
xiongli1@guosen.com.cn
S0980519030002

证券分析师：朱松
zhusong@guosen.com.cn
S0980520070001

联系人：黄浩峻
huanghaojun@guosen.com.cn

- **数据库是计算机系统的三大基础软件之一。**数据库管理系统是“按照数据结构来组织、存储和管理数据的仓库”，是信息化时代、大数据时代中各行各业不可或缺的重要基础软件。不同类别的数据库适用于不同的应用场景，常见的数据库分类方式：1) 按数据模型划分，可分为关系型数据库、非关系型数据库（NoSQL，包括键值型、文档型、图、对象型等）；2) 按应用场景划分，可分为OLTP事务型数据库、OLAP分析型数据库、HTAP混合型数据库；3) 按架构模型划分，可分为集中式数据库和分布式数据库；4) 按部署模式划分，可分为本地数据库和云数据库。
- **数据平台从BI向AI过渡。**数据平台主要是指数据分析平台，其分析内部和外部其它系统生成的各种原始数据，对这些数据进行各种分析挖掘以生成衍生数据，从而支持企业进行数据驱动的决策。整个数据平台的发展来看，就是随着企业信息化和数字化的逐渐推进，从数据库，数据仓库，数据湖到数据湖仓逐渐演进的。
- **国内数据库市场广阔，未来将逐步走向集中。**目前根据信通院的统计数据，2020年国内的数据库规模为240.9亿元，占全球数据库市场规模的比例为5.2%，未来到2025年的时候，国内的数据库规模将占到全球数据库规模的12.5%。目前国内的数据库市场呈现百花齐放状态，主要分为国内厂商和海外厂商两大类，其中国内数据库厂商主要包括传厂商、初创厂商、云厂商以及跨界厂商，海外厂商，主要包括Oracle、IBM及云厂商。目前，国产市场目前处于跑马圈地状态，未来竞争格局有望走向集中，观测的指标有人员数量、产品性能、服务质量和兼容性。
- **国内数据库呈现4大发展趋势。**趋势主要分为：1) 国产化，一方面政策环境较为友好，尤其是对核心技术领域的支持政策较多，另一方面我国的数据库开发水平逐步提升，国内厂商的市场份额从2007年的不足5%到2020年的48%，取得了较大进展；2) 开源数据库，目前国内主流的产品都已经开源，如PingCAP的TiDB产品、百度的Apache Doris、华为的openGauss产品等都已开源；3) 数据库云化，随着云计算基础设施的完善，越来越多的数据库服务在云端提供，从海外数据库厂商的份额变化可以得到佐证；4) 多模数据库，面对日益丰富的数据类型，在一个数据库里面实现多模型数据的存储和管理，即多模数据库，在业内得到广泛讨论。
- **数据库相关公司梳理。**数据库公司主要涉及拓尔思、星环科技、PingCAP、华为、达梦、腾讯、云和恩墨、易鲸捷、巨杉等。
- **风险提示：**研发成果商业化转化不确定性的风险；数据库厂商恶性竞争的风险；数据库核心成员流失的风险；国产替代进度不及预期的风险。

1、什么是数据库

- 1.1 数据库是计算机系统的三大核心基础软件之一
- 1.2 数据库发展史：数据库与信息技术的发展相互促进
- 1.3 数据库分类：不同类别数据库适用于不同场景

2、什么是数据平台

- 2.1 数据平台发展史：从BI到AI，数据时代带动底层架构持续迭代
- 2.2 数据平台分类：从数据仓库、数据湖到湖仓一体

3、数据库的竞争格局

- 3.1 数据库市场空间广阔
- 3.2 国内数据库市场的主要玩家
- 3.3 国内数据库市场有望走向集中

4、数据库的未来发展趋势

- 4.1 趋势一：国产化
- 4.2 趋势二：开源和商业数据库并存
- 4.3 趋势三：数据库云化
- 4.4 趋势四：多模数据库

5、数据库相关公司梳理

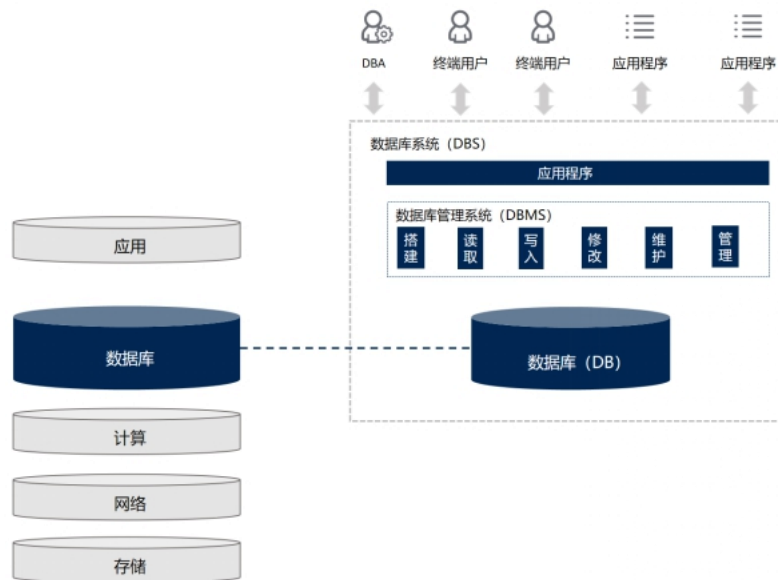
- 5.1 拓尔思：以“语义智能+”为战略，布局搜索引擎数据库赛道
- 5.2 星环科技：是国内领先的企业级大数据基础软件开发商
- 5.3 PingCAP：专注于企业级开源分布式数据库的公司
- 5.4 华为GaussDB：基于openGAUSS企业级分布式关系型数据库
- 5.5 达梦数据库：高性能数据库管理系统
- 5.6 TDSQL：企业级分布式数据库
- 5.7 云和恩墨：智能的数据技术提供商
- 5.8 易鲸捷：中国分布式数据库的领军企业
- 5.9 巨杉数据库：湖仓一体，金融级分布式数据库

1、什么是数据库

1.1 数据库是计算机系统的三大核心基础软件之一

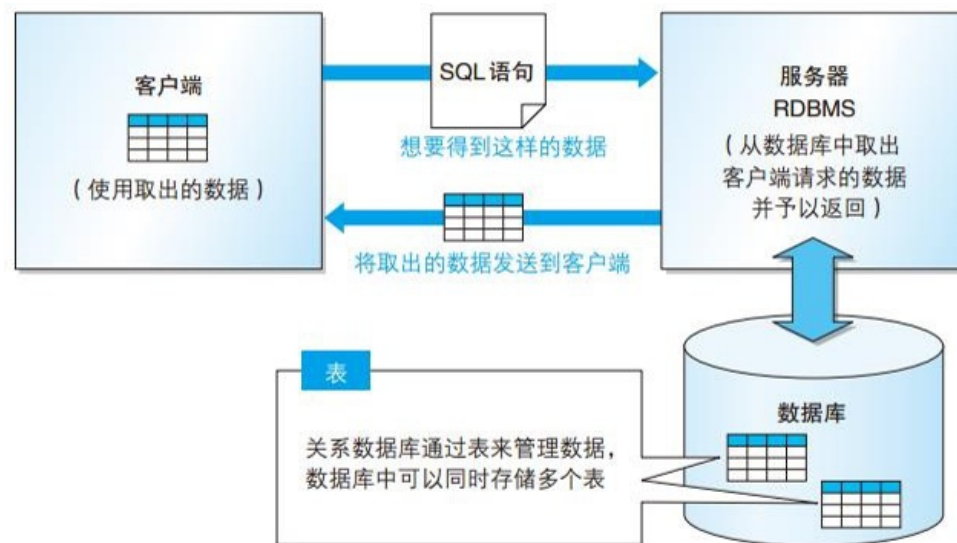
- **数据库是计算机系统的三大核心基础软件之一。**数据是数据库中存储的基本对象，包括数字、图像、音频等形式，在进行逐级抽象后存储在数据库中，通常由数据库管理系统（DBMS）来控制，DBMS充当数据库与其用户或程序之间的接口，允许用户检索、更新和管理信息的组织和优化方式。在现实中，数据、DBMS及关联应用一起被称为数据库系统，通常简称为数据库。数据库管理系统是“按照数据结构来组织、存储和管理数据的仓库”，是信息化时代、大数据时代中各行各业不可或缺的重要基础软件。
- **SQL在1970年被提出，是目前关系数据库标准的结构化查询语言。**SQL（Structured Query Language，结构化查询语言）是用于管理关系数据库管理系统（RDBMS），SQL的范围包括数据插入、查询、更新和删除，数据库模式创建和修改，以及数据访问控制。SQL由1970年提出，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，目前几乎所有的关系数据库都使用 SQL 编程语言来查询、操作和定义数据，进行数据访问控制。

数据库架构示意图



资料来源：艾瑞咨询、国信证券经济研究所整理

如何使用SQL语言查询数据库

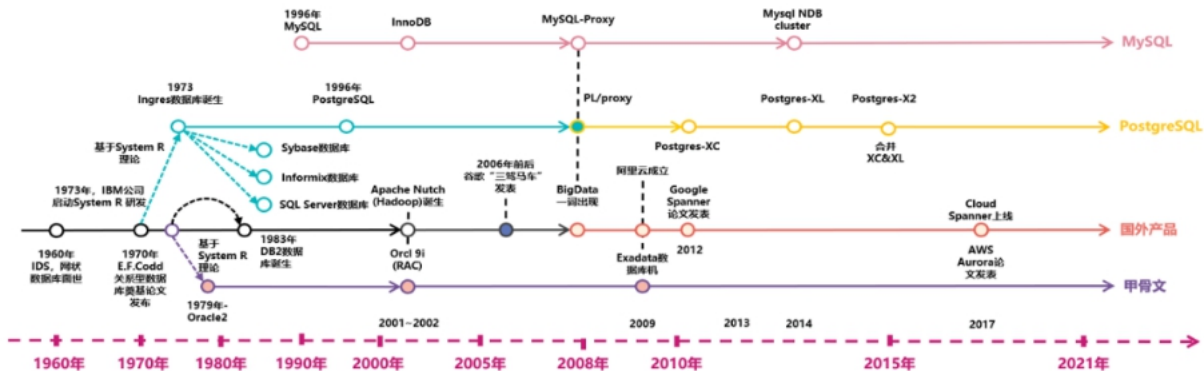


资料来源：CSDN、国信证券经济研究所整理

1.2 数据库发展史：数据库与信息技术的发展相互促进

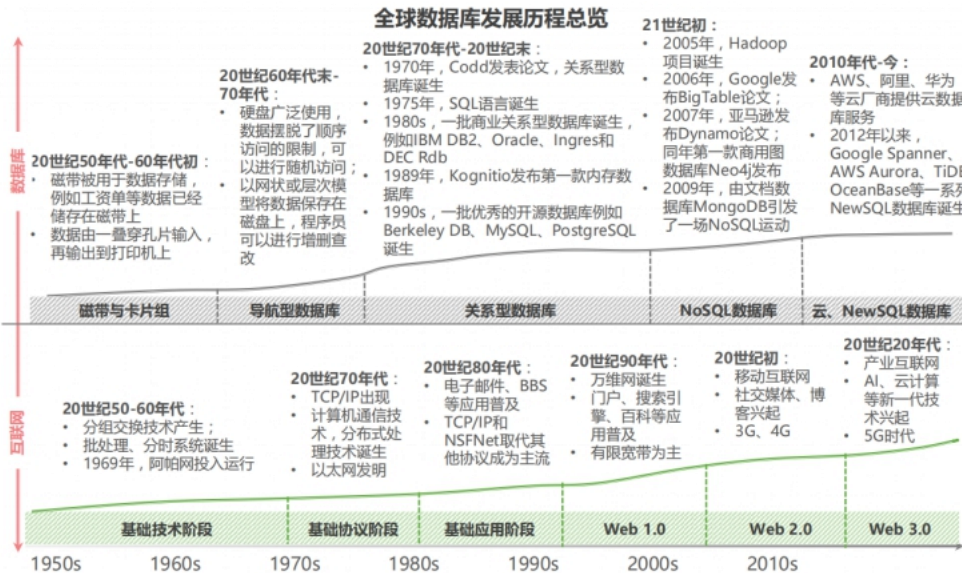
- **关系型事务数据库诞生于1970年。**1970年，IBM实验室的Edgar Frank Codd发布论文奠定关系型数据库的基石；Ingres原型在1974年诞生，为后续大量基于其源码开发的PostgreSQL、Sybase、Informix和Tandem等著名产品打下坚实基础。1977年，Oracle前身SDL成立，并于1978年发布Oracle第一个版本。1986年，美国国家标准局（ANSI）数据库委员会批准SQL作为数据库语言的美国标准并公布标准SQL文本。
- **OLAP分析型数据库与数仓在1990年前后出现，数据湖诞生于2010年。**1988年，IBM公司的研究员提出了数据仓库（Data Warehouse）新概念，数据仓库之父W. H. Inmon在1991年出版数据仓库经典作品《构建数据仓库》，标志着数据仓库概念的确立，在1993年Codd提出了OLAP（On-Line Analysis Processing联机分析处理）理念，2003年Grennplum公司成立并推出MPP数仓产品，之后将其开源，再到2012年第一个云原生数仓公司Snowflake成立；谷歌自2003开始发布了关于GFS、MapReduce 和BigTable三篇技术论文，Hadoop项目成立，后成为数据湖的重要组件，2010年由Pentaho的创始人兼首席技术官James Dixon提出数据湖概念。
- **NoSQL兴起于移动互联网的Web 2.0时代。**1998年Carlo Strozzi率先提出NoSQL概念，2009年MongoDB正式推出了文档型数据库MongoDB 1.0，在移动互联网快速发展下，带动非关系型数据库需求快速爆发。

数据库主要技术流派的发展路线



资料来源：中国信通院、国信证券经济研究所整理

全球数据库发展历程



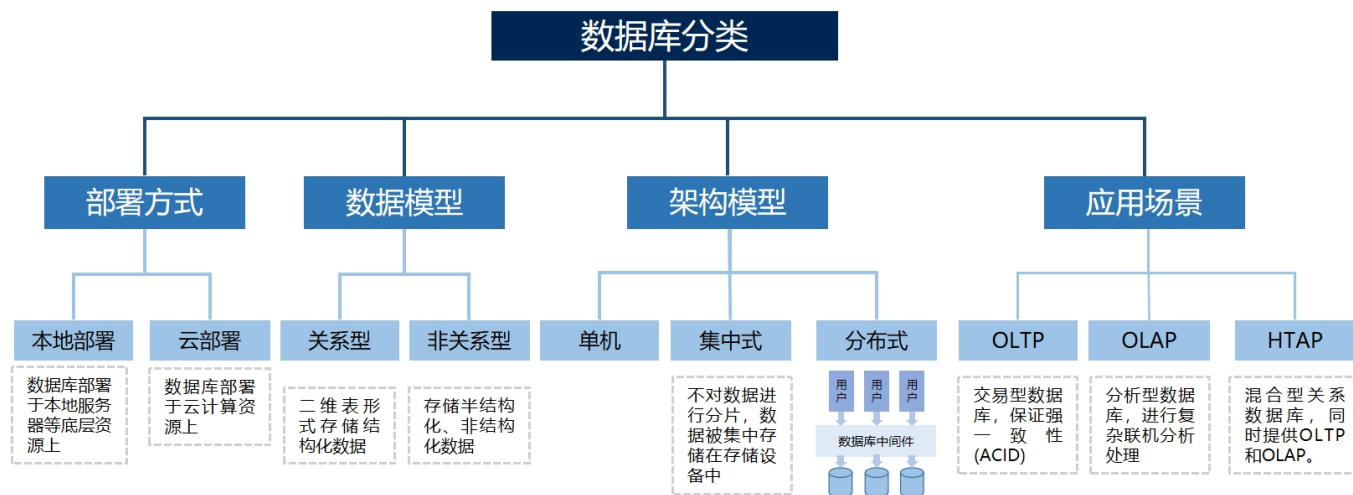
资料来源：艾瑞咨询、国信证券经济研究所整理

1.3 数据库分类：不同类别数据库适用于不同场景

不同类别的数据库适用于不同的应用场景，常见的数据库分类方式如下：

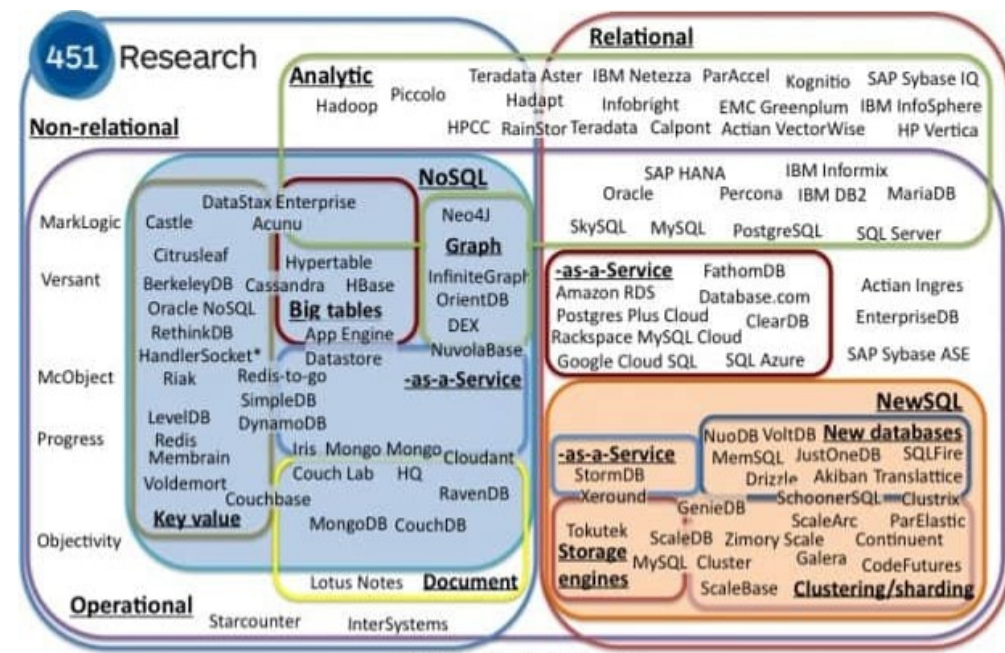
- 按数据模型划分，可分为关系型数据库、非关系型数据库（NoSQL，包括键值型、文档型、图、对象型等）；
- 按应用场景划分，可分为OLTP事务型数据库、OLAP分析型数据库、HTAP混合型数据库；
- 按架构模型划分，可分为集中式数据库和分布式数据库；
- 按部署模式划分，可分为本地数据库和云数据库；

