

2022 年中国数据库研究报告

数据库 | 研究报告

核心摘要：

中国数据库市场规模：据艾瑞统计，2021 年中国数据库市场总规模达 286.8 亿元，同比增长 16.1%。中国数据库市场虽受疫情影响，但整体稳步增长，由多方面因素促成：1) 信创利好；2) 需求驱动；3) 国产厂商厚积薄发；4) 国内用户对基础软件的 IT 支出，尤其是国产数据库的付费意愿逐年提升。

在数据库市场供给侧：分布式架构的技术特性趋于融合，云化部署大势所趋，使用门槛不断降低；数字化业务场景不断丰富，业务需求驱动 HTAP 技术成为重要的技术底座；在开源数据库领域，国产开源时间不长，整体表现尚有提升空间；国产数据库借信创东风，迎来了前所未有的发展机会。

在数据库市场需求侧：在数字化转型浪潮下，企业用户的数据已初具规模，业务场景和数据结构越来越复杂，对于数据分析的需求日趋强烈。数据库部署方式呈现多元化，其中分布式架构和云部署的渗透率显著提升。随着国产数据库的技术创新以及产品性能提升，国产产品的头部效应明显，用户体验仍有待进一步完善。

数据库行业趋势洞察：数据库出海成为国产厂商的全新增量逻辑，欧美线路和东南亚线路是国产厂商出海的优选项；云原生进入 2.0 时代，实现了云资源对数据库的细粒度资源拆解和优化编排，云托管数据库进一步向原生式数据库推演；数据库与大数据技术的边界不断模糊，呈现融合趋势。

