

2021年 平凯星辰PingCAP企业前景探析： 云化开源领步，商业化露锋芒

Corporate Prospect of PingCAP: Database leader featured by cloud-based and open-source, initializes commercialization stage

PingCAPの企業展望：
クラウドベースのオープンソースを特徴とするデータベース
リーダーは、商業化段階を初期化します

报告标签：分布式数据库、开源、云原生

主笔人：胡竣杰

报告说明

在2015年，数据库市场仍被Oracle等巨头垄断，PingCAP的三位极客创始人瞄准了数字化转型趋势与当代MySQL架构无法满足未来需求，以开源立旗，以解放企业生产力为初心，站在NewSQL的技术制高点上潜心耕耘工业级的技术实现，重新定义数据库基础设施。

2021年，PingCAP团队已带出了TiDB、TiKV、Chaos Mesh®等明星开源项目，累积了数据库领域工程师极高的口碑和流行度，在D轮融资以2.7亿美元刷新全球数据库赛道的融资历史，兼具技术、人气和资本，步入商业化元年，蓄势待发。

在2021DevCon大会上发布了TiDB Cloud服务和Ti-Star计划，在企业软件服务赛道上加速。待商业化长考之后，PingCAP或将引领未来数据服务平台化趋势。愿早修得屠龙技，但辟荆棘留初心。

报告提供的任何内容（包括但不限于数据、文字、图表、图像等）均系头豹研究院独有的高度机密性文件（在报告中另行标明出处者除外）。未经头豹研究院事先书面许可，任何人不得以任何方式擅自复制、再造、传播、出版、引用、改编、汇编本报告内容，若有违反上述约定的行为发生，头豹研究院保留采取法律措施、追究相关人员责任的权利。头豹研究院开展的所有商业活动均使用“头豹研究院”或“头豹”的商号、商标，头豹研究院无任何前述名称之外的其他分支机构，也未授权或聘用其他任何第三方代表头豹研究院开展商业活动。

01 PingCAP企业综述

在商业化目的的考量之上，秉承对传统 Hacker 精神的贯彻，PingCAP坚定开源，吸引了国内外技术专家加入，是以开源基因嵌入的产品技术驱动型公司。PingCAP 坐拥 TiKV 与 Chaos Mesh®两个CNCF项目，具备全球先进的工业级分布式架构理论实践。

02 数据库行业分析

企业数据规模激增对分布式架构的需求、业务在线化对数据处理时效性的要求、高IT人工成本体现全托管价值、IT基础设施上云趋势、应用架构弹性化改造趋势、开源趋势是近期数据库产业的六大驱动因素。

2020年中国数据库市场规模达188.43亿元，并将以21.1%的增速在2025年达到682.3亿元

03 技术路径与技术趋势洞察

TiDB是NewSQL时代中第一个开源项目，采用 Shared-Nothing 架构；通过TiKV、TiFlash实现实时HTAP；TiDB支持多云部署，与K8s深度适配，提供TiDB Cloud全托管云数据库即服务，无缝衔接本地与云端体验，成为云原生的、自动驾驶的新一代云数据库。

04 PingCAP业务生态分析

面对混合多云时代的发展大趋势，PingCAP作为纯数据库厂商，以云平台中立性构筑竞争优势。

PingCAP以先进的产品技术为基座和核心驱动力，以开源为长期战略，以混合生态和开源运营连接移动互联网时代红利。TiDB以领先的开源表现处于开源产品的结晶期发展阶段，采用商业开源订阅模式和DBaaS的开源商业模式

目录

◆ PingCAP企业综述	06
• 公司背景信息	07
• 公司团队	08
• 产品与服务	09
• 公司发展历程	10
◆ 数据库行业分析	11
• 政策分析	12
• 数据库产业驱动因素分析	13
• 中国数据库市场环境分析	15
• 开源创新理念的发展	16
◆ PingCAP技术路径与技术趋势洞察	17
• 分布式架构：NewSQL与Shared-nothing	18
• HTAP混合负载：TiKV与TiFlash	19
• 云原生：K8s与TiDB Cloud	20
◆ PingCAP业务生态分析	21
• 云中立属性契合多云部署需求	22
• Ti-Star计划促进开放混合生态释放	23
• 开源商业化模式	24
◆ 名词解释	26
◆ 方法论	27
◆ 法律声明	28

CONTENTS

◆ Summary of PingCAP	---	06
• Background information	---	07
• Team	---	08
• Products and Services	---	09
• History of Development	---	10
◆ Database Industry in China	---	11
• Policy Analysis	---	12
• Driving Factors	---	13
• Market Environment	---	15
• Open-source innovation	---	16
◆ PingCAP Technology Insight	---	17
• Distributed Architecture: NewSQL and Shared-nothing	---	18
• HTAP: TiKV and TiFlash	---	19
• Cloud Native: K8s and TiDB Cloud	---	20
◆ PingCAP Business Ecology Analysis	---	21
• Cloud Neutrality meets the needs of Multi-cloud Deployment	---	22
• Ti-Star Plan promote Hybrid Eco	---	23
• Open-source Commercialization	---	24
◆ Terms	---	26
◆ Methodology	---	27
◆ Legal Statement	---	28