数据库的分类方式



资料来源：沙利文、国信证券经济研究所整理

数据库的生态图谱



资料来源：451 Research、国信证券经济研究所整理

(1) 按数据结构，可以分为关系型数据库和非关系型数据库

按管理数据的结构，数据库可分为关系型数据库和非关系型数据库：

- **关系型数据库是指采用二维表格的关系模型来组织数据的数据库系统。**关系型数据库是由二维表及其之间的联系所组成的一个数据组织。直接使用通用的SQL语言，使得操作关系型数据库非常方便，关系型数据库遵循ACID规则。
- **非关系型数据库是用于非关系模型来组织数据的数据库系统，如键值、列、文档、图形等。**非关系型数据库在1998年被首次提出，非关系数据库试图去摆脱传统关系数据库的约束限制，比如像数据的一致性、合并内存中的数据处理以及简化数据模型。NoSQL数据库用非结构化数据的来快速存储和检索，不使用为关系数据库提供支撑的结构化数据图表，这与关系型数据库结构化数据是不同的。NoSQL数据库放宽或取消了一些ACID的规则，以达到更好的性能和更大的灵活性。NoSQL数据库认为 ACID 属性的要求过于严格，大多数NoSQL数据库遵循BASE原则来代替。

关系型数据库与非关系型数据库		
对比	关系型数据库（SQL）	非关系型数据库（NoSQL）
代表数据库	Oracle、Mysql等	Redis、MongoDB、Hbase等
遵循规则	ACID	BASE（Basically Available -基本可用、Soft-state-软状态、Eventual Consistency -最终一致性）
存储数据的格式	基础类型	存储格式是key-value形式、文档形式、图片形式等等，所以可以存储基础类型以及对象或者是集合等各种格式
扩展性	join这样的多表查询机制的限制导致扩展性有限	数据之间没有耦合性，所以非常容易水平扩展
事务	对于安全性能很高的数据访问要求得以实现	无事务处理或者弱事务
数据一致性	支持强一致性、最终一致性等	通常强调数据的最终一致性
SQL	支持SQL语言，可用于复杂查询	不使用SQL或者不仅仅是SQL
读写效率	海量数据的高效率读写的问题	无需经过sql层的解析，读写性能很高

资料来源：华为云、国信证券经济研究所整理

(1) 按数据结构，可以分为关系型数据库和非关系型数据库（续）

非关系型数据库适用于超大规模和高并发的场景。根据数据类型，可以分为键值存储数据库、列存储数据库、文档型数据库、图形数据库。随着web2.0的兴起，超大规模和高并发的社会型网络服务类型的动态网站对数据库高并发读写、可扩展性和高可用性的要求，以及对海量数据存储和访问在效率上的需求，非关系型数据库已经成为一个“互联网经济”首选方案。常见的NoSQL数据库包括键值存储数据库、列存储数据库、文档型数据库、图形数据库等

- **键值存储：**数据被存储为键-值对（键-阵列对），键值存储是所有NoSQL数据库中最简单的数据库。典型代表：Redis、Memcached、Apache Ignite、Riak等；
- **列存储：**也可称为柱状存储，列式存储其中包含连续存储值数列的行的集合，所有的列可通过列式排列，它们将数据列存储在一起，而不是行。典型代表：Hbase、Cassandra、Scylla等；
- **文档存储：**可用于管理，检索和存储文字文档的集合，将每个键与称为文档的复杂数据结构配对，文档可以包含键数组对、键值对甚至嵌套文档。典型代表：MongoDB、Apache CouchDB、ArangoDB、Couchbase、Cosmos DB、IBM Domino、MarkLogic、OrientDB等；
- **图形存储：**这些存储关于图形、网络的信息，例如社会关系、路线图、交通链接。典型代表：Neo4J、InfoGrid、Infinite Graph。

常见的非关系数据库类型及特点

分类	代表产品	典型应用场景	数据模型	优点	限制性
键值数据库	Redis、MemCahed	缓存用户信息，会话信息，配置文件，购物车等	Key指向Value，通常基于Hash table实现	查找速度快	数据无结构化，字符串或者二进制数据
列式数据库	HBase、Cassandra	日志 博客平台	列族式存储	查找速度快，可容易分布式扩展	不适合随机更新，不适合做有删除和更新的实时操作
文档型数据库	MongoDB、DouchDB	日志，可以存储不同模式的日志信息。基于弱模式的数据分析	和K-V类似，Value的数据结构要求不严格，无需预先定义表结构	表结构可变，扩展性好，适合非结构化对象	有些产品不支持事务操作
图数据库	Neo4J、InfoGrid、Infinite Graph	推荐引擎 关系图谱	图结构	借助图论算法处理特定领域问题	非图领域的应用受限

资料来源：华为云、国信证券经济研究所整理

(2) 按**应用场景**，可以分为OLAP、OLTP和HTAP三类

数据库按其应用场景可分为OLTP事务型数据库、OLAP分析型数据库、HTAP混合型数据库三类：

- **OLTP事务型数据库主要面向事务类场景，实时性要求高，数据量通常不大，面向一线业务人员。**OLTP (On-Line Transaction Processing, 联机事务处理)，主要面向事务型应用的场景，OLTP系统主要使用关系模型，保证强一致性，保证事务的ACID特性，面向一线业务人员，支持多并发、实时、快速地增删查改，例如银行交易、零售电商、车票预订等；
- **OLAP分析型数据库主要面向分析类场景，历史数据为主，数据量大，面向分析与管理人员。**OLAP (On-Line Analytical Processing, 联机分析处理)，主要面向分析型应用的场景。OLAP系统可以高速多维分析来自数据仓库、数据集市或者数据湖的数据，可使用关系型或者非关系型的数据库，主要面向分析师和管理者，支持对历史数据的复杂分析操作，从而赋能企业商业智能决策。

OLTP事务型数据库与OLAP分析型数据库的对比

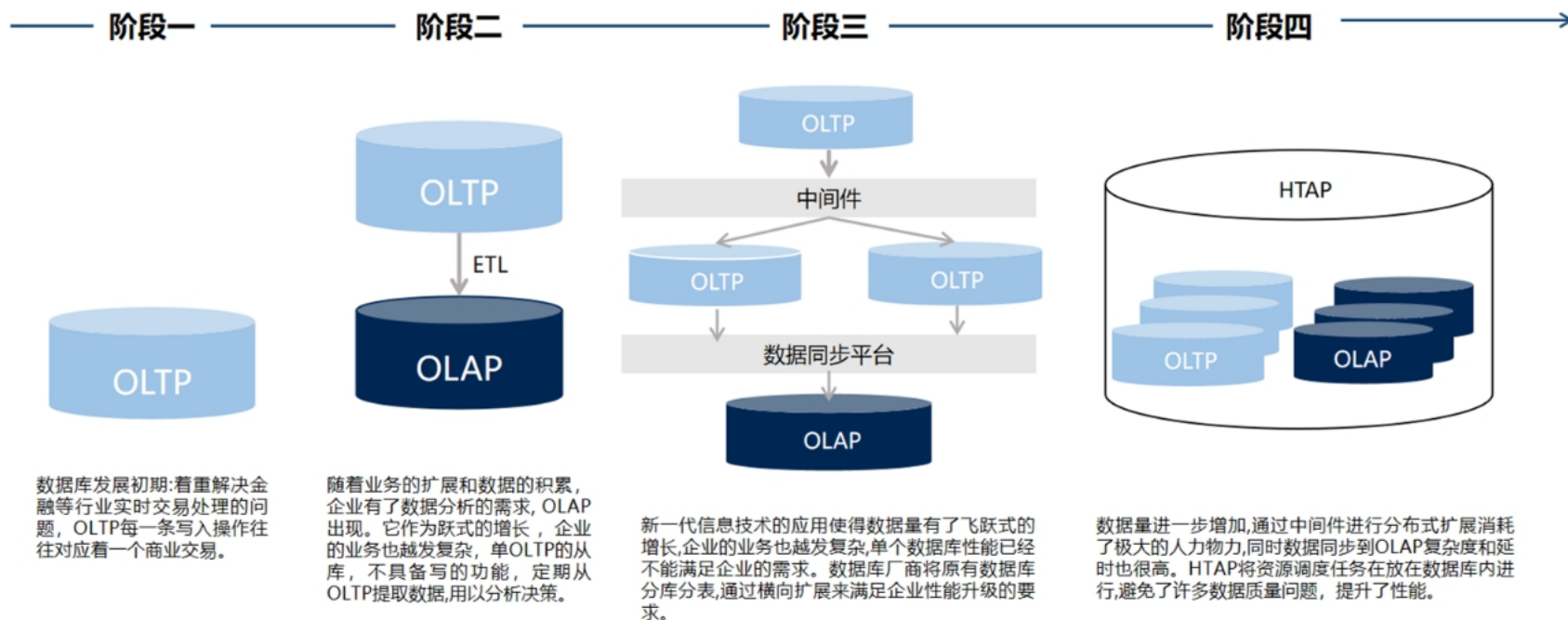
	OLTP	OLAP
产品定位	支持实时交易数据的存储、更新、共享	通过数据分析现状, 发现趋势, 支持决策
操作	基于INSERT, UPDATE, DELETE命令	基于SELECT命令聚合数据用以分析
数据量	实时数据, 通常较小	聚合历史数据, 较大
并发访问量	高	低
响应时间	毫秒	秒, 分钟或者小时(取决于处理的数据量)
存储结构	通常为行存储	通常为列存储
备份	需要定期备份以确保业务连续性	可以以OLTP数据库重新加载丢失的数据, 以代替常规备份
可视化	日常业务交易列表视图	多维视图满足分析需求
典型适用场景	快速处理高并发、小批量的数据	使用复杂的查询处理大量数据
主要用户	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：艾瑞咨询、国信证券经济研究所整理

(2) 按应用场景，可以分为OLAP、OLTP和HTAP三类（续）

- **HTAP是混合了OLTP和OLAP场景。**HTAP (Hybrid transaction/analytical processing, 混合事务/分析处理) 是混合OLTP和OLAP业务同时处理的系统，在对新旧数据进行OLAP分析的情况下增加事务的处理来对数据进行更新。HTAP概念由2014年Gartner公司提出，HTAP是一种新兴的应用体系结构，它打破了事务处理和分析之间的“墙”，它支持更多的信息和“实时业务”的决策。
- **HTAP基于分布式架构，支持弹性扩容，可按需扩展吞吐或存储，轻松应对高并发、海量数据场景。**随着数据价值的进一步挖掘，企业对数据库系统也提出了新的要求。现阶段企业为满足交易处理和分析的需求，往往采用OLTP+OLAP的组合方案。但二者之间往往存在时延，无法满足企业实时分析的需求；同时管理两个平台往往需要组建两支团队，运维成本高。HTAP混合型数据库基于新的计算存储框架，能够同时支撑OLTP和 OLAP 场景，避免传统架构中大量数据交互造成的资源浪费和冲突。

数据库从OLTP发展到HTAP



资料来源：艾瑞咨询、国信证券经济研究所整理

(3) 按系统架构，可以分为集中式与分布式数据库

- **集中式数据库：**是一种仅在单个位置存储、定位和维护的数据库。这种类型的数据库是从该位置本身修改和管理的。因此，该位置主要是任何数据库系统或集中式计算机系统。通过互联网连接 (LAN、WAN 等) 访问集中位置。该集中式数据库主要供机构或组织使用。
- **分布式数据库：**由多个相互连接并分布在不同物理位置的数据库组成。因此，分布式数据库可以独立于其他物理位置管理存储在各种物理位置的数据，不同物理位置的数据库之间的通信是由计算机网络完成的。
- **集中式数据库VS分布式数据库：**集中数据库是企业的最初选择，可以利用位于系统中心的服务器统一管理所有的共享资源，并处理来自用户的请求。单机数据库积累了大量的实践经验，在强一致性、稳定性、迁移成本和运维管理方面都更胜一筹，而且各资源独立，应用隔离性好，数据安全性高。分布式数据库在灵活性和扩展性方面具有优势，一方面分布式给予了每个部门根据其应用程序的特定需求选择软硬件的自由，不必因为共享IT架构而做出妥协；另一方面分布式IT架构天生自带可扩展属性，能够根据业务规模实现无限弹性扩展。

集中式与分布式数据库优缺点对比

比较基础	集中式数据库	分布式数据库
定义	它是一个仅在一个位置存储、定位和维护的数据库	它是一个由多个数据库组成的数据库，这些数据库相互连接并分布在不同的物理位置
访问时间	多用户情况下的数据访问时间更多的是在一个集中式数据库中。	在分布式数据库中，多用户情况下的数据访问时间更少
数据管理	由于整个数据都存在于同一位置，因此该数据库的管理、修改和备份更容易	这个数据库的管理、修改和备份非常困难，因为它分布在不同的物理位置
视图	该数据库为用户提供了一个统一的、完整的视图	由于它分布在不同的位置，因此难以为用户提供统一的视图
数据一致性	该数据库与分布式数据库相比具有更高的数据一致性	该数据库可能有一些数据复制，因此数据一致性较差
特性	如果数据库发生故障，用户将无法访问数据库	在分布式数据库中，如果一个数据库出现故障，用户可以访问其他数据库
成本	集中式数据库的成本更低	这个数据库非常昂贵
维护	易于维护，因为所有数据和信息都可以在一个位置获得，因此易于访问和访问	由于数据和信息分布在不同的地方，因此很难维护。因此，需要检查数据冗余问题以及如何保持数据一致性
高效	集中式数据库效率较低，因为数据查找变得相当复杂，因为数据和信息存储在特定位置	分布式数据库比集中式数据库更有效，因为数据在多个位置拆分，这使得数据查找变得简单且耗时更少
响应速度	与分布式数据库相比，响应速度更快	与集中式数据库相比，响应速度较慢

资料来源：CSDN、国信证券经济研究所整理

(4) 按部署模式，可以分为本地数据库与云数据库

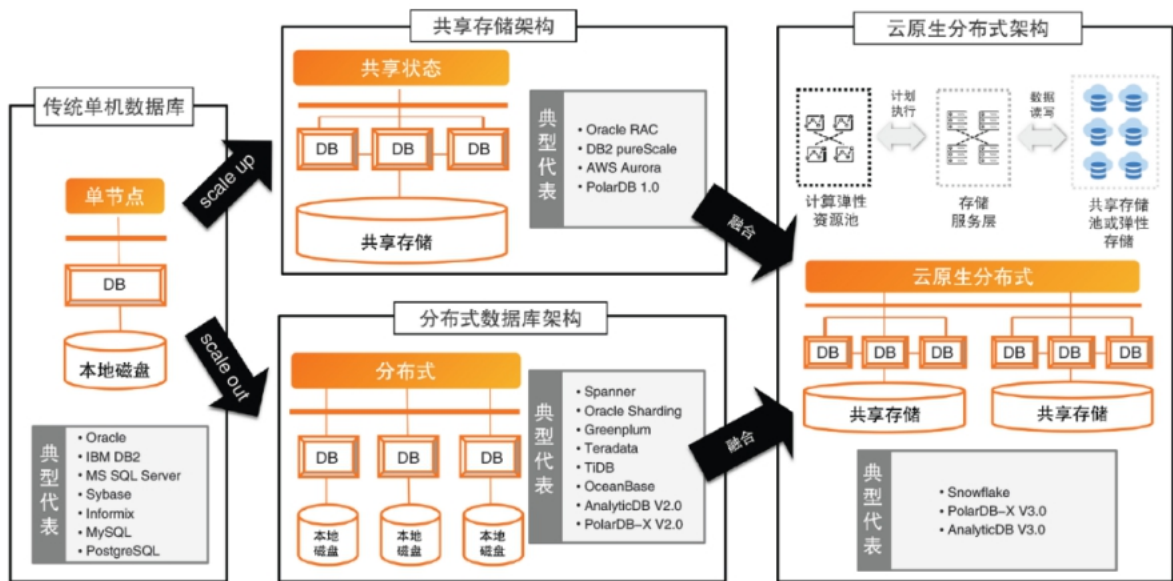
- **云数据库主要可以分为云厂商的传统数据库和云原生数据库。**云数据库是在云计算的大背景下发展起来的一种新兴的共享基础架构的方法，它极大地增强了数据库的存储能力，消除了人员、硬件、软件的重复配置，让软、硬件升级变得更加容易。现阶段云数据库主要包括两种：一种是托管在云厂商上的“传统”数据库，例如阿里云、腾讯云上的MySQL、PostgreSQL、MongoDB、Redis等；一种是基于云环境的云原生数据库，例如AWS的Aurora、阿里云的Lindorm和PolarDB等。
- **云数据库极大地利用了云计算“资源池化”的优势，在成本、可用性、易用性、扩展性和并行处理方面较传统数据库有绝对优势。**云数据库即开即用，用户可以根据自身的业务情况弹性开支、灵活调整；无需从头采购基础软硬件，无需考虑专业人员（DBA）部署，节省了人力物力；同时云数据库大多支持热备架构，可以实现故障秒级自动切换，备份、恢复更加灵活。但同时，由于云环境的特性、产品的不成熟性和市场的混合部署需求，云数据库在数据质量、数据迁移、数据融合、性能优化和规范标准方面仍有改进的空间。