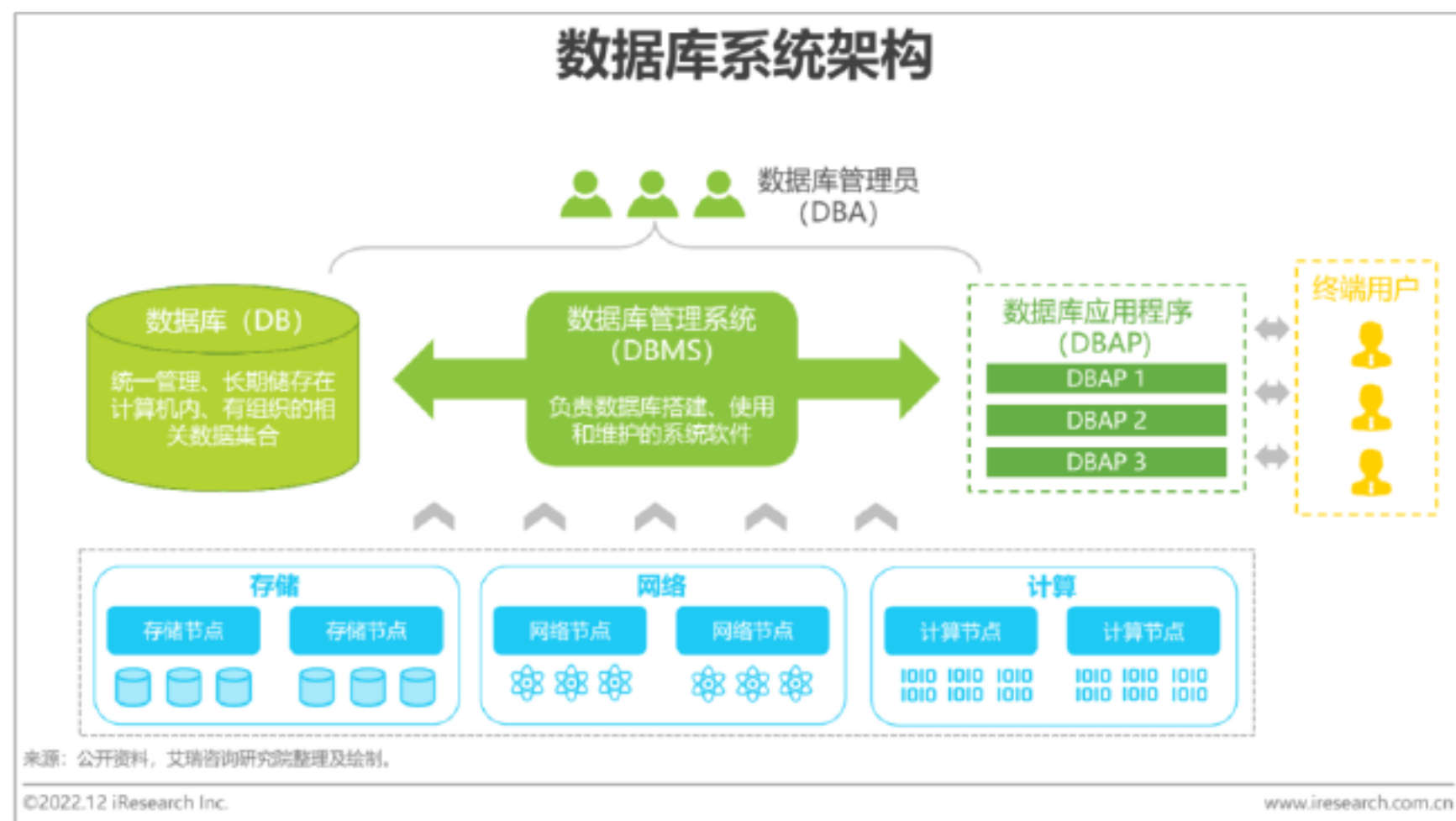
01

数据库概述

定义

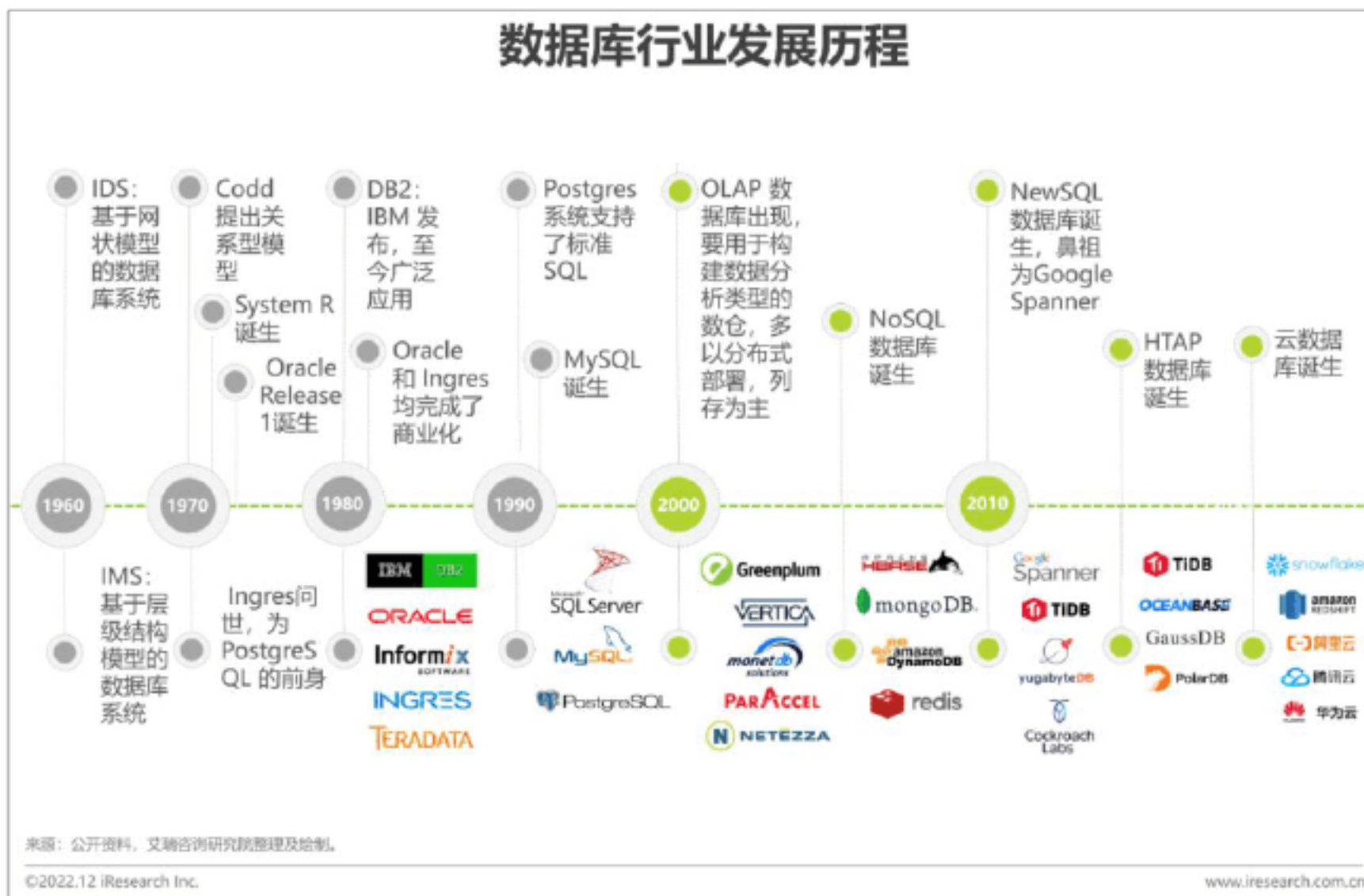
数据库是由 DBMS 统一管理的，有组织、共享的数据集合

数据是描述事物的符号记录，具有多种表现形式，包括文字、图形、图像、声音和语言等。数据库是相互有联关系的若干数据的集合，特点是数据间联系密切、冗余度小、独立性较高、易扩展，并且可为各类用户共享。数据库管理系统（DBMS）是负责数据库搭建、使用和维护的系统软件，通过组织、索引、查询、修改数据库文件，实现数据定义、组织、存储、管理以及数据库操作、运行、维护等主要功能。围绕 DBMS 构建包含硬件操作系统、数据库管理系统、数据库管理员以及相关机制配套在内的数据库系统（DBS），才能有组织、动态地存储大量相关数据，方便多用户访问。



发展历程

硬件技术、应用场景等快速变化，行业格局百花齐放



分类

数据库按照不同维度，分类方法多种多样



按部署模式分类



来源：艾瑞咨询研究院根据公开资料整理及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

按架构分类



按功能分类



来源：艾瑞咨询研究院根据公开资料整理及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

按商业模式分类



按存储介质分类



来源：艾瑞咨询研究院根据公开资料整理及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

02

数据库供给侧分析

分布式现状

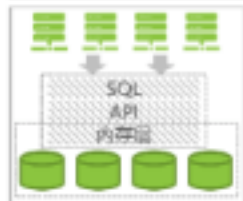
技术特性趋于融合，云化部署大势所趋，使用门槛不断降低

分布式数据库是通过计算机网络将物理分散的多个数据库单元连接组成的逻辑上统一的数据库。目前分布式数据库的技术路线分为三类：分库分表+中间件方案、共享存储分布式数据库、去中心化的分布式数据库。当下分布式数据库市场产品繁多，但其技术特性逐渐趋同，从两个角度分析：其一，系统架构。存储采用开源或者自研内核，用一主多从的数据副本保证数据完整性，计算通过 SQL 语句解析分发和结果汇聚，管控用于保证各模块的高可用和故障切换等管理功能。其二，开源和自研两条研发线路。开源主要基于 MySQL 和 PostgreSQL，难点在于源代码改造的技术难度大；自研周期长且技术要求高，国内完全自研产品屈指可数，如 TiDB 和 OceanBase 等。未来分布式数据库的核心竞争要素为分布式事务的稳定性、产品全面性以及生态工具的完备性。

分布式数据库技术线路分类



- 方案：分库分表+中间件，下层单机数据库提供存储，其上封装中间层补足分布式能力
- 优势：兼容性好，学习成本低，用户门槛低
- 劣势：代码修改成本高，计算节点扩展性、全局事务能力及高可用不佳，机器冗余度高



- 方案：共享存储分布式，共享存储集群，提供动态扩容的分布式高性能存储，计算节点独立
- 优势：兼容性好，应用无需改造；事务性能好，读写响应快；共享存储一致性高
- 劣势：依赖共享存储系统，扩展性有限，难以做到跨地域高可用，架构可改造性较低