云厂商托管的数据库与云原生数据库



资料来源：艾瑞咨询、国信证券经济研究所整理

从传统单机本地数据库到云原生分布式数据库



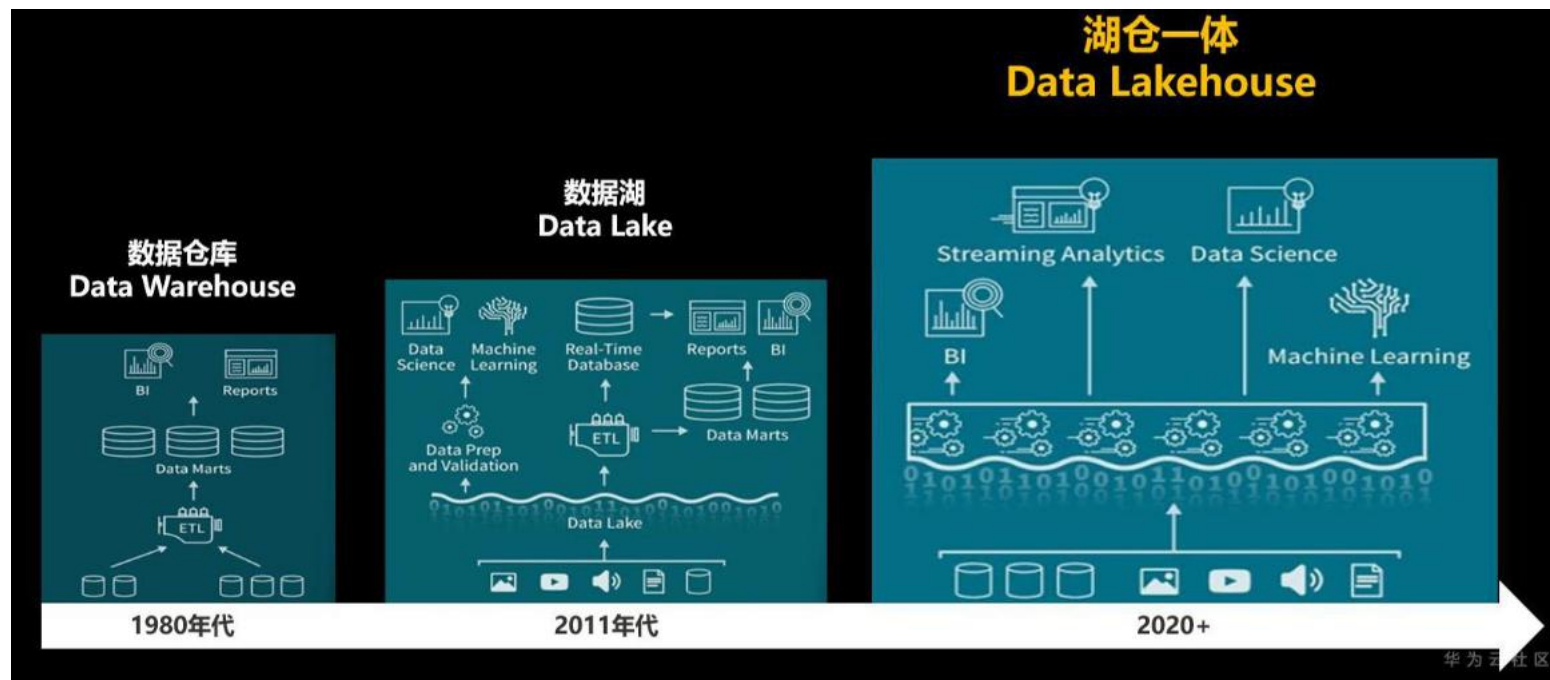
资料来源：阿里云、国信证券经济研究所整理

2、什么是数据平台

2.1 数据平台发展史：从BI到AI，数据时代带动底层架构持续迭代

- **数据时代带动数据平台底层架构从数据库、数据仓库、数据湖到湖仓一体。**数据平台主要是指数据分析平台，其分析内部和外部其它系统生成的各种原始数据，对这些数据进行各种分析挖掘以生成衍生数据，从而支持企业进行数据驱动的决策。整个数据平台的发展来看，就是随着企业信息化和数字化的逐渐推进，从数据库，数据仓库，数据湖到数据湖仓逐渐演进的。
- **大数据平台架构经历过三个阶段的技术演进。**大数据平台架构发展至今，已经经历了三个阶段的技术演进：随着数据量的不断增长，企业产生对数据分析的需求，从上世纪八九十年代的数据仓库（Data Warehouse）开始普及，对分析型场景提供非常好的支持，数仓无法支持半结构化以及非结构化数据，从而满足数据科学类或者机器学习类等高级分析场景，因此数据湖（Data Lake）概念于2011年被提出，但随着数据量的不断增长，数据湖缺乏数仓的数据管理特性，导致数据沼泽现象出现，因此在2020年湖仓一体（Data Lakehouse）被提出，旨在融合数仓与数据湖的优势。

从数据仓库、数据湖到湖仓一体



资料来源：华为云、Databricks、国信证券经济研究所整理

2.2 数据平台分类：从数据仓库、数据湖到湖仓一体



- **数据仓库（Data Warehouse）**：自上世纪80年代末出现以来，伴随MPP架构的出现，也使得系统能够处理更大规模的数据量。但是随着大量的非结构化、半结构化数据产生，数据仓库劣势显现出来；
- **数据湖（Data Lake）**：数据湖自2010年末出现，可以存储各类格式的原始数据存储库，但是企业对数据访问灵活性、高性能的需求从未降低，因此多系统共存是企业常态，如一个数据湖，加上多个数据仓库等，架构复杂性也为运维带来困难；
- **湖仓一体（Lakehouse）**：湖仓一体构自2020年出现，其建在数据湖低成本的数据存储架构之上，又继承了数据仓库的数据处理和管理功能。

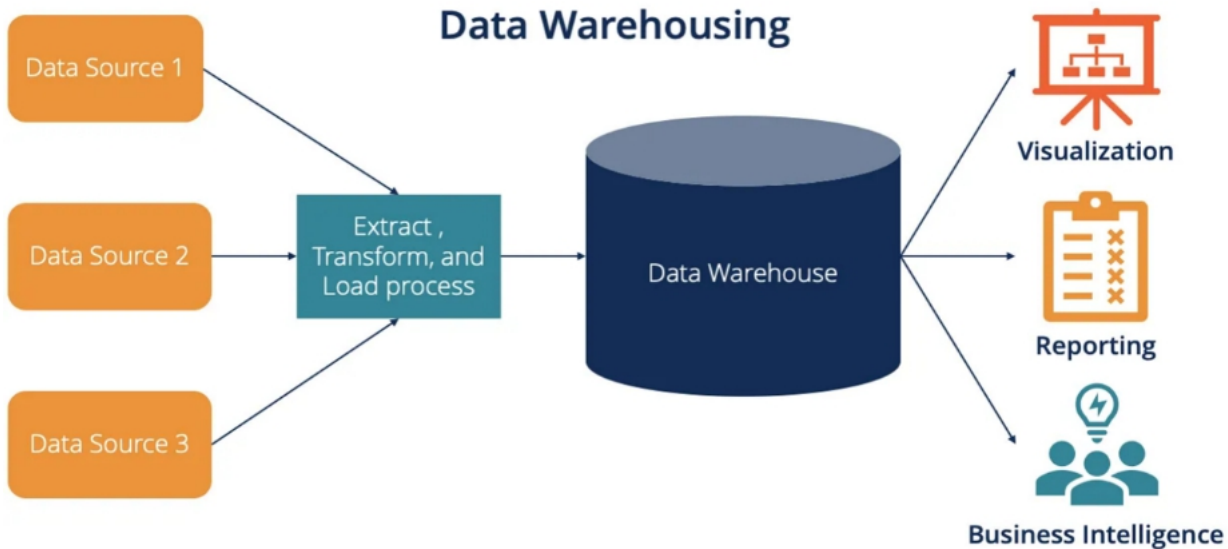
数据仓库、数据湖与湖仓一体的对比			
技术指标	数据仓库（Data Warehouse）	数据湖（Data Lake）	湖仓一体（Lakehouse）
成本	存储计算绑定，成本高	存储计算成本低，使用运维成本高	存算分离成本低，运维成本低
数据格式	封闭、专属	开放、通用	开放、通用
数据内容	数据类型单一，结构化为主	数据类型丰富	数据类型丰富
数据访问方式	SQL为主，少量支持API	开放API，直接读取数据SQL、R、Python等	开放API，直接读取数据SQL、R、Python等
数据可靠性	高质量、高可靠性、事务隔离好	数据质量一般，容易形成数据沼泽	高质量、高可靠性、事务隔离好
监管和安全	细粒度数据监管和权限控制	较差的数据监管和权限控制	细粒度数据监管和权限控制
性能	存储和计算深度优化，性能高	优化度不高，性能差	通过缓存、索引、分部等方式优化，性能高
扩展性	扩展性一般，扩展成本高	低成本、高扩展性	低成本、高扩展性
灵活度	预先建模，Schema-on-Write，灵活度低	无需提前建模，Schema-on-Road，灵活度高	无需提前建模，Schema-on-Road，灵活度高
使用场景	BI、SQL应用和报表等有限场景	数据科学、机器学习等丰富场景，性能对BI支持一般	数据科学、机器学习、BI等丰富场景

资料来源：CSDN、国信证券经济研究所整理

(1) 数据仓库 (Data Warehouse)：助力信息产业走向BI时代

- **数据仓库带动信息产业从运营式系统转向决策支持型系统。**随着数据库的大规模应用，为了研究数据之间的关系，人们越来越多的需要使用联机分析处理OLAP (On-Line Analytical Processing) 进行数据分析。但不同数据库间很难做到数据共享，数据之间的集成与分析也存在非常大的挑战。数据仓库之父比尔·恩门于1990年提出数据仓库 (Data Warehouse)。数据仓库主要功能是将OLTP经年累月所累积的大量数据，通过数据仓库特有的数据储存架构进行OLAP，最终帮助决策者能快速有效地从大量数据中，分析出有价值的信息，提供决策支持。自从数据仓库出现之后，信息产业就开始从以关系型数据库为基础的运营式系统慢慢向决策支持系统发展。
- **数据仓库主要应用于OLAP，支持商务智能和查询分析。**根据甲骨文对数据仓库的定义，数据仓库是一种面向商务智能 (BI) 活动 (尤其是分析) 的数据管理系统，它仅适用于查询和分析，通常涉及大量的历史数据。在实际应用中，数据仓库中的数据一般来自应用日志文件和事务应用等广泛来源。数据仓库系统的主要应用主要是OLAP，支持复杂的分析操作，侧重决策支持，并且提供直观易懂的查询结果。

数据仓库示意图



资料来源：AWS、国信证券经济研究所整理

数据库与数据仓库的对比

维度	数据仓库	数据库
数据来源	多数据源	单数据源
数据标准化	非标准化Schema	高度标准化的静态Schema
数据读取优势	针对读操作进行优化	针对写操作进行优化

资料来源：华为云、国信证券经济研究所整理

(2) 数据湖 (Data Lake)：面向AI时代，兼容各种数据格式

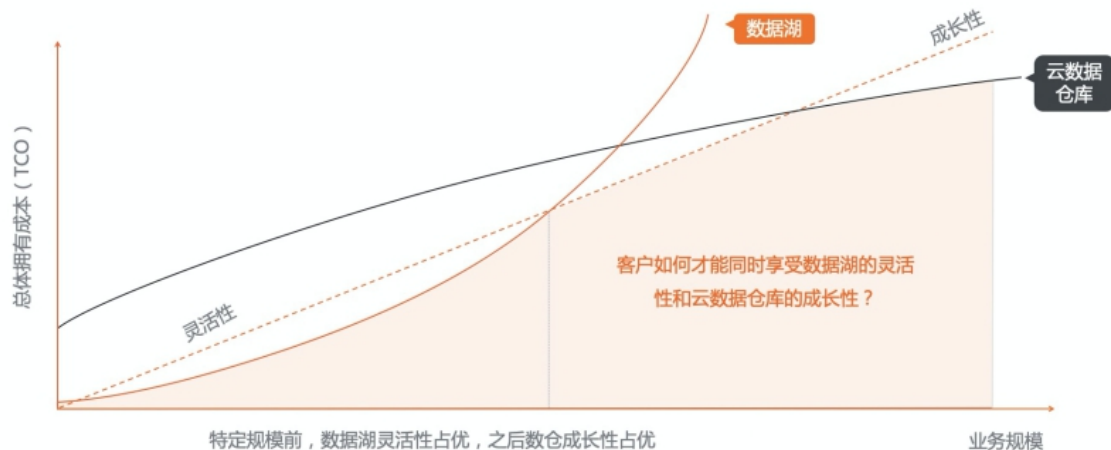
- **数据湖面向AI时代，可以存储结构化、半结构化和非结构化数据。**数据湖是大数据时代的产物，最早是由Pentaho的创始人James Dixon于2010年10月提出的，数据湖将数据以自然/原始格式存储在对象存储或文件系统中，数据湖通常把企业所有数据统一存储，包括源系统中的原始数据副本，传感器数据，社交媒体数据等，也包括转换后的数据，以用于支撑报表，可视化，高级数据分析和机器学习等。数据湖中的数据包括关系数据库的结构化数据(行与列)、半结构化的数据(CSV，日志，XML，JSON)，非结构化数据(电子邮件、文件、PDF)和二进制数据(图像、音频、视频)。
- **数据仓库VS数据湖：数仓主要支撑BI场景，数湖主要支撑AI场景。**数据仓库更关注数据使用效率、大规模下的企业级需求，数据由于是经过统一但开放的服务接口进入数据仓库，使其拥有更高的性能、闭环的安全体系、数据治理的能力等；数据湖则通过开放底层文件存储，给数据入湖带来了最大灵活性，数据类型没有过多要求。由于开放底层文件存储，也给上层引擎带来更多灵活度，各种引擎可根据自己针对的场景随意读写数据湖中存储的数据。但是数据湖虽然容易落地，但随着湖中数据量的增多，一旦数据治理措施不善，数据湖容易退化为数据沼泽，数据管理难度加大。

数据仓库与数据湖的对比

特性	数据湖	数据仓库
数据源	来自IoT设备、网站、移动应用程序、社交媒体和企业应用程序的非关系和关系	来自交易系统、运营数据库和业务线应用程序的关系
Schema	在分析时写入（读取型 Schema）	在DW实现之前设计（写入型Schema）
性价比	低成本存储获得较快的查询结果	较高成本的存储获得最快的查询结果
数据质量	任何可以或无法进行监管的数据（例如原始数据）	可作为重要事实依据的高度监管数据
用户	数据科学家、数据开发人员、业务分析师	业务分析师
分析	机器学习、预测分析、数据发现和描述	批处理报告、BI和可视化

资料来源：华为云、国信证券经济研究所整理

数据仓库与数据湖在灵活性和成长性的比较



资料来源：阿里云、国信证券经济研究所整理

(3) 湖仓一体 (Data Lakehouse)：旨在结合数仓与数据湖的优势

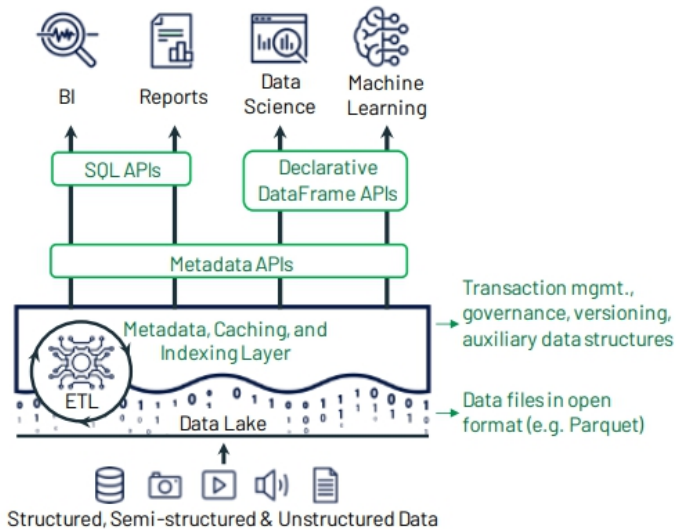
- **数据湖仓一体 (Data Lakehouse)** 目的是想结合数据仓库与数据湖的优势，同时满足BI与AI两类场景需求。数据湖仓在2020年被Databricks提出，由于数据仓库主要存储的是以关系型数据库组织起来的结构化数据，数仓中的数据存储结构与其定义的schema是强匹配的，而数据湖可以存储任何类型的数据，存储其中的数据不需要满足特定的schema。企业在搭建数据平台时，既有BI需求也有AI需求，因此，数据湖仓一体的目的是在想要结合数据仓库与数据湖的优势。数据湖仓是在数据湖的基本架构上，通过一系列以表格式Table Format 为代表的新技术解决了数据湖泊的各种传统痛点，将数据仓库和数据湖泊功能融合在一起，它构建在数据湖低成本的数据存储架构之上，又继承了数据仓库的数据处理和管理功能，并直接支持BI和AI的各种数据分析场景的新型架构。
- 湖仓一体架构将数据仓库的数据结构和管理功能与数据湖的低成本存储和灵活性相结合。湖仓一体通过提供单一通用的数据存储平台来满足所有业务数据需求来减少数据重复，也能消除了维护多个数据存储系统的成本和时间，但是由于其相对较新且不成熟，目前在成本和性能上还不足以与传统成熟的大数据存储解决方案竞争。

亚马逊智能湖仓一体示意图



资料来源：AWS、国信证券经济研究所整理

Databricks的数据湖仓一体示意图



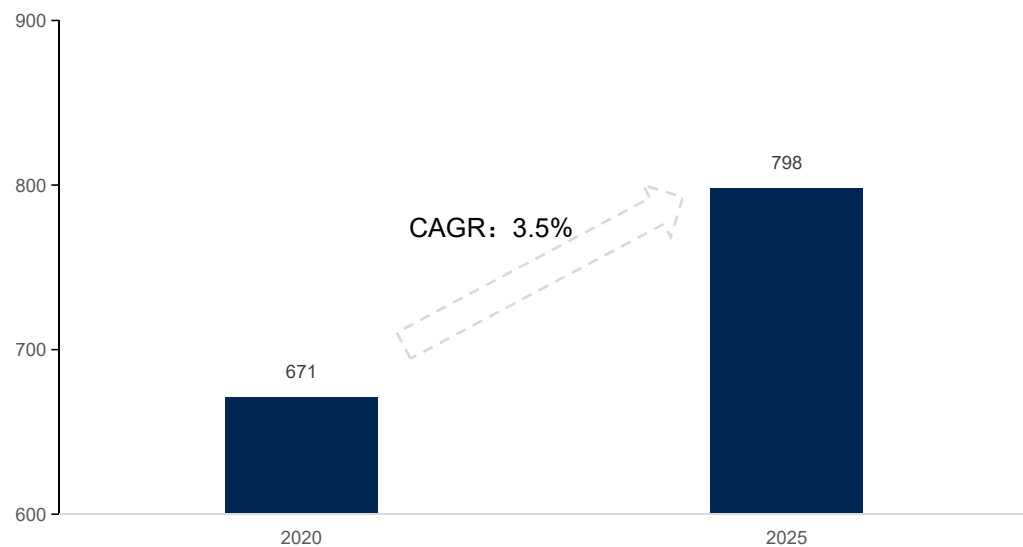
资料来源：Databricks、国信证券经济研究所整理

3、数据库的市场竞争格局

3.1 数据库市场空间广阔

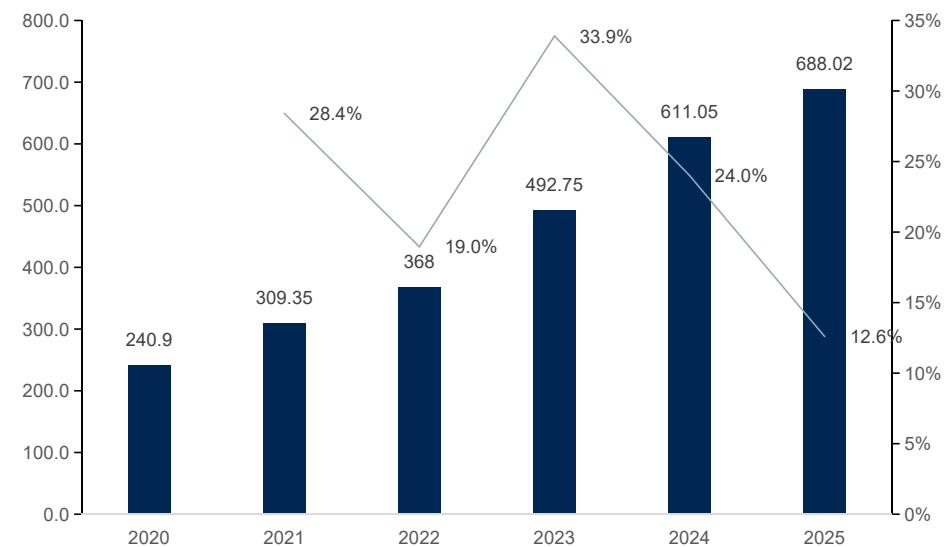
- **国内数据库市场增速较全球而言较快。**根据信通院的统计数据，全球数据库的市场规模在2020年为671亿美元，到2025年有望达到798亿美元，年均复合增速3.5%；而相比之下，国内的数据库市场相对较小，有望从2021年的309.35亿元增长到2025年的688.02亿美元，年均复合增速约23.4%，增速远高于全球数据库市场增速。
- **国内数据库市场占全球市场较小，未来发展空间较大。**目前根据信通院的统计数据，2020年国内的数据库规模为240.9亿元，占全球数据库市场规模的比例为5.2%，未来到2025年的时候，国内的数据库规模将占到全球数据库规模的12.5%（这一数据基本与中国IT总支出占全球IT总支出的比例较为接近）。

全球数据库市场规模及增速（单位：亿美元）



资料来源：信通院、国信证券经济研究所整理

国内数据库市场规模及增速（单位：亿元）

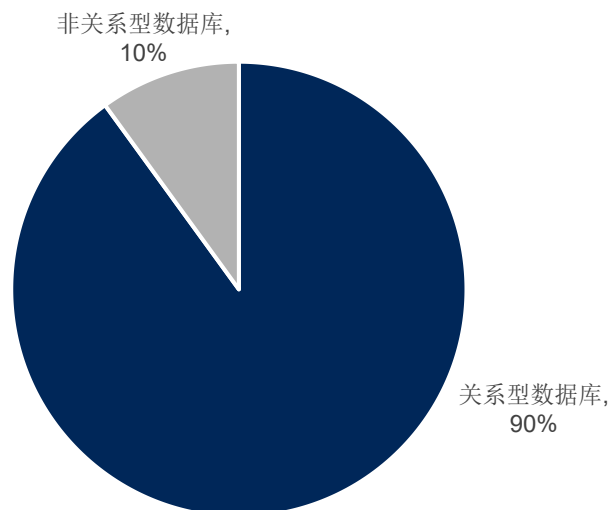


资料来源：信通院、国信证券经济研究所整理

3.1 数据库市场空间广阔

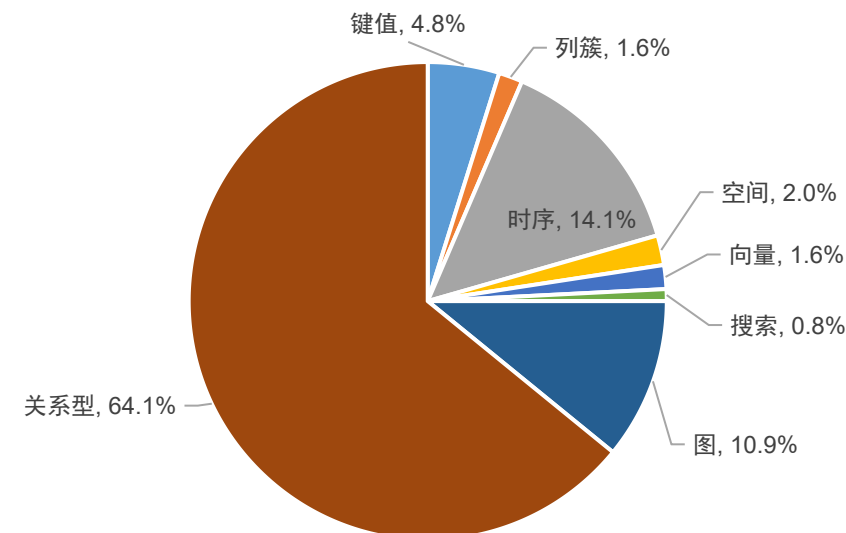
- **国内的市场还是以关系型数据库为主。**根据艾瑞咨询的统计数据，目前国内的数据库市场中，关系型数据库还是仍然占据主导地位，从市场占有率方面来看，关系型数据库市场占比达到90%左右，中国top的厂商都是以关系型数据库产品为主，主要原因是关系型数据库具有业务价值高、应用场景广的优势。非关系型数据库的市场占比为10%左右，主要是来自托管在公有云上的NoSQL开源产品。从具体的数据库产品来看（来自墨天轮统计），国内的数据库产品数量有248款，其中关系型数据库的产品占了159款，占比64.1%，其余的比如图数据库、时序数据库、键值数据库、列簇数据库、向量数据库、搜索数据库等也呈现百花齐放的状态。
- **全球市场中关系型数据库依旧占比较高。**目前根据IDC的统计，关系型数据库的份额优势明显，关系型数据库目前占据约80%的体量，其中全球关系型数据库中Oracle数据库占42%的市场份额。未来随着全球数据量的逐步增加、数据类型的逐渐丰富，除了关系型数据库之外的其他类型数据库将得到充分发展。

国内数据库市场中关系型数据库占主导



资料来源：艾瑞咨询、国信证券经济研究所整理

国内数据库市场中关系型数据库产品数量最多



资料来源：墨天轮、国信证券经济研究所整理

3.2 国内数据库市场的主要玩家

- **国内数据库市场呈现百花齐放态势。**目前国内的数据库市场主要分为国内厂商和海外厂商两大类，具体来说：1）国内数据库厂商，主要包括了传统数据库厂商、初创厂商、云厂商以及跨界厂商，传统数据库厂商成立时间较早，初创厂商则大多数是由中大型厂商的核心员工出来创办的居多，云厂商则利用自身云计算的优势加入行业，跨界厂商则主要依旧自身行业优势参与研发数据库；2）海外数据库厂商，主要分为开源和商业版，其中在商业数据库中，传统的Oracle、IBM等厂商依旧占据较大优势，新的云厂商借助自身云优势也获得较多市场份额，开源数据库中主要以非关系型数据库为主，如MongoDB、Redis、MySQL、PostgreSQL等发展势头良好。

中国数据库市场的主要玩家

国内数据库厂商

传统厂商



初创厂商



云厂商



跨界厂商



海外数据库厂商

商业数据库



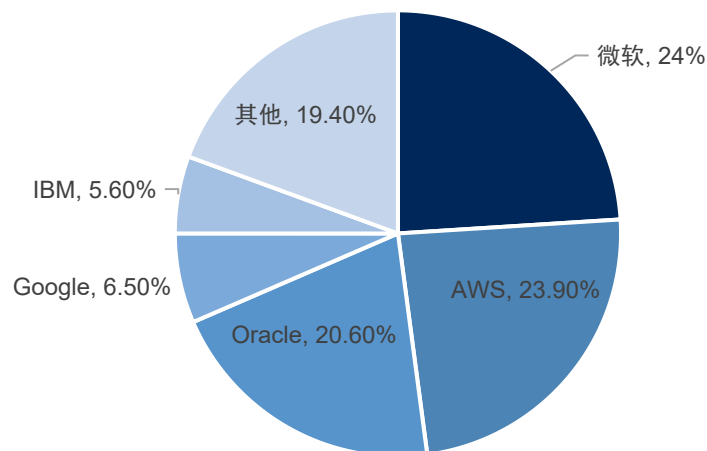
开源数据库



3.2 国内数据库市场的主要玩家

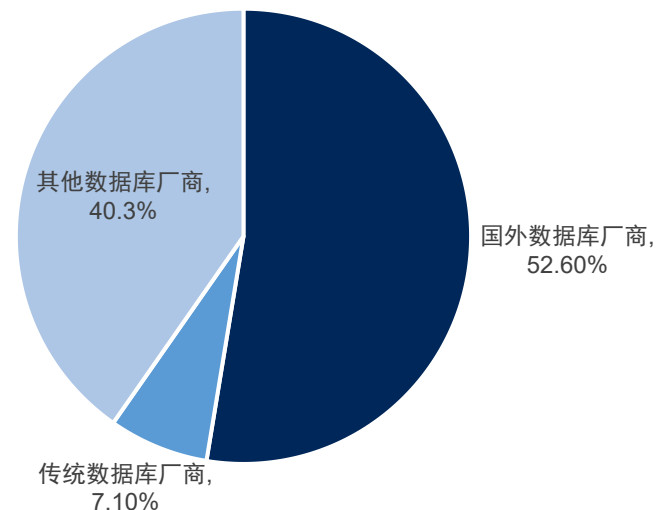
- **从市占率来看，国内数据库厂商的发展空间较大。**全球数据库的竞争格局中，由于国内市场占全球的比例过小（根据信通院数据，估算2020年国内数据库市场占全球5.2%）所以总体来说，微软的数据库目前占比第一，达到24%，AWS、Oracle、Google、IBM紧随其后，分别占比23.9%、20.6%、6.5%、5.6%，主要是因为海外的云化过程相比国内而言较早，所以随着云业务进一步扩展带来云化数据库服务也逐渐增长，最终使得微软、AWS等云厂商市场份额超过传统的Oracle。
- 就国内而言，根据2020年艾瑞咨询的统计数据，目前国内的数据库厂商主要包括达梦数据库、南大通用、神舟通用、人大金仓等传统数据库厂商，传统数据库厂商的市场份额是逐步上升的，国外的数据库厂商主要以Oracle、微软、AWS为主，其他的数据库厂商比如华为、阿里、PingCAP等都逐步在拓展各自的市场。

全球数据库市场竞争格局



资料来源：华经产业研究院、国信证券经济研究所整理

国内数据库市场竞争格局（2020年）

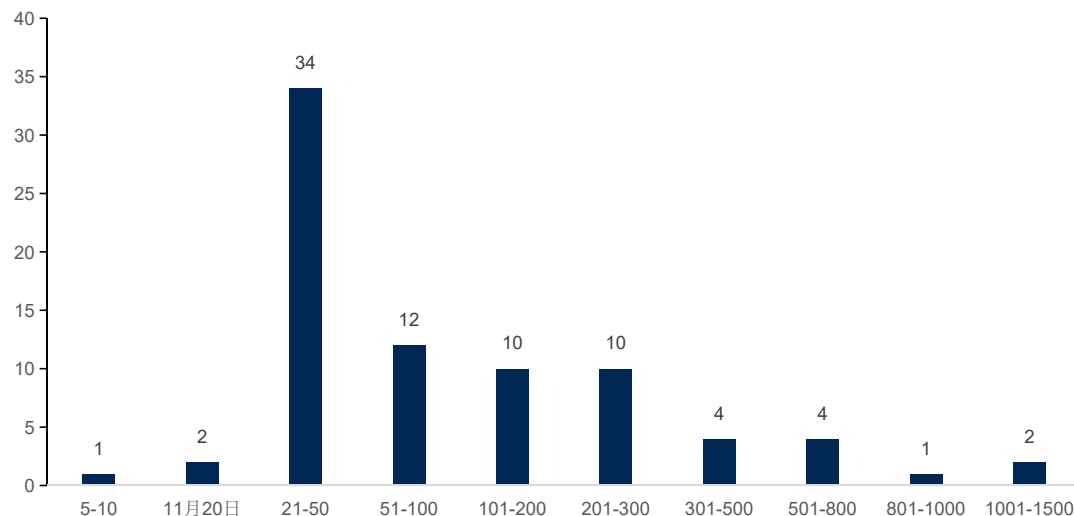


资料来源：艾瑞咨询、国信证券经济研究所整理

3.3 国内数据库市场有望走向集中

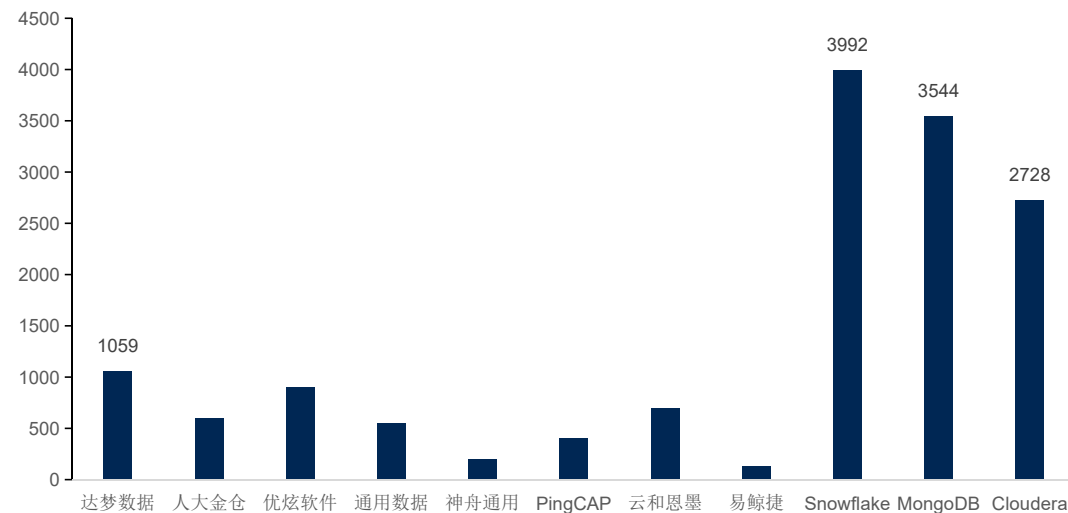
- **国产数据库厂商人员相比海外巨头较少。**目前国产数据库厂商总体人员数量较少，根据信通院的统计，目前我国对应的国产数据库厂商约有80家，其中大多数的人员都在100人以内，其中人员在21-50人规模的公司占比最高，人员在800人以上的仅有3家（信通院数据截止到21年6月）。相比之下，海外的数据库巨头人员数量较高，snowflake有3992人，MongoDB有3544人，Cloudera有2728人，Oracle有143000人，虽然这些云数据库公司的人员中营销人员占比较高，但总体的研发和技术支持人员也远高于国内数据库厂商。
- **国产市场目前处于跑马圈地状态，未来竞争格局有望走向集中。**未来国内数据库市场会逐步走向集中，观测的指标有：1）人员数量，目前国内处于快速替代期，产品的迭代、生态构建和营销推广都处于投入期，需要大量的人员；2）产品性能，在某些OLTP场景的替代还有一定的难度，技术的快速迭代并性能逐步接近Oracle等传统巨头是取得市场的关键；3）服务质量，目前国产替代的核心领域金融、电信、能源、政务、制造等，遇到问题时对数据库厂商及时提供技术支持能力非常看重；4）兼容性，包括了数据迁移工具的完备以及对于Oracle数据库和SQL语言的深度兼容，最大程度降低原有应用程序迁移到新数据库的难度和复杂度。

国内数据库公司人员统计分布



资料来源：信通院、国信证券经济研究所整理

部分国产数据库公司人员数量（单位：人）



资料来源：信通院、各公司官网、达梦招股说明书、国信证券经济研究所整理

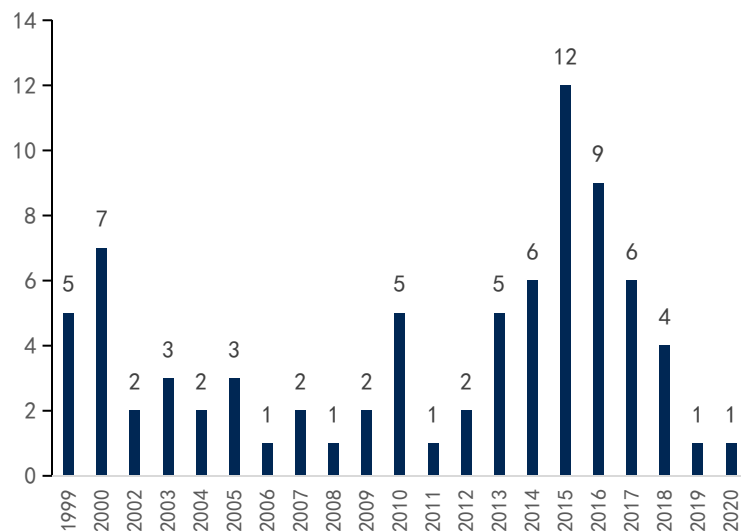
4、数据库的发展趋势

4.1趋势一：国产化

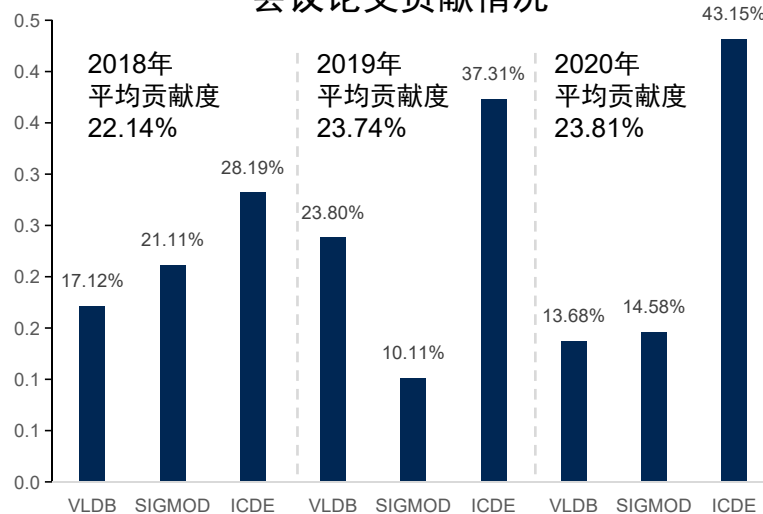
- **国产数据库的企业主要成立于2000年和2015年前后。**根据中国信通院的统计（截止到2021年6月），我国的数据库公司主要成立时间分别集中在2000年和2014-2017年左右，一方面是因为2000年前后是我国数据库发展的萌芽阶段，当时成立的传统数据库厂商有达梦、人大金仓、南大通用、神舟通用等，另一方面在2014-2017年之间，我国由于互联网时代的到来数据量激增以及去IOE趋势盛行，所以成立的数据库国产厂商较多。
- **经过十几年的发展，我国的数据库研发水平逐步提升。**VLDB、SIGMOD、ICDE三大会议是公认的数据库领域权威会议，根据信通院的统计来看，我国的数据库发展水平逐步提升，2018-2020年的三大会议的平均论文贡献度分别为22.14%、23.74%、23.81%。同时，从全球各国数据库领域的发论文数量和质量来看，美国和我国发文数量较高，分别达到1319、1141篇，领先其他国家，但是在质量层面还有待加强。