- 方案：去中心化的分布式，各节点独立计算和存储功能，且不共享数据 (shared-nothing)，架构存算分离，通过共识算法保证副本高可用
- 优势：架构高度解耦，按需扩展，无数量和规模限制，数据一致性的安全机制更好，高可用高兼容
- 劣势：对硬件要求较高，上下游生态适配度仍需进一步完善

来源：艾瑞咨询研究院根据公开资料整理及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

分布式数据库发展方向

1 技术特性融合，能力补齐
分布式数据库的不同技术路线逐渐融合，取长补短，丰富产品能力，集中体现于两方面：第一，分布式事务的一致性提升，全周一致性协议保护加强；第二，分布式场景下数据分片智能化，多副本细粒度控制能力得到加强。

2 云化与云原生设计
分布式数据库与云环境加速融合。一方面，分布式的架构优势可将传统部署与公有云资源结合，解决私有部署灵活度低、成本高等问题；另一方面，云环境可低成本加强分布式架构的高容灾和算力扩展，融合云基础资源实现云原生。

3 软硬件结合，异构芯片混合部署
以多核 CPU、异构计算（如 DSP、GPU、ASIC、FPGA）持久化内存等为代表新硬件，为分布式数据库提供丰富的想象空间。通过异构芯片的混合部署，让分布式数据库在不同层级上提供锁定模式的安全机制，提供分离模式的性能。

4 单体分布式一体化数据库架构
在单机部署下实现分布式数据库的完整功能，从技术上突破分布式数据库的单机性能瓶颈，极大降低分布式数据库的部署门槛。无论用户自身规模大小，采用单机部署或水平扩展部署，一套数据库系统便可满足核心系统要求。

来源：艾瑞咨询研究院根据公开资料整理及绘制。

©2022.12 iResearch Inc.

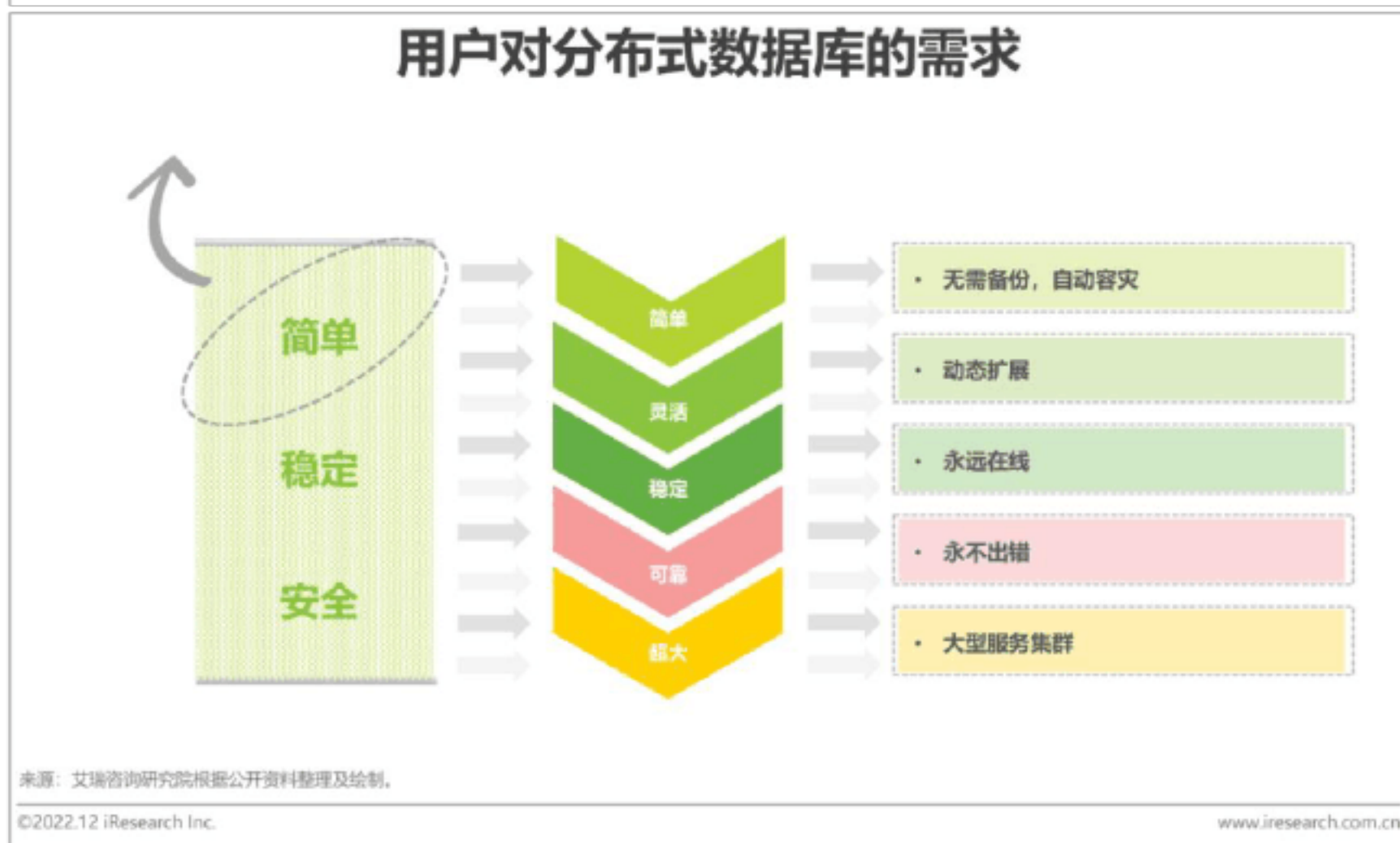
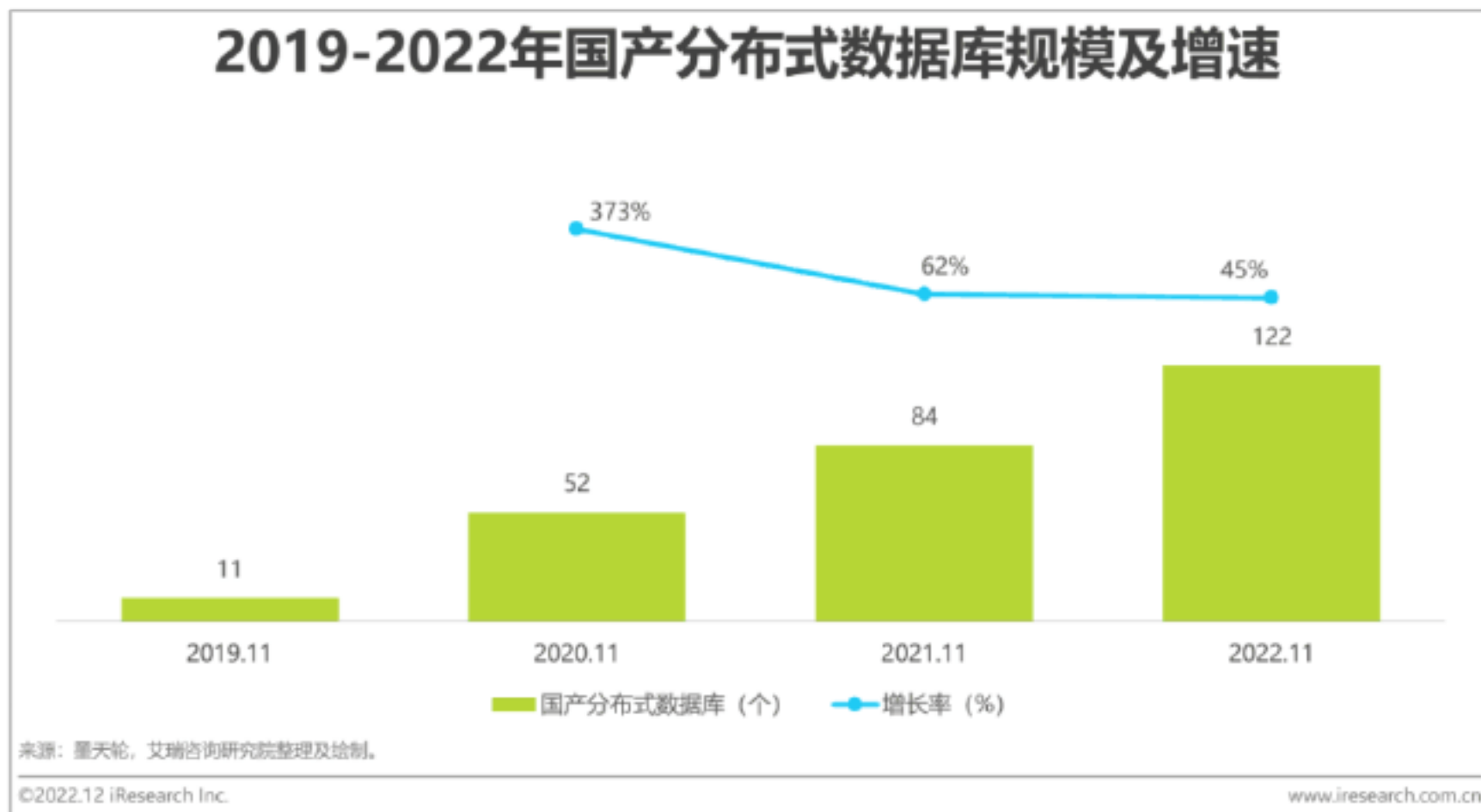
www.iresearch.com.cn

分布式前瞻

加强底层故障隐患感知，提升异常处置能力，实现开箱即用

墨天轮数据显示，国产分布式数据库自 2019 年以来保持高速增长，热度持续高涨。而通过用户调研了解到，面对分布式数据库的故障时，运维人员往往束手无策。主要原因是相较集中式数据库，分布式数据库对 IT 基础设施可靠性要求更高了，其核心代码主要覆盖了 SQL 实现和数据存储，而未能自动感知各种对数据库稳定性、性能、并发能力有重大影响的故障隐患，也难以在代码层面对此类问题进行处理，从而实现故障自动规避。未来分布式数据库应提升基础数据探

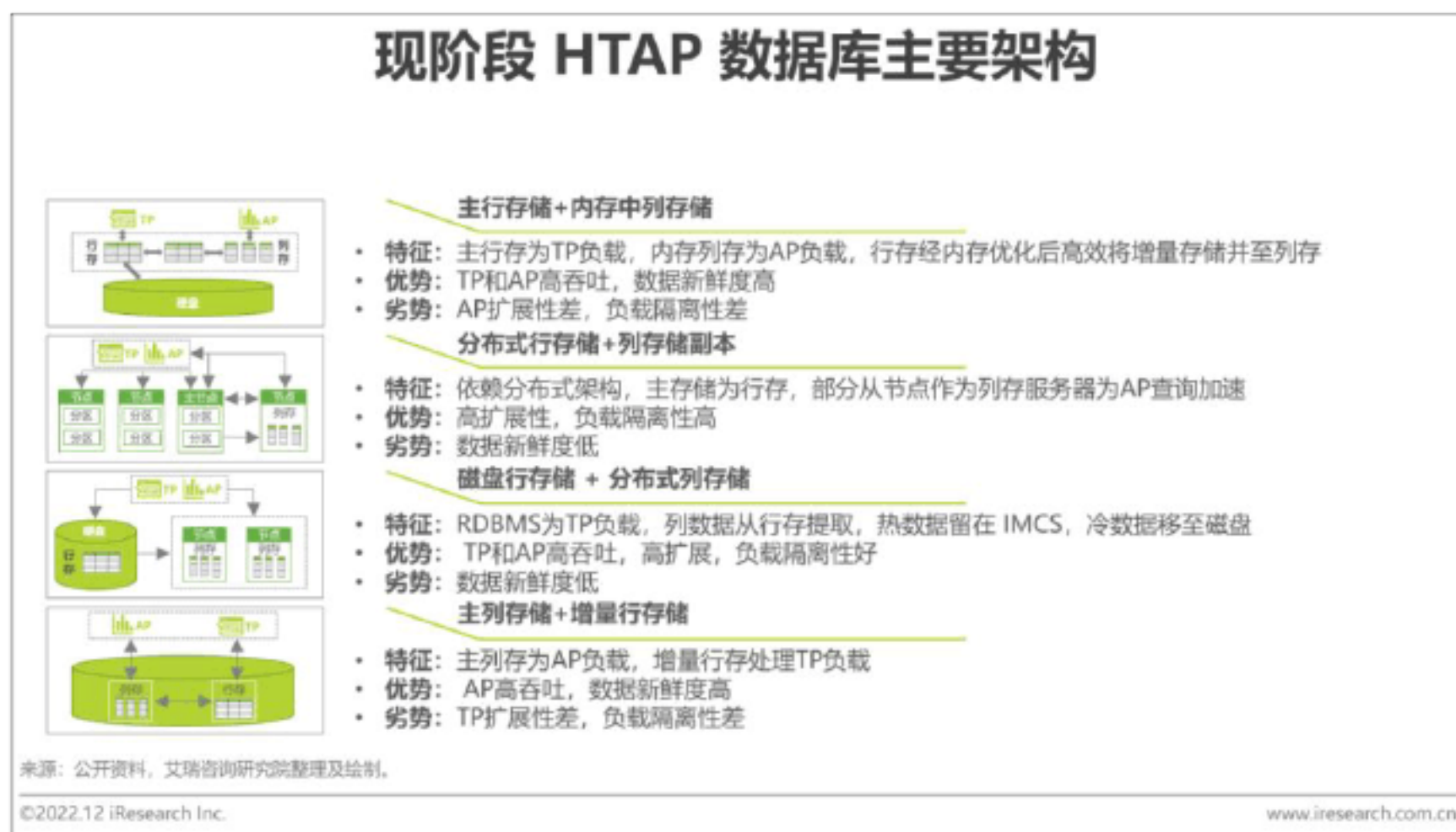
测和分析能力，随时针对出现的异常隐患提前处置，实现无需运维人员过多干预的高效自治运行，让分布式数据库从 IT 工程化产品变为真正开箱即用的通用型产品。