国内数据库的发展水平逐步提升

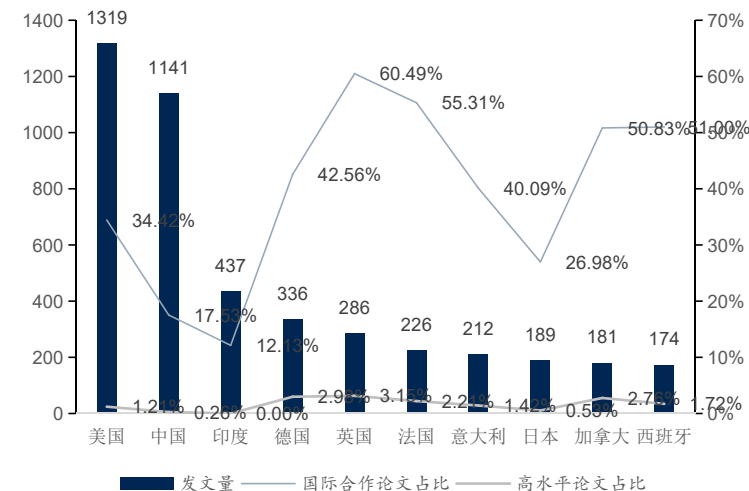
历年数据库公司成立数量



2018-2020 年中国高校及企业学术会议论文贡献情况



全球各国数据库领域发文量及质量



资料来源：中国信息通信研究院、国信证券经济研究所整理

4.1趋势一：国产化

- **国产化政策持续推动，数据库作为核心基础软件受重视。**从政策方面来看，近些年国家出台了众多的支持国产数据库的发展政策，尤其是在近几年随着国际形势的变化，国产化趋势愈发明显，其中在2020年发布的《关于扩大战略性新兴产业投资培育壮大新增长点增长极的指导意见》中提出要加大关键软件等核心技术的攻关，在2022年8月份国资委召开的中央企业关键核心技术攻关大会上，也强调要加大原创性技术投入，聚焦“卡脖子”问题取得更多突破性成果。
- **国内阿里先行，率先提出去IOE。**2009年，王坚带队在利用开源数据库MySQL和平民化的服务器来替换Oracle和小型机；2011年7月，淘宝“商品库”完成从Oracle到AliSQL的迁移；2012年，淘宝最后一台小型机宣布下线；2013年，Oracle数据库被完全清除出淘宝核心系统，所有数据库从Oracle迁移到AliSQL数据库；2013年，淘宝广告系统的Oracle数据库下线；2015年，OceanBase替换了支付宝支付系统中的Oracle数据库；2016年，支付宝总账全面用OceanBase替换了Oracle数据库。从上面的历程可以看出，阿里在国内的企业中是去IOE比较彻底的。

国内支持政策梳理

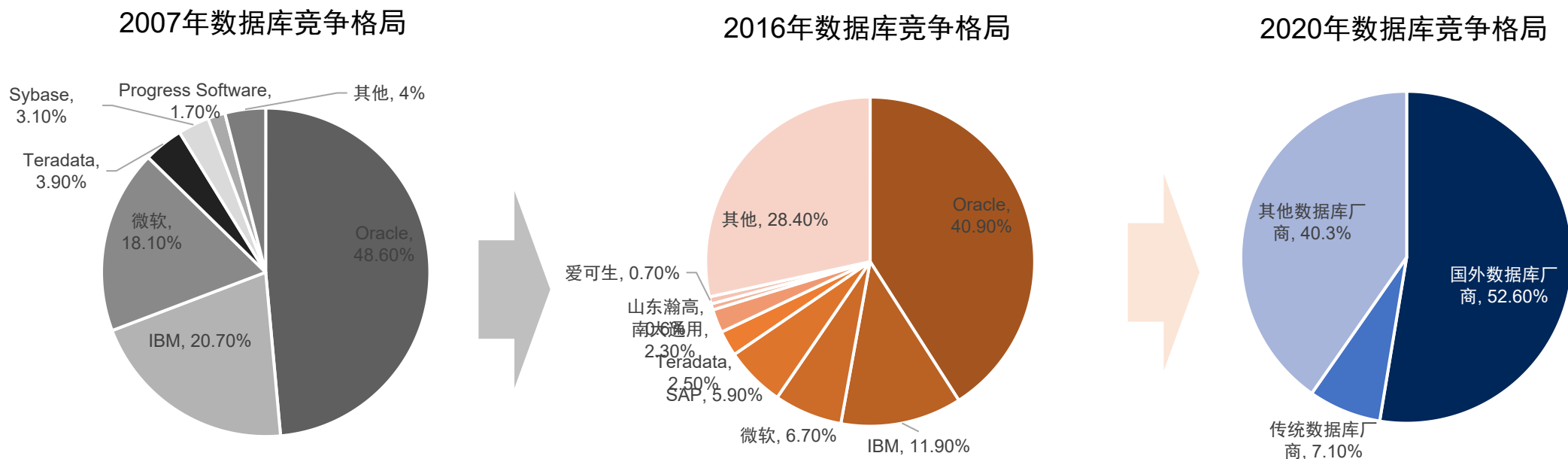
时间	事件	重点内容
2020年8月	《关于新时期促进集成电路产业和软件产业高质量发展若干政策的通知》	为进一步优化集成电路产业和软件产业发展环境，深化产业国际合作，提升产业创新能力和发展质量，加速国内科技产业建设，推动国产替代进程。
2020年9月	《关于扩大战略性新兴产业投资培育壮大新增长点增长极的指导意见》	加大5G建设投资，加快5G商用发展步伐，加快基础材料、关键芯片、高端元器件、新型显示器件、关键软件等核心技术攻关，大力推动重点工程和重大项目建设，积极扩大合理有效投资。
2021年7月	中共中央政治局召开会议	要强化科技创新和产业链供应链韧性，加强基础研究，推动应用研究，开展补链强链专项行动，加快解决“卡脖子”难题，发展专精特新中小企业。
2021年12月	《“十四五”大数据产业发展规划》	到2025年底，大数据产业测算规模突破3万亿元的增长目标，以及数据要素价值体系、现代化大数据产业体系建设等方面的新目标。
2022年1月	《“十四五”数字经济发展规划》	规划指出，要加快推动数字产业化，增强关键技术创新能力，提升核心产业竞争力。提升核心产业竞争力方面，要着力提升基础软硬件、核心电子元器件、关键基础材料和生产装备的供给水平，强化关键产品自给保障能力。
2022年3月	《“十四五”国家信息化规划》	规划指出，以开源生态构建为重点，打造高水平产业生态；以软件价值提升为抓手，推动数字产业能级跃升；以科技创新为核心，推动网信企业发展壮大。
2022年6月	《关于加强数字政府建设的指导意见》	提高自主可控水平。加强自主创新，加快数字政府建设领域关键核心技术攻关，强化安全可靠技术和产品应用，切实提高自主可控水平。
2022年8月	国资委召开中央企业关键核心技术攻关大会	加大原创技术投入，不断提升自主创新能力，聚焦“卡脖子”问题取得更多突破性成果。

资料来源：前瞻产业研究院、国信证券经济研究所整理

4.1趋势一：国产化

- **数据库的国产化趋势已经形成，未来替代空间较大。**我国数据库市场在早年间，基本上由海外厂商把持，根据《中国数据库追梦之路》的表述，2007年国内数据库的竞争格局中Oracle一家独大，市场占比达48.6%，紧随其后的是IBM、微软、Teradata、Sybase等，基本上整个市场均被海外厂商占据，很重要的原因是自中国21世纪加入WTO后各个行业需求井喷，而国外的数据库厂商率先进入国内市场，此时国内的数据库厂商还未做好准备。根据中国产业信息网、艾瑞咨询的数据显示，经过数十年的发展，国内的数据库厂商逐步在国产化的需求下逐步有了一定的市场份额，在2016年海外数据库厂商的市场份额下降到68%左右，到了2020年进一步下降到52.6%。虽然我国的数据库厂商在这些年中获得了较大的发展，但是从总体的市场占比上还是远远落后于海外数据库厂商。未来随着在党政以及教育、电力、金融、能源、电信、石油、交通、医疗、航天航空等领域的国产化需求不断释放，国产数据库厂商的市场份额有望进一步提升。

国内数据库市场竞争格局演化过程

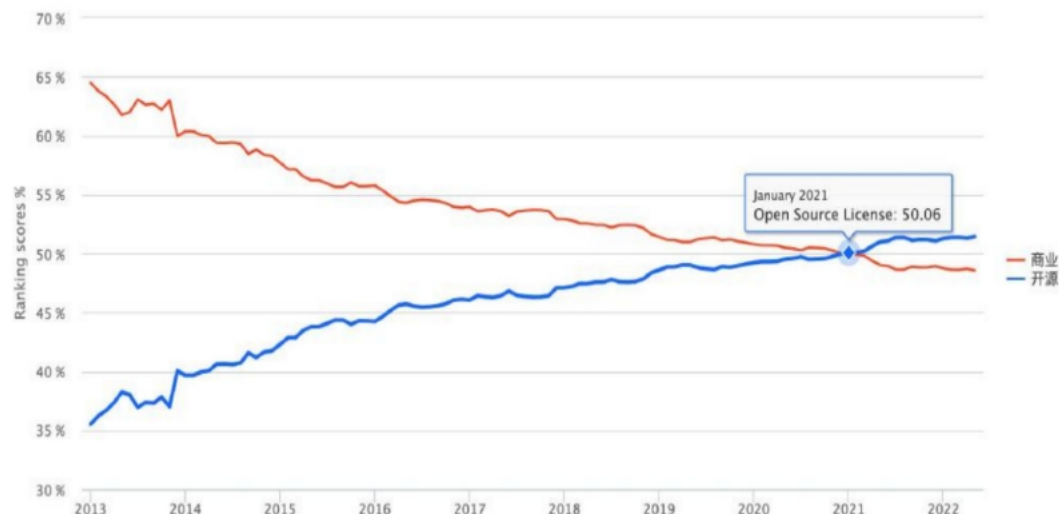


资料来源：《中国数据库追梦之路》、中国产业信息网、艾瑞咨询、国信证券经济研究所整理

4.2趋势二：开源和商业数据库并存

- **全球市场中开源数据库快速增长，国内排名靠前的数据库大多为开源数据库。**目前，在全球数据库市场中，开源数据库相比于商业数据库而言进展快速，根据DB-Engines的统计，在2021年的时候开源数据库的得分就已经超过了商业数据库的得分，呈现较强的发展趋势。就国内而言，根据墨天轮2022年11月的数据库排名，国产数据库中排名前十的大部分都是开源数据库，比如说PingCAP的TiDB（2015年开源）、蚂蚁金融的Oceanbase（2021年开源）、华为的openGauss（2019年开源），还有PolarDB、TDSQL等，大多数排名靠前的都是开源数据库，只有达梦、GaussDB、人大金仓等属于商业数据库。
- **国内开源数据库大多来源于MySQL和PostgreSQL。**目前国内的很多开源数据库都来自于海外的MySQL和PostgreSQL，比如基于MySQL的国内数据库有阿里云的PolarDB、腾讯云的TDSQL等，基于PostgreSQL的有华为的GaussDB等，总体来说，国内的开源数据库大多数是基于MySQL和PostgreSQL开发而来。MySQL遵循GPL协议，对于开发者来说使用了遵循GPL协议的数据库也要开源，而PostgreSQL遵循的是BSD协议，允许使用者修改和重新发布代码，也允许使用或在BSD代码上开发商业软件进行发布和销售，对开发者较为友好。

2021年开源数据库得分超过商业数据库



资料来源：DB-Engines、国信证券经济研究所整理

国内数据库排行榜（2022年11月）

排行	上月	半年前	名称	模型	属性	三方评测	生态	专利	论文	得分	上月	半年前
1	1	↑ 2	TiDB +	关系型	HP X B	BP	GO O C E	15	23	580.67	-31.78	-6.22
2	2	↑ 3	OceanBase +	关系型	HP X B	BP BP BP	GO O C E	137	17	574.56	-9.56	+11.45
3	3	↑↑ 5	达梦 +	关系型	TP X B	BP	GO O C E	381	0	497.01	-59.11	+34.23
4	4	↓↓↓ 1	openGauss +	关系型	TP X B	BP BP	GO O C E	562	65	486.17	-47.16	-102.85
5	↑ 6	↓ 4	GaussDB +	关系型	X B HP	BP BP	GO O C E	562	65	402.19	-33.20	-91.98
6	↓ 5	6	PolarDB +	关系型	X B HP	BP BP	GO O C E	512	26	388.53	-47.77	-10.83
7	7	↑ 8	人大金仓 +	关系型	TP X B	BP	GO O C E	232	0	387.93	-43.44	+52.21
8	↑ 9	↓ 7	GBase +	关系型	TP X B	BP BP BP	GO O C E	152	0	259.13	-16.04	-125.77
9	↓ 8	9	TDSQL +	关系型	X B HP	BP BP	GO O C E	39	10	251.47	-28.23	-17.68
10	10	10	AnalyticDB +	关系型	AP X B	BP BP BP	GO	480	28	172.02	-20.50	-22.03

资料来源：墨天轮、国信证券经济研究所整理

4.2趋势二：开源和商业数据库并存

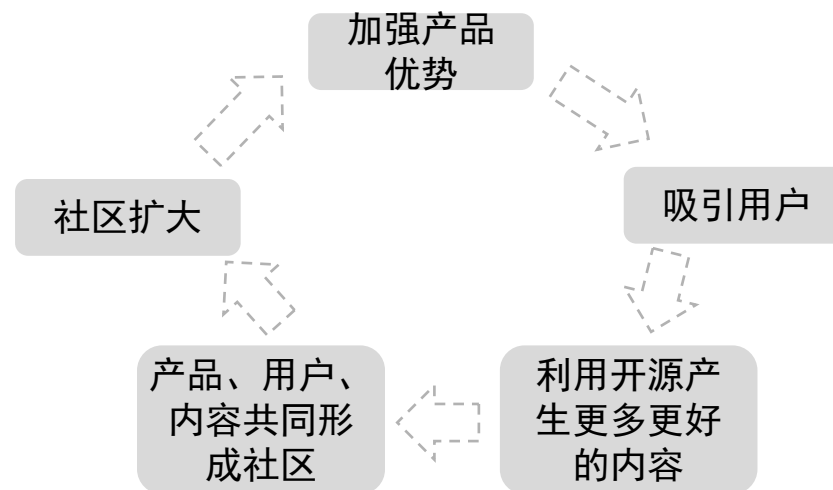
- **开源数据库具备开发成本低、迭代快的优势。**相比于商业数据库而言，开源数据库在开发成本、迭代速度等方面有明显的优势。由于开源数据库的代码是开源的，所以开发者在获得代码进行开发的成本较低，不用像商业数据库一样缴纳高昂的许可费用；同时，由于开源数据库社区凝聚了众多极客开发者，所以更新迭代的速度相对而言较快，尤其是对于新的功能或者应用的试验会较为迅速。商业数据库虽然代码不公开，且收取较高的许可费用，但是会为客户提供较为优质的服务，包括开发、运维以及技术支持等。
- **开源数据库利用生态优势获得较快的进步速度。**开源数据库凭借生态优势获得快速的发展，首先凭借领先的产品性能吸引用户，有了用户之后可以产生更多更好的内容，进而形成社区，随着社区的扩大的可以进一步加强产品的迭代速度和性能，从而形成正向循环。开源数据库和商业数据库之间并不矛盾，开源产品以最快的速度进行宣传和迭代，聚拢众多开发者，而商业产品则可以为客户提供比较完善的服务，因为大多数的客户都没有独立开发和修改数据库的能力。

开源数据库和商业数据库对比

指标	重点	例子	成本	社区	源代码	技术支持	许可证	支持	安装和更新
开源数据库	在开放源码的数据库中，任何人都可以轻松查看它的源代码	MySQL、PostgreSQL、MongoDB	免费的或成本是可承担的	社区一般可以看到分享和修改数据库的代码	源代码开放，故存在编码错误的风险	开源数据库技术支持有限，主要靠社区	许可证免费	开源数据库的支持主要靠社区	安装和更新由用户自己决定
商业数据库	商业数据库是只为商业目的而创建的。	Oracle、DB2、Splunk	一般需要付出高昂的费用购买	社区看不到也不能交流或者修改数据库的代码	未经授权无法接触到数据库源代码	由厂家提供完善的服务	有高昂的许可费用	可以额外从厂商那里购买服务	安装和更新由软件供应商管理

资料来源：CSDN、墨天轮、国信证券经济研究所整理

开源数据库发展的正向循环



资料来源：《中国数据库追梦之路》、国信证券经济研究所整理

4. 2趋势二：开源和商业数据库并存

- **国内主要的数据库产品已开源。**目前国内主流的产品都已经开源，比如PingCAP的TiDB产品于2015年开源、百度的Apache Doris于2017年开源、华为的openGauss产品于2019年开源、涛思数据的TDengine于2019年开源。一个开源数据库产品的进步速度取决于生态构建的丰富与否，主要的观测指标有开源时间、社区贡献人数、已解决问题数量等，比如TiDB数据库社区从2015年开源后已经解决了9043个问题，有741个贡献者，再比如华为的openGauss在2019年开源后已经解决了1056个问题，有165个贡献者。开源数据库的发展有较大程度依赖社区生态的完善。
- **基于国内开源数据库的国产数据库逐渐丰富。**目前国内开源生态逐渐丰富，不少的厂商都基于目前主流的国产开源数据库开发出了自己的数据库产品，比较明显的是的华为的openGauss生态，基于开源高斯数据库，人大金仓开发了AnnoDB，云和恩墨开发出了MogDB，海量数据开发出了Vastbase，神舟通用开发了神通数据库，虚谷伟业开发了有蓉数据库，东华软件开发了东华龙腾数据库，超图软件开发了Yukon数据库。

注：github开源代码托管平台为全英文且托管在美国；gitee开源代码托管平台为全中文且托管在国内。

国内数据库开源情况梳理									
名称	贡献者	开源地址	开源时间	已处理Issue	已合并Pull请求	Watch	Star	Fork	Contributors
TiDB	PingCAP	github.com/pingcap/tidb	2015年	9043	22110	1300	31200	5100	741
Apache Doris	百度	github.com/apache/incubator-doris	2017年	2733	5454	227	4300	1400	314
TDengine	涛思数据	github.com/taosdata/TDengine	2019年	2426	9417	679	18300	4300	181
Nebula Graph	悦数科技	github.com/vesoft-inc/nebula	2020年	1399	2439	170	7400	827	79
StarRocks	鼎石科技	github.com/StarRocks/starrocks	2021年	1376	3787	82	2500	529	104
云树Shard	曼可生	github.com/actiontech/dble	2016年	1109	1988	76	925	310	29
Apache HAWQ	偶数科技	github.com/apache/hawq	2015年	-	1505	95	648	320	76
openGauss	华为	gitee.com/opengauss/opengauss-server	2019年	1056	1167	277	817	655	165
OceanBase	OceanBase	github.com/oceanbase/oceanbase/	2021年	484	234	91	4300	963	126
RadonDB	青云	github.com/radondb/radon	2018年	294	444	86	1600	214	9
gStore	北京大学	github.com/pkumod/gStore	2014年	65	46	38	515	187	26
AliSQL	阿里巴巴	github.com/alibaba/AliSQL	2016年	61	1	497	4500	856	3
TDSQL	腾讯云	github.com/Tencent/TBase	2019年	58	17	74	1200	237	18
PolarDB	阿里云	github.com/ApsaraDB/PolarDB-for-PostgreSQL	2021年	45	100	56	2300	350	55
IvorySQL	瀚高	github.com/IvorySQL/IvorySQL	2021年	28	56	10	83	47	63
PolarDB	阿里云	github.com/ApsaraDB/galaxysql	2021年	23	13	30	1000	199	15
SequoiaDB	巨杉科技	github.com/SequoiaDB/SequoiaDB	2015年	3	68	22	272	111	-
GreatSQL	万里数据库	github.com/GreatSQL/GreatSQL	2021年	0	0	2	8	40	-