HTAP 现状

业务驱动 HTAP 技术成为重要底座，与分布式技术天然适配

随着业务系统接入数据源复杂性提升，混合负载需求愈发强烈，数据库技术正向多源异构、高实时并发、多 SQL 标准接口的方向演进。过去，HTAP 数据库的技术架构以扩展主行存技术，在行存基础上加列存的方式为主，典型代表有 SQL Server、Oracle 和 L-stor 等。随着分布式技术不断成熟，天然适配 HTAP 数据库，于是开启了分布式的架构实现，满足了高并发需求，典型代表有 SingleStoreMySQL、Heatwave 和 Greenplum。我们认为未来 HTAP 仍应是在 OLTP 数据库的基础上，引入原生分布式架构和低成本存储引擎以扩展大数据量 OLAP 能力。最后，为更好支持 OLAP 的数据开发和建模能力，提升实时分析，HTAP 支持物化视图和外部表，并与各类数据开发工具和 BI 工具适配对接。





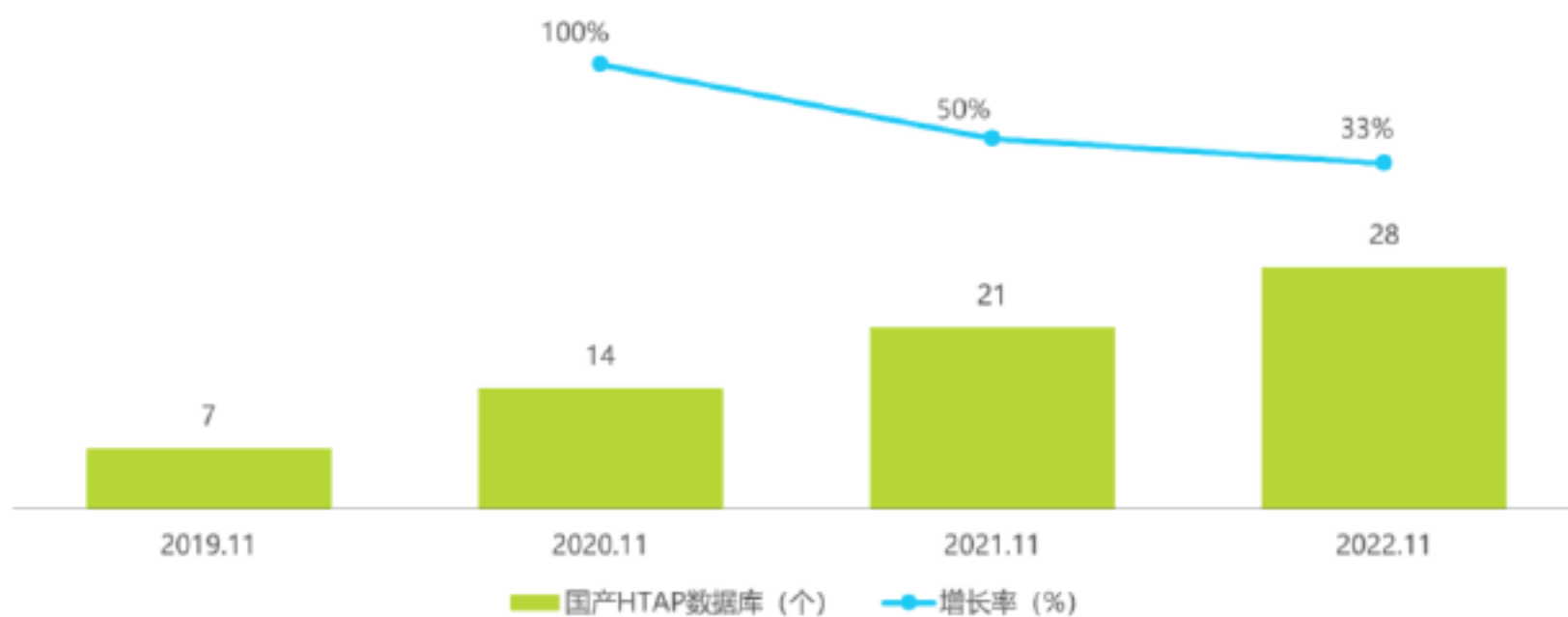
HTAP 前瞻

运用内置流处理能力进一步加强 AP 和 TP 间的融合

墨天轮数据显示，自 2020 年以来国产 HTAP 数据库发展迅猛，几乎成为新兴数据库的必选项。Gartner 也指出 HTAP 已经成为全球范围内新一代数据库的入场筹码，HTAP 能力成为数据库重要选项。随着数据技术的不断创新，未来 HTAP 数据库会进一步加强 AP 和 TP 之间的连接和融合，在数据库内部实现 AP 和 TP 之间的内置流

(Streaming) 处理能力。通过将 ETL 工作内置于数据库当中，让 HTAP 数据库同时具备 AP、TP 和流 (Streaming) 能力。用户可以按需创建各类表，运用流处理能力将表连接，从而获得简单、便捷的数据处理能力。数据库技术的进一步融合会打破当下数据栈的割裂状态，HTAP、流批一体、湖仓一体等技术趋势最终会让数据集集中在简单易用、安全可靠、高性价比的数据平台。

2019-2022年国产 HTAP 数据库规模及增速

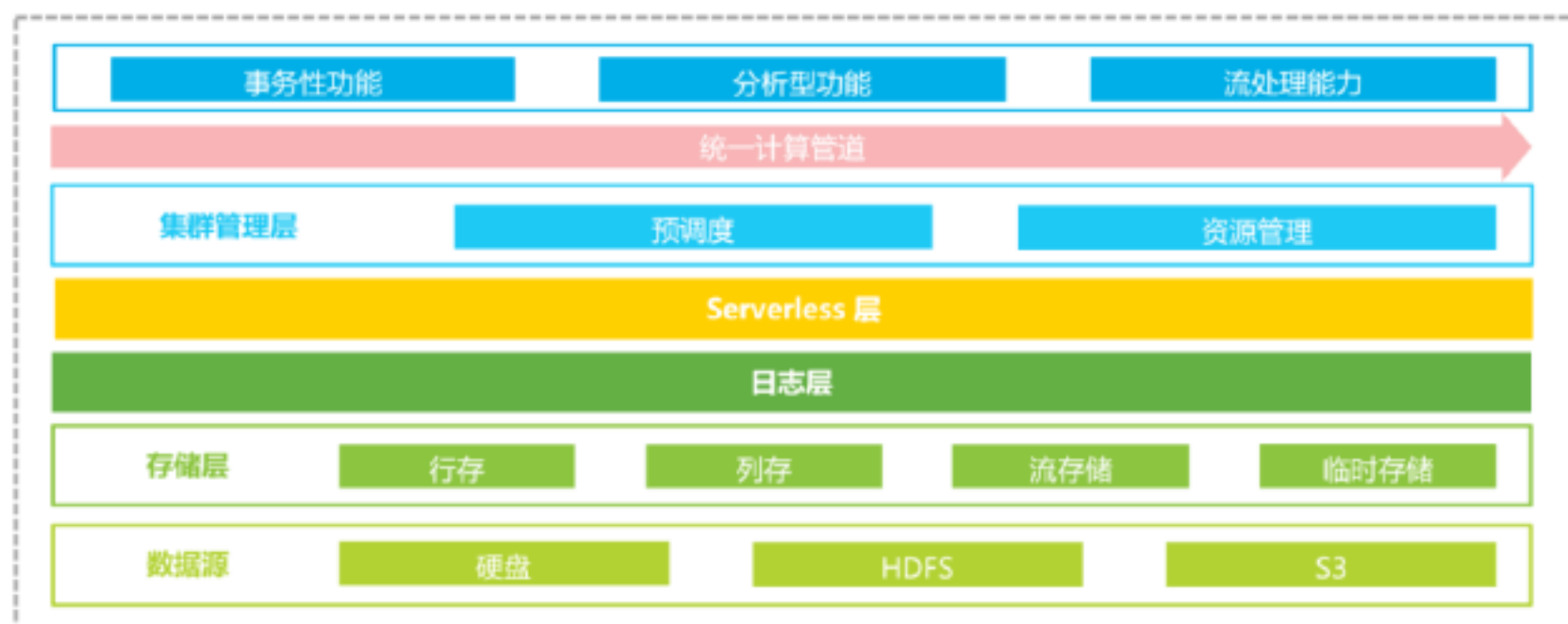


来源：墨天轮，艾瑞咨询研究院整理及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

基于内置流处理能力的 HTAP 架构



来源：艾瑞咨询研究院根据公开资料整理及绘制。

©2022.12 iResearch Inc.