资料来源：墨天轮、国信证券经济研究所整理

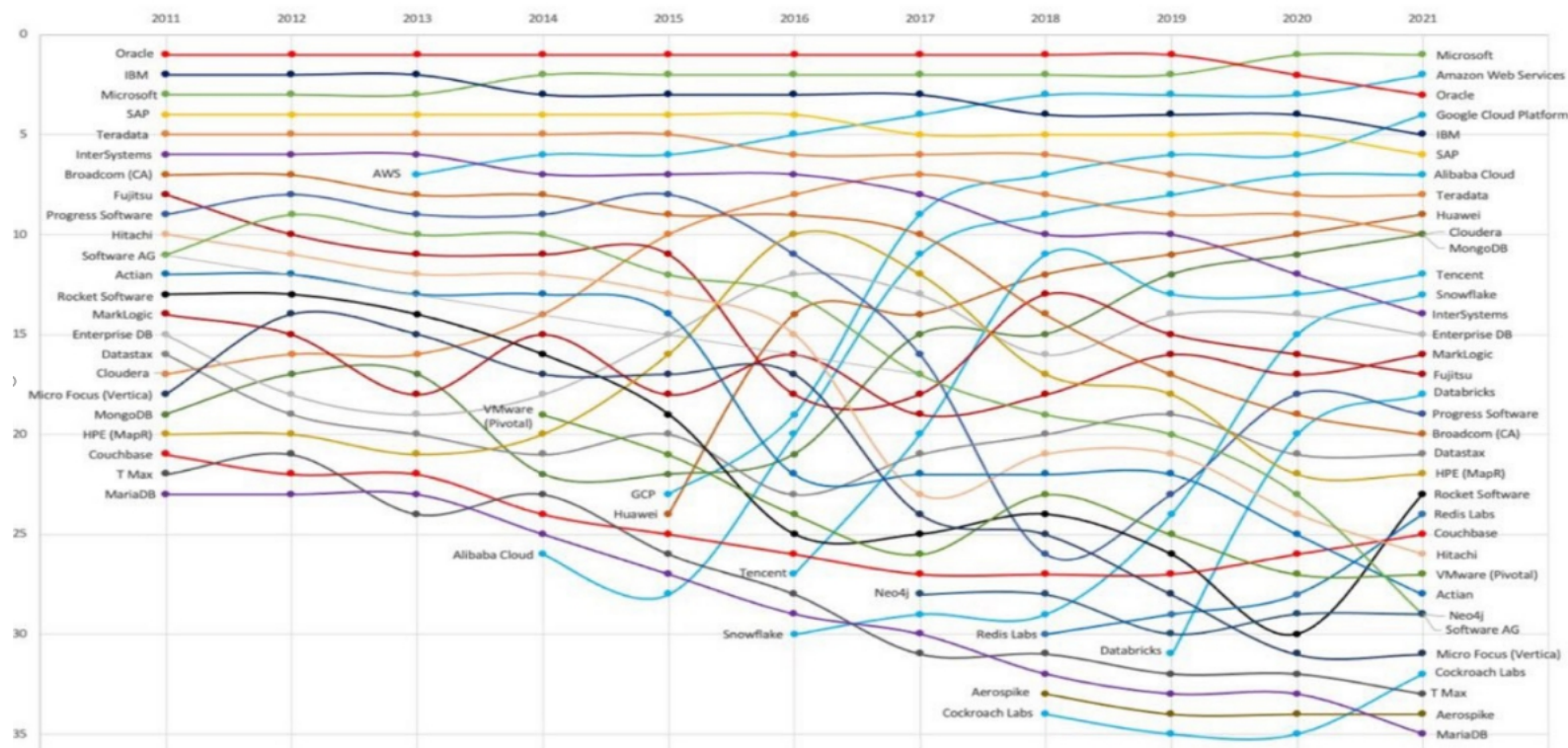
基于openGauss开源数据库的产品一览	
人大金仓，AnnoDB：	基于openGauss的企业级关系型数据库
云和恩墨，MogDB：	基于openGauss的完善增强的企业发行版数据库
海量数据，Vastbase：	基于openGauss开发的企业级关系型数据库
神舟通用，神通数据库：	以openGauss为基础开发的企业级关系型数据库
虚谷伟业，有蓉数据库：	基于openGauss内核、适配鲲鹏体系的数据库
东华软件，东华龙腾数据库：	以openGauss为底座的数据库
华为云，GaussDB：	基于openGauss形成的企业级分布式数据库
超图软件，Yukon：	基于openGauss的二三维一体化空间数据库

资料来源：各公司官网、墨天轮、国信证券经济研究所整理

4.3趋势三：数据库云化

- **云计算改变数据库市场格局，云数据库崛起。**根据Gartner对近十年的全球数据库市场的份额追踪，最为显著的变化是云计算厂商改变了数据库的市场格局。在2021年，全球数据库市场的前五名为Oracle、IBM、Microsoft、SAP和Teradata，而到了2021年全球市场的前五名为Microsoft、Amazon、Oracle、Google和IBM。Gartner每年都会做魔力象限，2013年数据库的领导者象限主要被传统巨头占据（Oracle、微软、IBM、SAP），到了2020年AWS、谷歌、微软等云厂商也进入领导者象限。微软凭借SQL Server以及Azure云数据库的快速成长实现了对Oracle的反超，这与海外市场云基础设施的发展有着密切联系。

2011-2021全球数据库市场份额变化

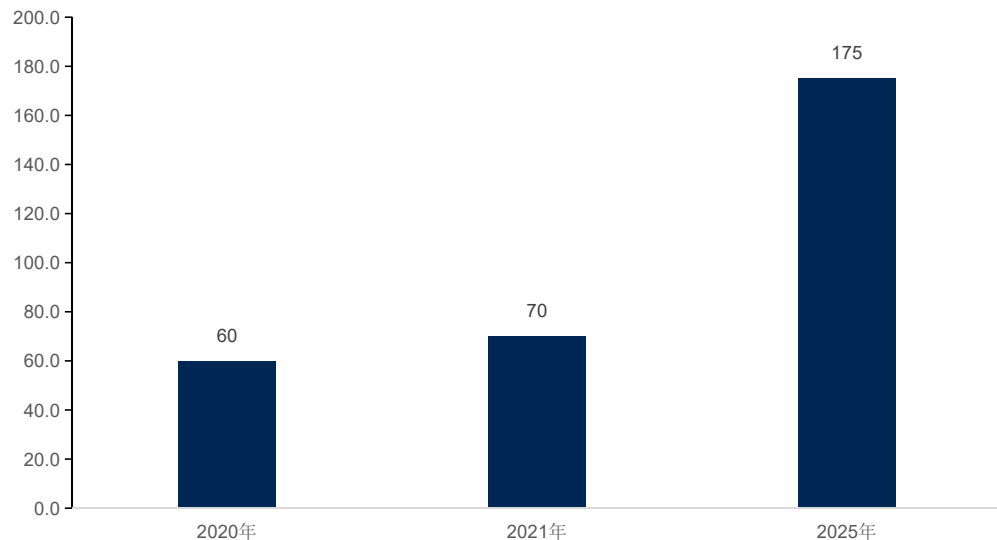


资料来源：Gartner、墨天伦、国信证券经济研究所整理

4.4趋势四：多模数据库

- **全球数据量快速增长，非结构化数据占比较高。**根据IDC的统计数据，目前全球的数据量呈现快速增长态势，2020年数据量达到60ZB，2021年达到70ZB，而到了2025年基本上会进一步上升到175ZB。在一项IDC面向企业的数据调查中，企业中的数据有80%是非结构化数据，所以对应的数据存储需求就越来越多样化，也延伸出了很多非关系型数据库的需求，比如键值数据库、宽表数据库、文档数据库、图数据库、内存数据库、时序数据库等。
- **多模数据库实现一库多用。**面对日益丰富的数据类型，一方面可以用多个专用数据库去进行管理，为此AWS为用户提供了数十种专用数据库，另一方面还可以在一个数据库里面实现多模型数据的存储和管理，即多模数据库，这方面诸如Azure Cosmos DB（微软）、ArangoDB（德国数据库公司）、SequoiaDB（巨杉数据库）、Lindorm（阿里云推出）等多模数据库表现较好。未来在云化的架构下，多模型数据管理是一种新的趋势，避免了应用程序面对多类型数据而频繁访问多个独立数据库的效率低下，国内的厂商的多模型数据库目前并不是很多。

全球数据量呈现快速增长态势（单位：ZB）



资料来源：IDC、国信证券经济研究所整理

主要的数据库存储类型

关系	键值	宽表	文档	内存	图	时序	分类帐
引用完整性，ACID 事务，写模式	高吞吐量，低延迟读写，无限扩展	以近乎无限的可扩展性存储大量数据	存储文档并快速访问和查询任何属性	以键查询，毫秒延迟	快速轻松的创建和找到数据之间的关系	按时间顺序收集、存储和处理数据	应用数据更改的完整的、不可变的、和可验证的历史记录
常见应用场景							
ERP, CRM, 财务	实时竞价、购物车、社交、产品目录、客户偏好	工业设备维护、车队管理和路线优化	内容管理、个性化、移动端	缓存、会话存储、排行榜、地理空间服务和实时分析	欺诈检测、社交网络、推荐引擎	物联网应用，事件跟踪	记录、供应链、医疗保健、注册和财务系统

资料来源：AWS、墨天轮、国信证券经济研究所整理

5、数据库相关公司梳理

5.1 拓尔思：以“语义智能+”为战略，布局搜索引擎数据库赛道

- 拓尔思是国内最早从事自然语言处理的企业之一。公司是国内最早从事自然语言处理（NLP）研发的企业之一，在NLP、知识图谱、OCR、图像视频结构化领域都具备自主可控的底层技术，处于行业领先地位。公司是最早从事中文全文检索技术研发的企业，具有全球领先的智能检索技术，现已将领先优势拓展到数据采集、治理、分析挖掘等大数据核心技术的全生命周期，所有底层技术均自主可控。
- 拓尔思主要面向内容安全和互联网空间治理、数字政府和数据智能三大板块。公司以“语义智能+”为发展战略，面向内容安全和互联网空间治理、数字政府和数据智能三大板块，覆盖从党政机关到企事业单位的众多行业市场。在内容安全和互联网空间治理板块，公司主要服务于媒体融合、内容审核、舆情分析以及版权维护等领域；在数字政府板块，主要服务于政府网站门户、政府大数据应用、专利大数据平台和服务、智慧公检法等领域；在数据智能板块，主要服务于金融智能风控和风险预警、企业大数据营销和智能客服、防务开源情报、科教知识服务等细分领域。

拓尔思打造的政府统一信息资源库平台



资料来源：拓尔思、国信证券经济研究所整理

拓尔思的AI+大数据产品生态体系



资料来源：拓尔思、国信证券经济研究所整理

5.1 拓尔思：以“语义智能+”为战略，布局搜索引擎数据库赛道



- **公司自研的TRS Hybase海贝大数据管理系统有望替代海外开源数据库搜索技术。**TRS Hybase海贝大数据管理系统是拓尔思自主研发的大数据管理软件，也是公司在大数据存储管理领域的主打产品，为大数据应用提供高效的数据存储、全文检索、分类统计等数据管理服务，支持几乎所有类型的数据，包括文本、数字、地理空间、图片等各种结构化、半结构化和非结构化数据。在信创适配方面，TRS海贝已完成与飞腾、龙芯、兆芯、海光、鲲鹏等CPU，麒麟软件、统信UOS、中科方德操作系统等主流信创厂商相关产品的全面兼容适配，并在政府、金融、安全等多个行业率先进行了实践应用。
- **公司是稀缺的国产全自研搜索引擎数据库标的。**根据IDC测算，当前全球搜索引擎数据库的市场规模已超过450亿美元，其中搜索系统、内容分析和认知/AI软件平台的市场规模约80亿美元。Elastic Search自发布以来迅速占领了全文搜索引擎市场，目前国内外很多企业已放弃自主研发，逐渐转投ES阵营。国内厂商大多数都以开源的ES作为搜索引擎数据库，小部分厂商在开源Lucene引擎的基础上进行封装，而海贝是目前国内少有的从底层分词算法到全文搜索引擎，以及上层系统都完全自研的纯国产搜索引擎数据库。

海贝与海外ES、Solr对比

	海贝	ES	SolrCloud
分词器	单一分词器处理全语种	多语种支持（分别处理）	多语种支持（分别处理）
查询语法	Lucene 语法/TRS 语法	JSON	Lucene 语法
多引擎	Lucene 引擎（Java） TRS 引擎(C)	Lucene（Java）	Lucene（Java）
易用性	全中文可视化数据管理后台	Kibana（英文）	较弱（英文）
冷热分区	支持	不支持	不支持
读写分离	支持	不支持	不支持
访问隔离	支持	不支持	不支持
索引加密	支持	不支持	不支持
安全控制	支持	不支持	不支持
国密算法	支持	不支持	不支持

资料来源：拓尔思、国信证券经济研究所整理

TRS海贝大数据管理系统典型应用案例



中国知识产权大数据与智慧服务系统

知识产权出版社为了给客户提供更好的服务，进行检索引擎选型，采用TRS Hybase建设中国知识产权大数据与智慧服务系统，针对1.3亿件全球专利，使用7台高配置PC服务器成功搭建。

了解更多

信用中国

“信用中国”网站提供全国企业（一亿多家）信用信息、信用代码、以及相关站内文章的公开查询服务。2018年开始迁移到海贝大数据管理平台，仅用6台服务器就满足了5000+的并发访问。

了解更多



资料来源：拓尔思、国信证券经济研究所整理

5.2 星环科技：是国内领先的企业级大数据基础软件开发商

- **星环科技是国内领先的企业级大数据基础软件开发商。**星环科技是国内领先的企业级大数据基础软件开发商，围绕数据的集成、存储、治理、建模、分析、挖掘和流通等数据全生命周期提供基础软件及服务，已形成大数据与云基础平台、分布式关系型数据库、数据开发与智能分析工具的软件产品矩阵，支撑客户及合作伙伴开发数据应用系统和业务应用系统，助力客户实现数字化转型。2022年6月，公司多个产品或子产品入选Gartner发布的《中国数据库管理系统供应商识别指南》，在识别的8类数据库管理系统产品中，公司入选产品覆盖其中7类，是覆盖超过7类或以上产品的四家厂商之一，以及覆盖多模数据库的四家厂商之一。
- **公司以大数据基础软件为核心，赋能客户数字化转型升级。**数据作为数字经济时代的生产要素，数字化基础设施主要可以分为基础软件、数据应用系统以及业务应用系统三层。星环科技主要为客户提供数字化基础设施底层、中间层的基础软件和技术服务，支持客户的技术团队及合作伙伴构建数据和业务应用系统，助力客户进行数字化转型。

数据库架构示意图



资料来源：星环科技、国信证券经济研究所整理

如何使用SQL语言查询数据库



资料来源：星环科技、国信证券经济研究所整理

5.2 星环科技：是国内领先的企业级大数据基础软件开发商



- **大数据基础平台（TDH）是公司自主研发的一站式大数据基础平台**，包括多个大数据存储与分析产品，能够存储PB级别的海量数据，可以处理包括关系表、文本、时空地理、图数据、文档、时序、图像等在内的多种数据格式，提供高性能的查询搜索、实时分析、统计分析、预测性分析等数据分析功能。TDH是通用的高性能大数据平台，提供标准的SQL开发接口，有着优秀的数据库兼容性，不仅可以帮助各个行业用户开发创新的数字化业务，还可以替代关系型数据库提升当前业务的应用效能。
- **以大数据基础平台为切入点, 打造数据全生命周期管理工具**。公司2013年发布大数据基础平台软件TDH 2.0，2014年发布数据智能分析工具Sophon早期版本Discover 3.0，2017年发布大数据开发工具TDS早期版本Transwarp Studio 5.0以及升级后的智能分析工具Sophon 1.0，2018年发布数据云平台TDC 1.0以及分布式分析型数据库ArgoDB 1.0，2019年发布分布式交易型数据库KunDB 1.0以及升级后的大数据开发工具TDS。公司围绕数据集成、存储、治理、建模、分析、挖掘和流通等数据全生命周期管理的各个阶段，研发了包括大数据与云基础平台、分布式关系型数据库、数据开发与智能分析工具等软件产品、软硬一体机产品及相关技术服务，实现“一站式”数据管理解决方案。