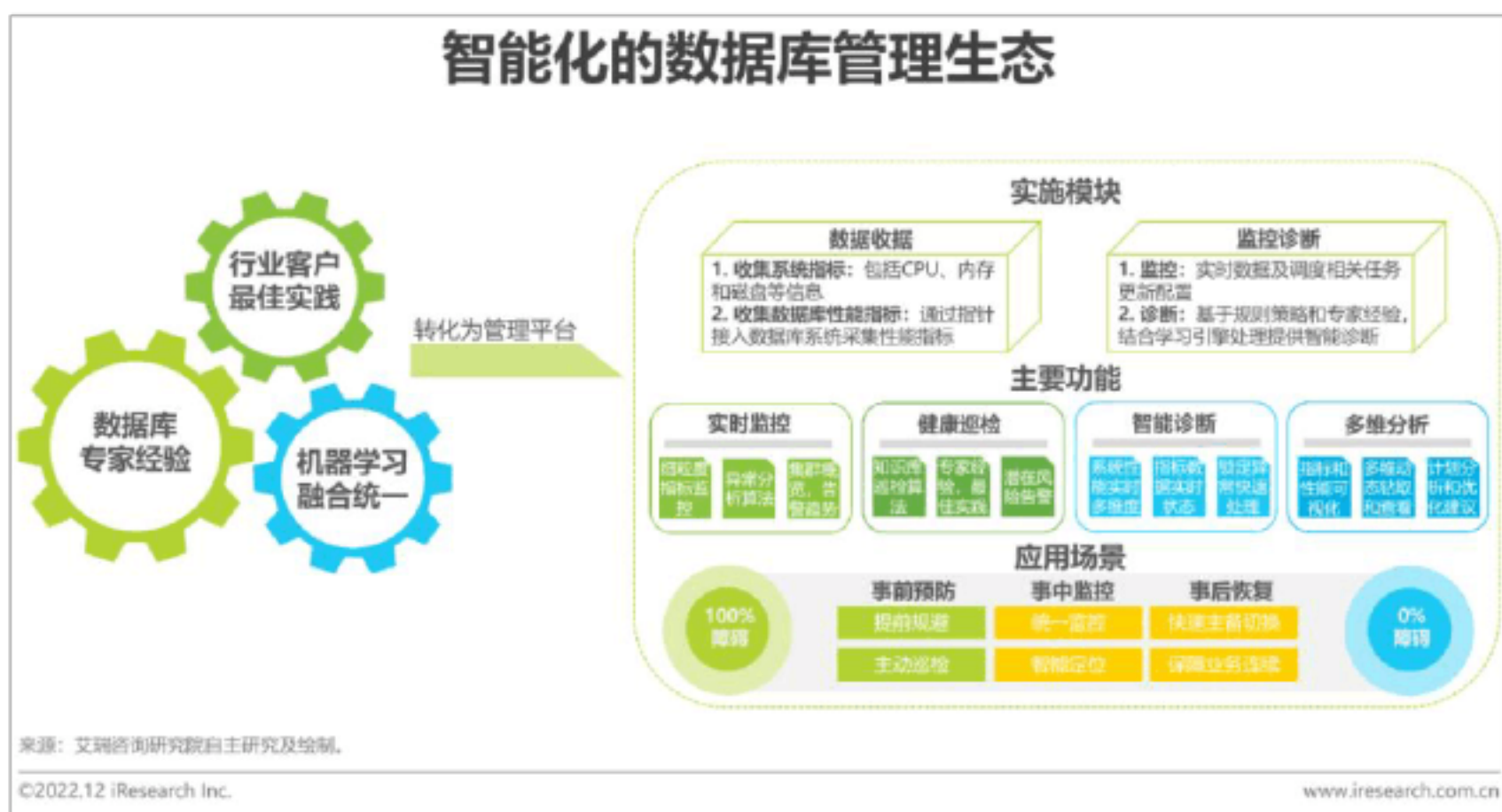
www.iresearch.com.cn

智能运维

融合智能技术实现复杂数据库环境的管理自治

DT 时代数据库技术架构和运行环境日趋复杂，种类从单一产品转变为混合型商业数据库和开源数据库组合，依靠人工运维显得捉襟见肘。作为智能化的数据库周边工具，数据库管理平台将机器学习与数据管理在功能上融合统一，利用机器学习增强系统设计开发，以标准

化、自动化、智能化的方式提供实时监控、健康巡检、智能诊断、多维分析等数据库管理服务。数据库管理平台的本质是数据库管理经验的代码化，核心方法论是云资源池化、分层解耦以及服务化，实现手段是基于微服务、分布式等云技术实现多元混合数据库环境的统一管理，目标是实现海量数据高安全、高可用、高性能的运维要求，助力数字化转型。

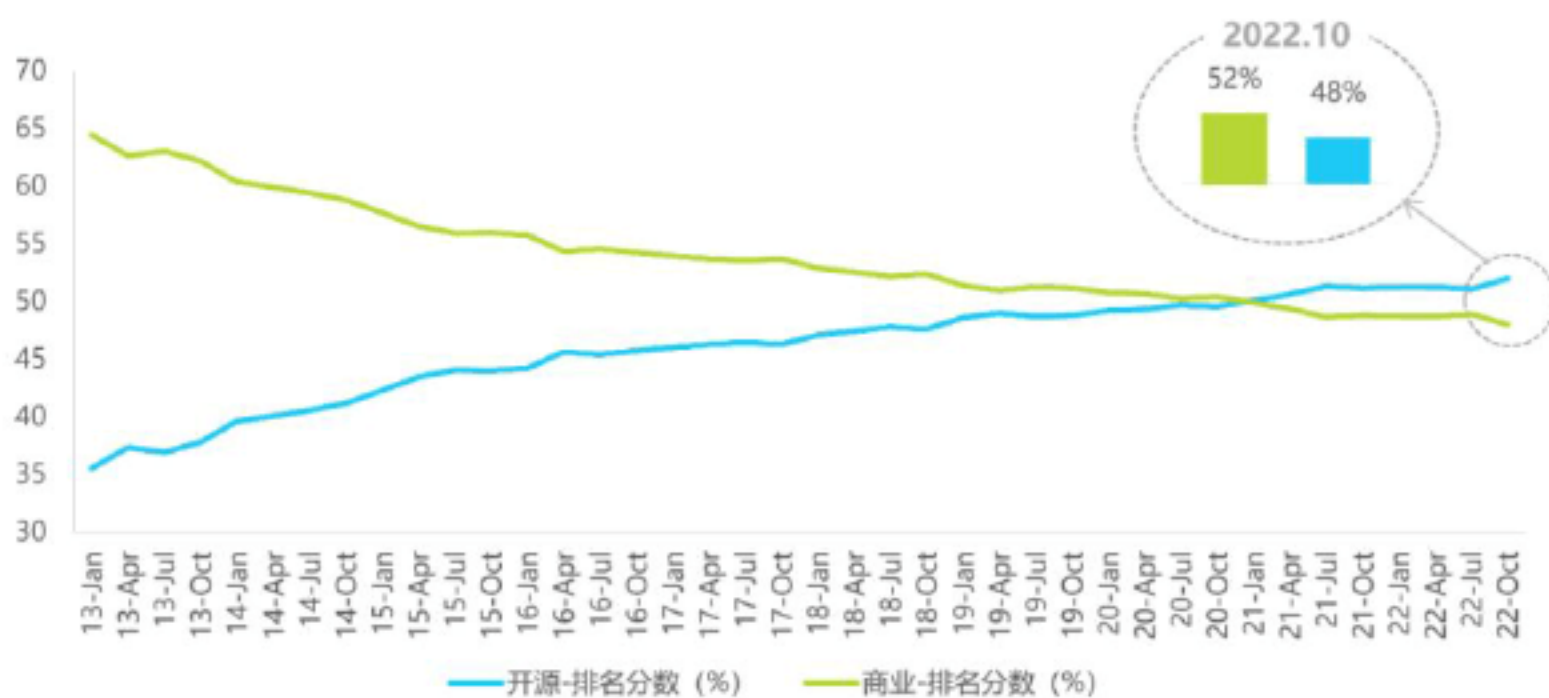


全球开源

热度持续，成为国产数据库走向国际化的重要途径

开源即开放源代码，用户可在源代码的基础上完成学习和修改。DB-Engines数据显示，截至2021年1月，开源数据库的全球部署首次超过商业数据库。此后，开源数据库凭借在成本、产品丰富度、社区活跃性等方面的突出优势，持续获得了全球的高关注度。开源成为驱动数据库技术和市场变革的重要力量。在此趋势下，开源成为数据库项目冷启动的最佳方式，也是构建国产自研数据库生态、加速技术迭代、让国产数据库走向国际化的重要途径。