星环科技的TDH产品架构示意图



资料来源：星环科技、国信证券经济研究所整理

星环科技产品体系发展历程

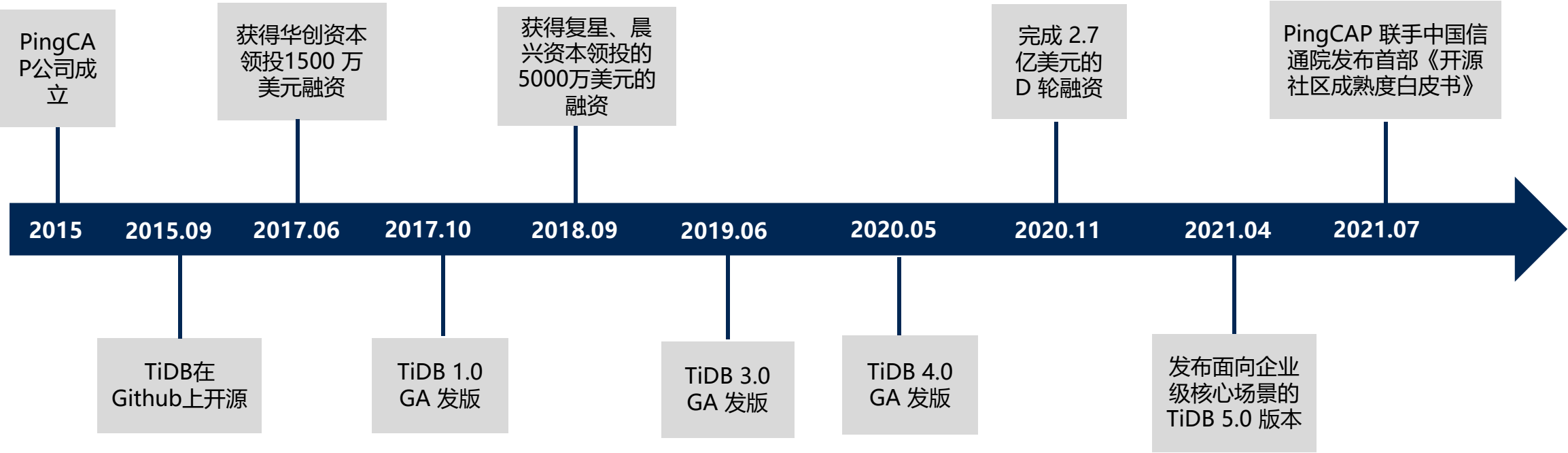


资料来源：星环科技、国信证券经济研究所整理

5.3 PingCAP：专注于企业级开源分布式数据库的公司

➤ **PingCAP是一家企业级开源分布式数据库厂商。** PingCAP成立于2015年，是一家企业级开源分布式数据库制造商。它提供开源分布式数据库产品、解决方案和咨询、技术支持和培训认证服务。公司致力于为全球行业用户提供稳定、高效、安全、可靠、开放、兼容的新数据服务平台，解放企业生产力，加快企业数字化转型升级。PingCAP以开源社区为依托，正在打造不分国界的全球产业生态，并将在企业数字化转型中提供更强大的助推力。公司创始人刘奇先后创建了Codis、TiDB、TiKV等知名开源项目，曾任豌豆荚、京东资深系统架构师，同时也是Go语言专家和Redis专家，擅长高并发、大规模、分布式数据库系统架构设计。2015年，刘奇与黄东旭和崔秋共同创立了PingCAP。

PingCAP发展历程



资料来源：墨天轮、国信证券经济研究所整理

5.3 PingCAP：专注于企业级开源分布式数据库的公司

- **PingCAP研发了多款TiDB相关产品。**目前TiDB的相关产品主要包括TiKV（键值数据库）、Zetta（开源NoSQL数据库）、Tidris（分布式NoSQL数据库）、Titan（基于TiKV的Redis兼容层的分布式实现）、JuiceFS（基于TiKV和S3的开源POSIX文件系统），丰富的产品线进一步拓宽了PingCAP的业务半径。
- **PingCAP最新一轮估值近30亿美金。**水电费公司从2015年成立就开始了融资之旅，从天使轮开始，基本上每年融资一次，到2021年7月的战略投资，共经历了6次融资，投资机构中不乏明星投资机构，最新一轮的估值已经达到30亿美金。目前PingCAP已经有全球员工400余人（2021年4月）。

TiDB相关产品

TiKV	是一个高度可扩展、低延迟且易于使用的键值数据库，广泛应用于在线服务服务。
TiDB	一个开源MySQL兼容NewSQL数据库，支持混合事务和分析处理(HTAP)工作负载
Zetta	一个开源的NoSQL数据库，支持Transaction和CloudSpanner之类的API
Tidris	一个分布式NoSQL数据库，提供Redis协议API(字符串、列表、哈希、集合、排序集合)，用Go编写
Titan	基于TiKV的Redis兼容层的分布式实现
JuiceFS	基于TiKV和S3的开源POSIX文件系统

资料来源：墨天轮、国信证券经济研究所整理

PingCAP融资历程

日期	融资轮次	估值金额	融资金额	投资机构
2021/7/20	战略投资	30亿美元		领投机构：红杉资本 跟投机构：GIC、贝塔斯曼亚洲投资基金、纪源资本、五源资本
2020/11/17	D轮		2.7亿美元	Coatue、云启资本、贝塔斯曼亚洲投资基金、时代资本、经纬中国、挚信资本、纪源资本、五源资本、昆仑资本、天际资本、晨曦投资、Access Technology Ventures
2018/9/12	C轮		5000万美元	云启资本、华创资本、五源资本、经纬中国、复星创富
2017/6/13	B轮		1500万美元	峰端资本、云启资本、经纬中国、险峰常青、华创资本
2016/9/21	A轮			峰端资本、云启资本、初心资本、经纬中国
2015/12/9	天使轮		数百万人民币	经纬中国

资料来源：墨天轮、国信证券经济研究所整理

5.4 华为GaussDB：基于openGAUSS企业级分布式关系型数据库

- **产品简介：**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，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基于统一架构，支持关系型与非关系型数据库引擎，满足政企全场景的数据智能管理需求，开启数据库极速与融合时代，加速政企智能升级。
- **核心业务：**GaussDB是一个产品系列，在整体架构设计上，底层是分布式存储，中间是每个DB特有的数据结构，最外层则是各个生态的接口，体现了多模的设计理念。具体产品包括：关系型数据库【GaussDB (for openGauss)、GaussDB (for MySQL)】、非关系型数据库【GaussDB (for Mongo)、GaussDB (for Influx)、GaussDB (for Cassandra)、GaussDB (for Redis)】、数据库工具及中间件（数据迁移DRS、应用结构转换UGO、分布式数据库中间件DDM）。

华为高斯数据库全景图

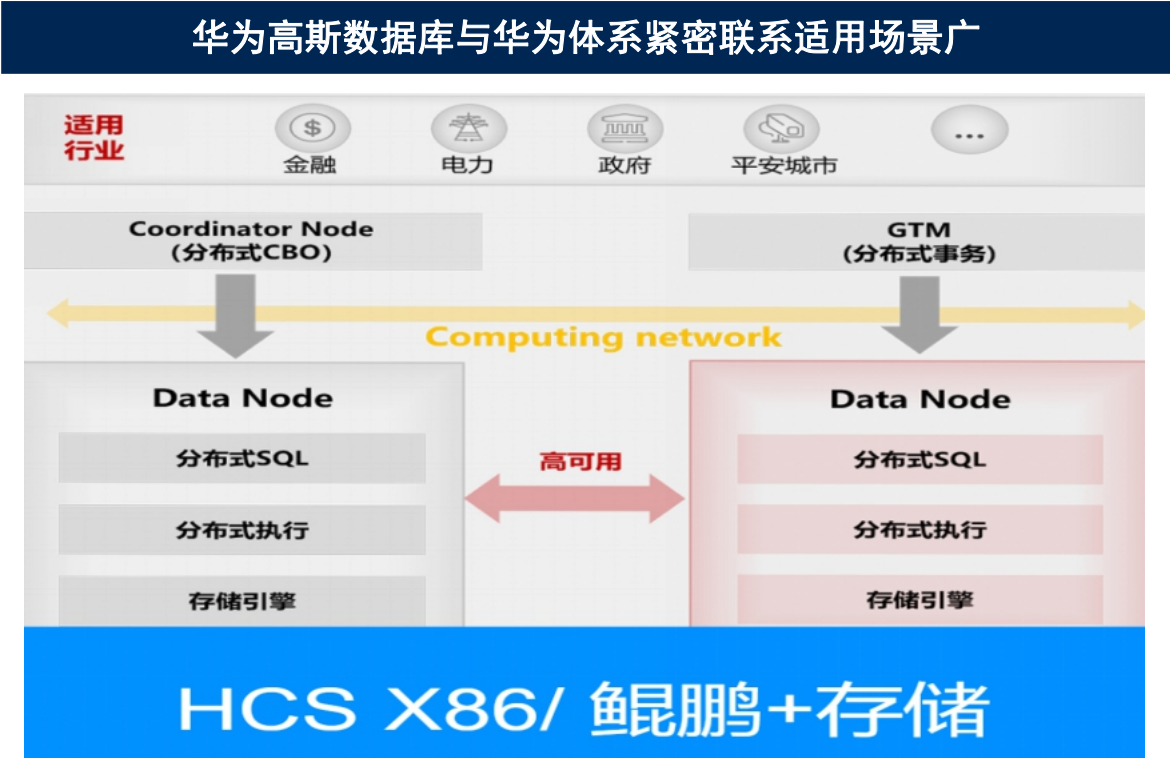
GaussDB全景图



5.4 华为GaussDB：基于openGAUSS企业级分布式关系型数据库



- **华为高斯数据库与华为体系紧密联系，适用场景广。** 华为GaussDB 在核心技术上跟传统数据库相比有巨大优势，可以解决很多行业用户的数据处理性能问题，为超大规模数据管理提供高性价比的通用计算平台，并可用于支撑各类数据仓库系统、BI（Business Intelligence）系统和决策支持系统，统一为上层应用的决策分析等服务。尤其是在华为的产品和业务体系内，从芯片到操作系统再到数据库，会比其他单纯数据库的厂商在解决方案上更加具有优势。目前高斯数据库在金融、政务、电力等重大领域应用较广。
- **华为高斯产品具有企业级复杂事务混合负载能力，能为企业提供功能全面，稳定可靠，扩展性强，性能优越的企业级数据库服务**



资料来源：墨天轮、国信证券经济研究所整理

华为高斯数据库的高性能

高性能—分布式并行执行框架

分布式执行框架会根据业务SQL生成最优执行计划，通过算子下推、并行执行等技术，提升分布式执行效率。算子下推：完美sharding的单节点执行、点查等不需要DN间数据交互的场景下，CN将SQL直接下发至DN执行；跨节点分布式执行、关联查询等需要DN间数据交互的场景下，CN将执行计划下发给DN，DN间通过Streaming算子完成数据聚合。并行执行：节点间支持DDL、DML等SQL语句并行执行；节点内支持基于数据页的并行查询。

高性能—分布式事务处理性能，GTM-Lite技术

GTM-Lite技术，在保证事务全局强一致的同时，提供高性能的事务处理能力，避免了单GTM的性能瓶颈。事务管理节点GTM，通过无锁原子操作提供CSN序号，不存全局单点瓶颈。全局数据结构NUMA化改造。通过工作进程NUMA绑核、全局数据结构NUMA分区化改造，减少跨核、跨处理器竞争冲突。

高可用—并行回放实现极致RTO，多层次冗余无单点故障以及跨AZ/Region容灾技术

并行回放实现极致RTO。通过日志流水线、批量回放和Block级物理并行恢复等关键技术，日志回放效率大幅提升，实现大压力下RTO<10s的极致高可靠。同城跨AZ双活：实现同城部署，实现应用层的就近数据访问、两地三中心：通过AB数据中心同城双活，使用跨Region的容灾能力、异地多活：支持多中心统一查询及全局一致读，整体资源利用率高。

高扩展—Scale-out在线横向扩展

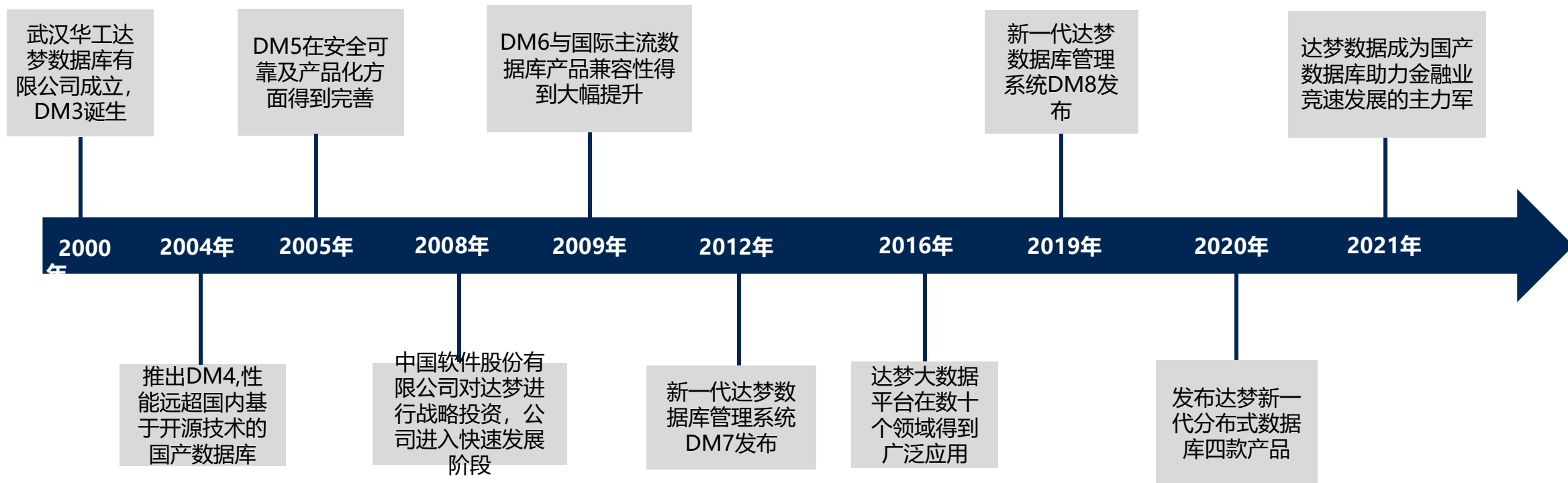
在线扩容。通过hashbucket聚集存储和多轮追增等技术，减少锁表时间，实现应用无感知的在线扩容。通过增加服务器，实现系容量和性能的线性Scale Out。支持1000+超大分布式集群的能力，可以轻松应对海量高并发数据处理和复杂查询场景的考验。

资料来源：墨天轮、国信证券经济研究所整理

5.5 达梦数据库：高性能数据库管理系统

- **达梦数据库：专注于数据库软件产品和服务。** 武汉达梦数据库股份有限公司成立于2000年，是国内领先的数据库产品开发服务商，国内数据库基础软件产业发展的关键推动者。公司为客户提供各类数据库软件及集群软件、云计算与大数据等一系列数据库产品及相关技术服务，致力于成为国际顶尖的全栈数据产品及解决方案提供商。公司专业从事数据库管理系统与大数据平台的研发、销售和服务，同时可为用户提供全栈数据产品和解决方案，前身是华中科技大学数据库与多媒体研究所，是国内最早从事数据库管理系统研发的科研机构。目前公司已经达到了1000人以上的规模，同时不断拓展自身的渠道，加强生态对于达梦数据库的支持。

达梦数据库发展历程



资料来源：公司官网、国信证券经济研究所整理

5.5 达梦数据库：高性能数据库管理系统

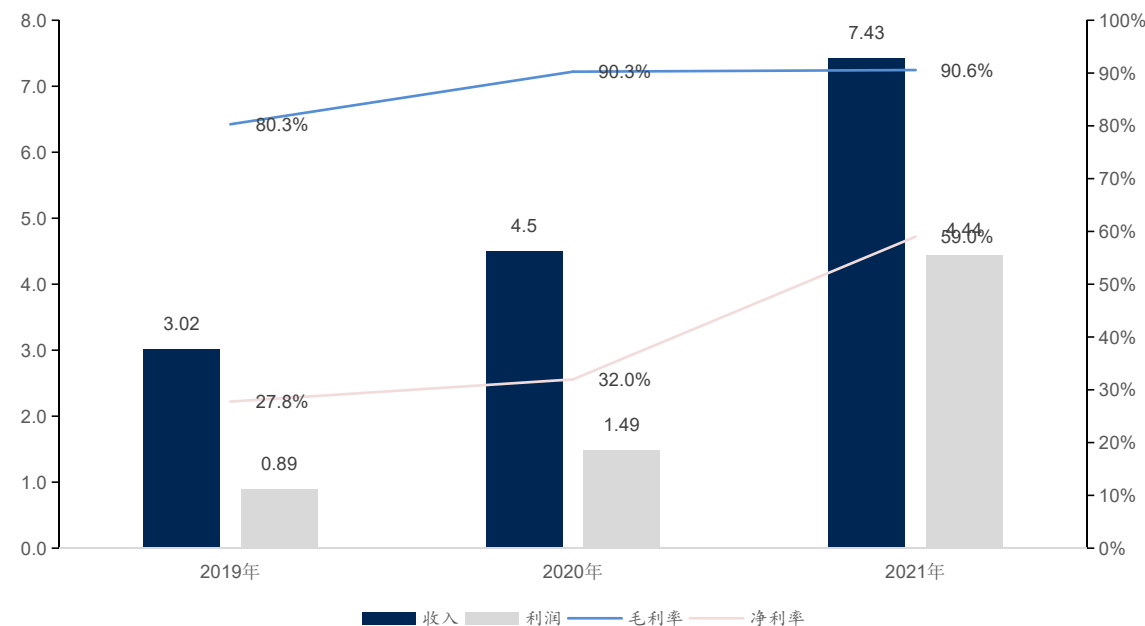
- **达梦数据库产品体系较为丰富。**达梦数据库目前的主要产品较为丰富，包括了数据库管理系统（DM8）、数据共享集群DMDSC、新一代分布式数据库DMDPC、数据守护集群DMDataWatch、分析型大规模数据处理集群DMMPP、读写分离集群DMRWC、数据交换平台软件DMETL和实时同步软件DMHS。这些丰富的产品线，最大程度的提升了公司产品的性能，同时为客户各种各样的应用场景实现提供了可能。
- **达梦收入和利润快速增长，盈利能力较强。**公司在2021年的收入已经达到了7.4亿元，同比增长65.4%，趁着国产化替代的春风营收快速增长，利润实现了4.4亿元，同比增长196.8%。同时，公司的盈利能力较强，毛利率和净利率分别为90.6%、59%，在同类公司中处于佼佼者。同属于传统数据库的人大金仓2021年实现收入3.4亿，利润实现3129万元。

达梦数据库的主要产品

达梦数据库管理系统（DM8）	新一代大型通用关系型数据库,全面支持 SQL 标准和主流编程语言接口/开发框架。
达梦数据共享集群DMDSC	适用于密集交易型场景的共享存储集群,是适合金融行业核心生产系统的解决方案。
新一代分布式数据库DMDPC	适用于金融科技、工业互联网、物联网场景的新一代分布式数据库,计算与存储分离。
达梦数据守护集群DMDataWatch	适用于高可靠环境场景下的数据守护集群,提供不间断数据库服务。
达梦分析型大规模数据处理集群DMMPP	提供高可用性和动态扩展能力,是超大型数据应用的高性价比通用解决方案。
达梦读写分离集群DMRWC	通过客户端来实现读、写事务的自动分离,减轻主机的负载,提高系统的并发能力,增强系统性能。
达梦数据交换平台软件DMETL	构建数据中心、数据仓库、数据交换和数据同步等数据集成类应用的理想平台。
达梦数据实时同步软件DMHS	广泛应用于应急系统、容灾备份、负载均衡、数据移植、联机维护、订阅分发、多业务中心等业务领域。

资料来源：公司官网、国信证券经济研究所整理

达梦数据库收入利润快速增长（单位：亿元）



资料来源：iFind、国信证券经济研究所整理

5.6 TDSQL：企业级分布式数据库

- **TDSQL为客户提供完整的分布式数据库解决方案。**腾讯云原有的TDSQL、TBase、CynosDB三大产品线将统一升级为腾讯云企业级分布式数据库TDSQL，全新升级后的腾讯云TDSQL涵盖分布式、分析型、云原生等多引擎融合的完整数据库产品体系。腾讯云数据库(TencentDB)是腾讯提供的高可靠、高可用、可弹性伸缩的云数据库服务产品的总称，在公有云和专有云领域提供全行业数据库解决方案，可轻松运维主流开源及商业数据库。公司产品可以提供业界领先的金融级高可用、计算存储分离、数据仓库、企业级安全等能力，同时具备智能运维平台、Serverless版本等完善的产品服务体系。
- **TDSQL已经服务了金融、证券、政府等重要领域客户。**目前腾讯云的数据库已经拥有超过50万客户，服务1000多家政府客户和2000多家金融客户，每天支撑数十亿笔的交易量。

TDSQL用户 客户案例



5.6 TDSQL：企业级分布式数据库



➤ **TDSQL主要分为MySQL版和PostgreSQL版（原TBase）。**TDSQL MySQL 版（TDSQL for MySQL）是腾讯打造的一款分布式数据库产品，具备强一致高可用、全球部署架构、分布式水平扩展、高性能、企业级安全等特性，同时提供智能 DBA、自动化运营、监控告警等配套设施，为客户提供完整的分布式数据库解决方案。TDSQL PostgreSQL版（TDSQL for PostgreSQL，原TBase）是腾讯自主研发的分布式数据库系统，具备高SQL兼容度、完整分布式事务、高安全、高扩展、多级容灾等能力，成功应用在金融、政府、电信等行业核心业务中。同时提供完善的容灾、备份、监控、审计等全套方案，适用于GB到PB级海量 HTAP 场景。

TDSQL MySQL版的核心性能优势

高度兼容 MySQL语法TDSQL MySQL 版兼容大多数常用的MySQL语法	
领先的分布式架构	TDSQL MySQL版为此提供三种建表方案，提供（分布式）事务特性，提供全局唯一数字序列，支持JSON等能力
不停机弹性扩展	目前单一片最大可支持6TB存储，如果性能或容量不足以支撑业务发展时，在控制台点击，即可自动升级完成。
强同步复制	TDSQL MySQL版默认采用主从架构，可确保99.95%以上可用性，目前强同步复制性能已基本等于异步复制。
支持数据库智能管家	TDSQL MySQL版现已在专有云方案支持数据库智能管家DBBrain，可提供7*24小时实时数据库服务
易于使用的托管部署	只需在腾讯云 TDSQL MySQL 版管理控制台中单击几下，即可在几分钟内启动并连接到一个可以立即投入生产的 TDSQL MySQL 版数据库。

TDSQL PostgreSQL版的核心性能优势

HTAP 双引擎	TDSQL PostgreSQL版具备事务和分析混合处理技术。提供了高效的OLTP能力和海量的OLAP同时处理能力，可降低业务架构复杂度和成本。
分布式事务一致性	TDSQL PostgreSQL版引入全局事务管理节点来管理分布式事务，通过拥有自主专利的分布式事务一致性技术来保证在全分布式环境下的事务一致性。
秒级在线扩容	用户可以在控制台操作实例实现在线自动扩容，且对业务影响时间可以控制在秒级。
去0理想选择	TDSQL PostgreSQL版推出兼容PostgreSQL版本和Oracle版本，其中兼容Oracle版本提供分布式和集中式两种形态，是去0的理想选择。
企业级数据安全	内核支持三权分立的体系，提供数据透明加密，数据脱敏访问，强制访问控制等多个层级的数据安全保障能力
便捷管理	提供备份、恢复、监控、迁移等全套解决方案，通过运维管理平台便捷管理集群。。

资料来源：公司官网、国信证券经济研究所整理

资料来源：iFind、国信证券经济研究所整理

5.7 云和恩墨：智能的数据技术提供商



- **公司简介：**云和恩墨由Oracle ACE总监盖国强先生创建于2011年，公司以“数据驱动，成就未来”为使命，是智能的数据技术提供商。盖国强是中国地区首位OracleACE，开发过基于Oracle数据库的大型ERP系统，后任职于北京某电信增值服务商企业，提供电信级数据库的规划与运维支持，对于Oracle内部技术等具有深入研究，是中国地区最著名的Oracle技术推广者之一，实践经验丰富，长于数据库诊断、性能调整与SQL优化等。目前，公司已经有700多名员工分布在国内外的34个地区。
- **核心产品与业务：**在数据承载（分布式存储、数据持续保护）、管理（数据库软件、数据库云管平台、数据技术服务）、加工（应用开发质量管控、数据模型管控、数字化转型咨询）和应用（数据服务化管理平台、数据智能、隐私计算数据联邦平台）等领域为各个组织提供可靠的产品、服务和解决方案，围绕用户需求，持续为客户创造价值，激发数据潜能。

云和恩墨战略架构总览



资料来源：墨天轮、国信证券经济研究所整理

5.7 云和恩墨：智能的数据技术提供商

- **公司已经具有面向数据的全栈自研产品。**云和恩墨设计和研发面向数据的端到端的软件产品，涵盖数据的承载（分布式存储、数据持续保护）、管理（数据库软件、数据库云管平台）、加工（应用开发质量管控）、应用（数据服务化管理、数据智能和数据联邦）等领域。公司在openGauss生态社区建设上持续投入、联合创新，于2020年9月发布MogDB数据库，成为了openGauss的主要企业版发行商之一。
- **公司从成立以来已经经历了4轮融资。**公司从2011年成立以来，目前已经融资到C轮，最新的一次融资是在2020年12月份，融资金额2亿元，由泓生资本、千合资本等领投。

云和恩墨全栈产品



资料来源：公司官网、国信证券研究所整理

云和恩墨融资历程

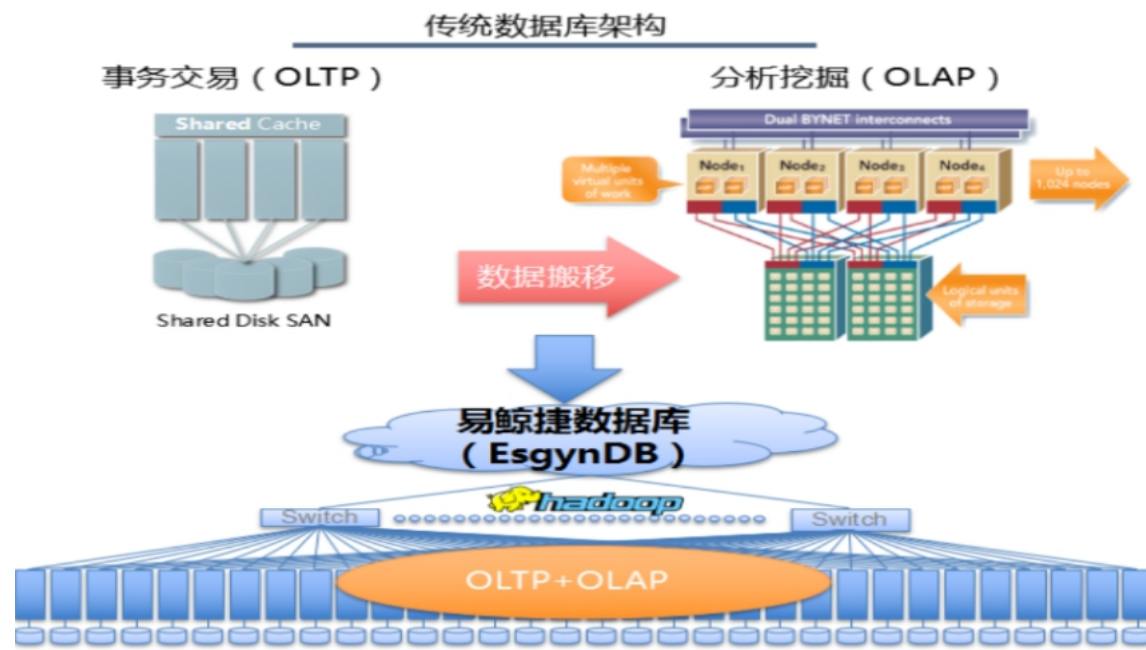
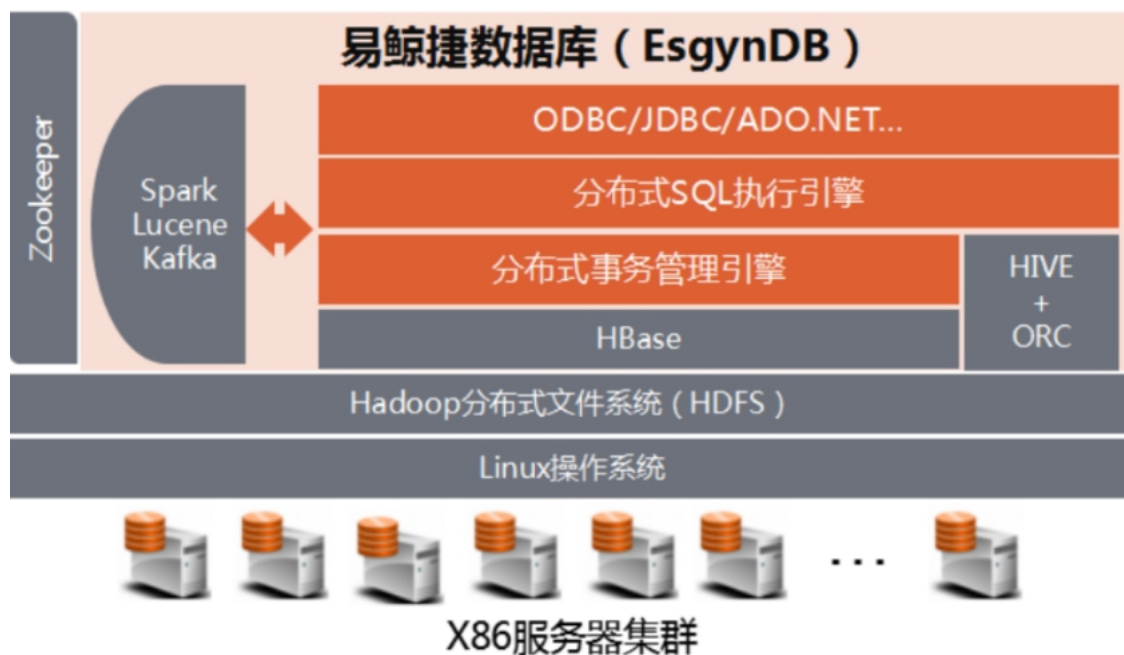
日期	融资轮次	融资金额	投资方
2020/12/31	C轮	20000万元	泓生资本、千合资本、杭实投资、中国互联网基金
2017/10/25	B轮	金额未知	亿联资本、国投创业
2016/12/26	A轮	2000万元	千合资本
2016/3/16	天使轮	1000万元	千合资本

资料来源：墨天轮、国信证券研究所整理

5.8 易鲸捷：中国分布式数据库的领军企业

- **易鲸捷：迈向大数据时代的分布式数据库。** 贵州易鲸捷信息技术有限公司成立于2015年，是一家中方独资的大数据初创公司，也是大数据（贵州）综合试验区唯一一家致力于融合型分布式数据库研发工作的公司，且是国内极少拥有自主核心技术的数据库厂商，研发团队源自天腾公司。公司的CEO武新博士为原华为云高斯数据库首席架构师、华为IT产品线数据平台部CTO，曾任南大通用CTO、CEO，是中国数据库领军人物之一。
- **公司的核心产品为EsgynDB数据库。** 公司核心产品EsgynDB是一款分布式的融合数据库，底层采用Hadoop架构，结合自主研发的SQL引擎，满足大数据时代的业务操作和数据分析需要，实现了性能的提高和扩展性的增加，满足高速的大量数据处理，支撑在线业务操作和实时数据分析；另一方面继承传统数据库的特点，提供完整的SQL和ACID事务支持，降低了开发难度和周期。

易鲸捷数据总体架构



资料来源：墨天轮、国信证券经济研究所整理

5.8 易鲸捷：中国分布式数据库的领军企业



- **公司打造了针对不同场景的多款数据库周边产品。**公司目前主要的产品线除了EsgynDB，还有：1）QianBase xTP，面向核心业务的可靠原生分布式数据库；2）QianBase MPP，面向OLAP&HTAP的数据处理平台；3）QianBase TP，经典的集中式数据库系统；4）QianBase HTAP，Hadoop大数据平台的HTAP数据库。公司所有知识产权自主可控，在技术上具备与全球顶尖技术竞争的實力且互相兼容，可支撑金融核心交易系统对于高并发、高可用及高可靠性的需求，彻底解决我国银行、保险、证券等金融核心交易系统的底层数据库技术问题。
- **公司已经盈利，获重要股东加持。**公司2020年实现收入2.3亿元，利润4550万元。同时，公司2022年4月获中国软件战略投资，正式被纳入中国电子业务体系，后续公司将逐步打造并形成以芯片、操作系统、数据库三大基础软硬件为基础的PEK生态体系。

易鲸捷主要产品线

QianBase xTP	面向核心业务的可靠原生分布式数据库
QianBase MPP	面向OLAP&HTAP的数据处理平台
QianBase TP	经典的集中式数据库系统

QianBase HTAP Hadoop大数据平台的HTAP数据库

易鲸捷融资历程

日期	融资轮次	融资金额	投资机构
2021/5/31	B++轮	未披露	中国软件
2020/12/28	B+轮	未披露	中天金融、德威资本
2020/6/29	B轮	未披露	高科信浚
2020/1/1	A++轮	未披露	南京高科
2018/7/12	A+轮	未披露	东信科创
2018/4/20	A轮	未披露	远东控股集团
2017/6/6	天使轮	未披露	贵阳创投、贵阳新区新兴产业发展基金
2016/8/19	种子轮	未披露	贵阳创投

资料来源：公司官网、国信证券经济研究所整理

资料来源：墨天轮、国信证券经济研究所整理

5.9 巨杉数据库：湖仓一体，金融级分布式数据库

- **SequoiaDB巨杉数据库是一款金融级分布式关系型数据库。**公司自研的原生分布式存储引擎支持完整ACID，具备弹性扩展、高并发和高可用特性，支持MySQL、PostgreSQL和SparkSQL等多种SQL访问形式，同时支持MongoDB引擎，适用于核心交易、数据中台、内容管理等应用场景。巨杉数据库联合创始人&CTO王涛，曾就职于IBMDB2数据库北美实验室，作为IBMDB2全球最高技术专家小组的成员，参与了IBM下一代大数据平台的架构规划，精通数据库内核及体系结构，是数据库领域国际顶尖的专家。王涛在在IBM多伦多实验室工作了八年后，选择回国创业，带领团队开发出了新一代分布式数据库——SequoiaDB巨杉数据库。
- **SequoiaDB数据库的主要技术特点：**1) 标准SQL支持，MySQL协议级兼容；2) 金融级分布式OLTP；3) 分布式架构；4) Multi-Model多模数据引擎；5) HTAP混合事务/分析处理；6) 数据安全与多活容灾。

巨杉数据库核心产品线

三大核心产品线

分布式数据库



SequoiaDB
实时数据湖

内容管理平台



SequoiaCM
多模数据湖

湖仓融合平台



SequoiaDP
湖仓一体

5.9 巨杉数据库：湖仓一体，金融级分布式数据库



- **巨杉主要产品线包括了分布式数据库、内容管理平台和湖仓一体平台。**公司主要产品线：1) SequoiaDB，基于巨杉自研的原生分布式内核，支持跨结构化、半结构化、非结构化的多模数据处理，提供一万以上并发连接下的万亿级数据实时SQL查询性能；2) SequoiaCM，基于SequoiaDB 分布式数据库，结合微服务化的基础内容管理体系，为存储、管理海量非结构化数据，提供非结构化数据处理应用程序编程接口、标签管理及检索、异构存储多级缓存等能力的内容数据管理平台；3) SequoiaDP，基于SequoiaDB分布式数据库的湖仓一体平台，兼容多种结构化数据、半结构化数据和非结构化对象数据，还提供包括流式计算、高性能列存分析引擎、跨引擎数据ACID一致性等能力。
- **公司目前已经融资了5轮。**公司从成立以来已经融资了5轮，最近的一轮是在2020年的D轮融资，由越秀产业基金、中金资本领投。

巨杉主要产品线

分布式数据库 SequoiaDB 实时数据湖	基于巨杉自研的原生分布式内核，支持跨结构化、半结构化、非结构化的多模数据处理，提供一万以上并发连接下的万亿级数据实时SQL查询性能，适用于历史数据平台、全量数据平台、实时数据中台等需要将海量多模数据用于对客服务，以优化业务效率，提升客户满意度的场景。
内容管理平台 SequoiaCM 多模数据湖	基于SequoiaDB 分布式数据库，结合微服务化的基础内容管理体系，为存储、管理海量非结构化数据，提供非结构化数据处理应用程序编程接口、标签管理及检索、异构存储多级缓存等能力的内容数据管理平台，应用于非结构化数据治理与需要超大容量、弹性扩展的内容管理系统的场景。
湖仓融合平台 SequoiaDP 湖仓一体	基于SequoiaDB分布式数据库的湖仓一体平台，除了具备多模能力，兼容多种结构化数据、半结构化数据和非结构化对象数据引擎以外，还提供包括流式计算、高性能列存分析引擎、跨引擎数据ACID一致性等能力。适用于需构建融合数据平台及在保持事务一致性的基础上，进行高性能实时分析的场景。

资料来源：公司官网、国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

巨杉主要融资历程

日期	融资轮次	融资金额	投资机构
2020/10/13	D轮	数亿人民币	越秀产业基金、中金资本、元禾重元
2018/9/16	C轮	数千万美元	启明创投、DCM Ventures、嘉实投资
2016/7/14	B轮	1000万美元	启明创投、DCM Ventures
2014/12/2	A轮	100万美元	启明创投
2014/9/24	Pre-A轮	数百万人民币	极客邦创投

资料来源：墨天轮、国信证券经济研究所整理

第一，研发成果商业化转化不确定性的风险；

第二，数据库厂商恶性竞争的风险；

第三，数据库核心成员流失的风险；

第四，国产替代进度不及预期的风险。

国信证券投资评级		
类别	级别	定义
股票投资评级	买入	预计6个月内，股价表现优于市场指数20%以上
	增持	预计6个月内，股价表现优于市场指数10%-20%之间
	中性	预计6个月内，股价表现介于市场指数±10%之间
	卖出	预计6个月内，股价表现弱于市场指数10%以上
行业投资评级	超配	预计6个月内，行业指数表现优于市场指数10%以上
	中性	预计6个月内，行业指数表现介于市场指数±10%之间
	低配	预计6个月内，行业指数表现弱于市场指数10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。 ， 本公司不会因接收人收到本报告而视其为客户。 未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。



国信证券
GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046 总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032