2013-2022年全球数据库流行趋势



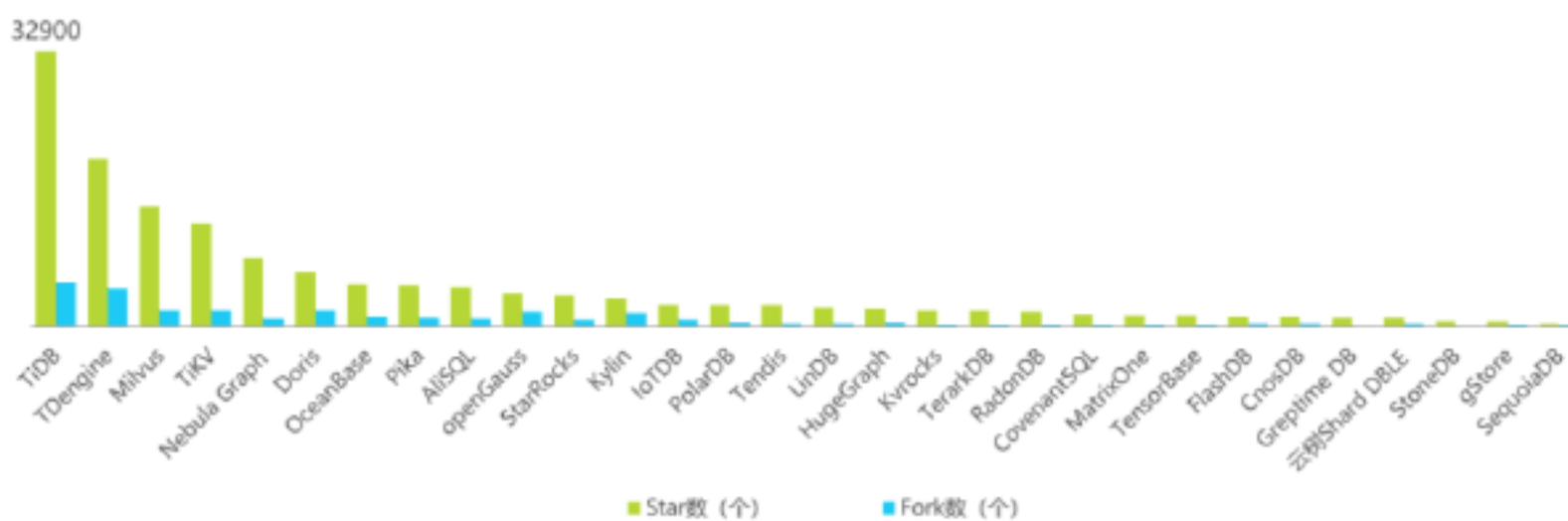
注释：1、DB-Engines排名是按照当前流行程度的排名，较实际使用情况具有一定的超前性，具体指标包括网站上系统提及的频率、Google trends、IT论坛上系统讨论的频率、提及系统的工作机会数量、专业网络中提及系统的配置文件数和系统在社交网络中的提及次数；
2、开源、商业-排名分数 (%) 为相对值。
来源：DB-Engines，艾瑞咨询研究院自主研究及绘制。
©2022.12 iResearch Inc. www.iresearch.com.cn

国产开源

国产数据库开源时间不长，整体表现有待提升

近年来国内数据库厂商逐步推进开源策略。但因国内数据库开源时间不长，运营有待加强，整体表现有较大提升空间。以 TiDB 为例，从开发者、使用者及代码角度分析，其开发者活跃、用户众多、代码变化较大，平均 2-3 年完成一次软件重构，侧面印证了上述特征。

国产开源数据库在 Github 的整体活跃度



来源：Github，艾瑞咨询研究院根据公开资料自主研究及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

国产开源数据库在 Github 的活跃度细分情况

项目	开发活跃度			使用者维度			代码维度		贡献者维度
	active	merge	open	active	close	new	新增	删除	数量
TiDB	604	392	212	599	271	328	123161	24239	1146
OceanBase	13	4	9	45	9	36	125473	28777	74
openGauss	44	-	-	198	1	1	-	-	121
TDengine	283	246	37	65	36	29	25943	9269	140
PolarDB	5	5	0	4	1	3	2908	368	41

来源: Github, 艾瑞咨询研究院根据公开资料自主研究及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

国产数据库开源的必要性和发展趋势



必要性 发展趋势

获取市场信任

基于国产数据库行业现状，只有选择开源，彻底打开“黑盒”，才能真正获取用户信任，打开市场局面

开发方式被重塑

开发者的用户体验得到高度关注，开发者的门槛不断降低，只需具备前端能力，其他都由全链路的可伸缩服务替代

构建成熟的数据库生态

单纯依靠个别技术或软件很难构建成熟、完善的产品，只有选择开源，依赖社区生态推进，才能打磨出出色的数据库产品

Serverless

基于云原生的 Serverless 生态和技术，各链路 Serverless 解决方案串联成全栈 Serverless 体验

满足业务场景敏捷性

闭源数据库厂商的视角相对滞后，产品反馈链条较长，只有选择开源，才能及时捕获业务场景的敏捷变化

从 SQL 到 API

将数据库用户定位至应用开发者，简化 API 封装和使用，拉近 SQL 到实际应用的距离，让应用开发友好性最大化

来源: 艾瑞咨询研究院根据公开资料自主研究及绘制。

©2022.12 iResearch Inc.

www.iresearch.com.cn

中国数据库产业图谱

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/487004113046006162>