图表目录

▪ PingCAP基本情况介绍	-----	07
▪ PingCAP股权结构及企业关系图	-----	07
▪ 公司研发团队成员介绍	-----	08
▪ PingCAP的产品与服务	-----	09
▪ PingCAP的发展历程	-----	10
▪ 各国或地区数据库产业建设与发展方针	-----	12
• 中国数据库产业部分相关政策	-----	12
▪ 中国与全球年数据产生量，2018年-2025年预测	-----	13
▪ 不同数据库类型的开源数据库流行度占比	-----	14
▪ 中国数据库市场环境梳理	-----	15
▪ 2016-2025年中国数据库市场规模及预测	-----	15
▪ TiDB的NewSQL路径选择	-----	18
▪ Shared-Nothing架构	-----	18
▪ TiDB架构图	-----	19
▪ Self-driving数据库的演进路径	-----	20
▪ TiDB的云上平台架构	-----	22
▪ 桥接社区和企业用户的开放混合生态	-----	23
▪ 面向企业用户的Ti-Star计划	-----	23
▪ PingCAP的开源商业化底层逻辑	-----	24
▪ 2020年中国开源项目活跃度排名	-----	25
▪ 开源社区成熟度曲线图	-----	25
▪ TiDB社区版与商业版的区别	-----	25
▪ TiDB Cloud 收费参考（AWS - 新加坡）	-----	25

Chap 1

PingCAP企业综述

- 公司背景信息
- 公司团队
- 产品与服务
- 公司发展历程

公司背景信息

- PingCAP 成立于 2015 年，是一家专注企业级开源分布式数据库独角兽公司，在产业和学术界受广泛认可，以独特的产品和生态价值，赋能国内外1500+企业数字化

PingCAP基本情况介绍

企业名称	平凯星辰（北京）科技有限公司 PingCAP
成立日期	2015年
具体地址	北京市海淀区西小口路66号中关村东升科技园·北领地C-1楼2层207
注册资本	10,000万美元

来源：企查查，头豹研究院

公司简介

PingCAP是一家企业级开源分布式数据库厂商，秉承开源是基础软件的未来这一理念，PingCAP 持续扩大社区影响力，致力于前沿技术领域的创新实现。其研发的分布式关系型数据库 TiDB 项目，是大数据时代理想的数据库集群和云数据库解决方案。9 个全球分支机构分布于中国、美国、新加坡、日本，服务客户遍布20余个国家，并已被超过1500家不同行业的领先企业应用在实际生产环境，涉及互联网、游戏、金融、政府、电信、制造业等多个领域。

PingCAP行业影响力导览

墨天轮 国产数据库排行 第一	CNCF 2019-2020年贡献 榜排名 全球第六 中国第一	Github 开源项目活跃度 中国第三	入选威斯康辛麦迪逊分校、普渡大学、卡内基梅隆大学课程	国际顶级会议收录三篇前沿学术论文
全新一代HTAP数据库架构 • 为规模化场景和实时业务形态而生	云原生 云中立 • 满足上云下云、跨云需要	开源带来自主 • 确保客户掌握核心数据主权	社区与开放生态 • 不断从社区获得创新数据解决方案，真实场景验证	一栈式数据服务生态 • 简化客户技术栈
10+银行与头部金融企业 • 网联交易、网贷核算 • 互联网业务中台 • 实时风控、反欺诈	三大运营商 • 支付与风控 • 征信与账单系统 • 精准营销	覆盖80%中国互联网独角兽企业 • 万亿级数据毫秒响应 • 支付、秒杀等场景 • 单个集群1000+节点	物流与零售 • 实时统计报表 • 数据中台 • 交易与报表业务隔离	海外头部支付企业与电商 • 在线支付 • 用户与订单管理 • 风控与营销

来源：PingCAP，头豹研究院

公司技术团队

- 在商业化目的的考量之上，秉承对传统 Hacker 精神的贯彻，PingCAP坚定开源，吸引了国内外技术专家加入，研发工程师为主要人员构成，是以开源基因嵌入的产品技术驱动型公司

公司研发团队成员介绍

	职务	姓名	相关介绍
创始人	创始人 CEO	刘奇	创建了 TiDB、TiKV、Codis 等知名开源项目。曾任豌豆荚，京东资深系统架构师。同时也是知名的 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
	联合创始人 CTO	黄东旭	曾就职于微软亚洲研究院，网易有道及豌豆荚，擅长分布式系统以及数据库开发，在分布式存储领域有丰富的经验和独到的见解。狂热的开源爱好者以及开源软件作者，代表作品分布式 Redis 缓存方案 Codis，以及分布式关系型数据库 TiDB
	联合创始人 CFO	崔秋	曾任职于搜狗、豌豆荚，资深广告系统工程师，长期从事广告系统基础组件相关的研究，现主要从事开源 HTAP 数据库 TiDB/TiKV 相关的设计和研发工作
研发技术专家	首席架构师 TiKV负责人	唐刘	分布式数据库专家，知名开源软件 LedisDB / RebornDB / Mixer 作者。曾任金山软件架构师、WPS云服务后台架构师
	TiKV工程师	刘聪	曾任金山软件架构师、金山快盘创始团队核心成员
	TiKV工程师	Brian Anderson	Rust编程语言联合创始人，曾任Mozilla研究员，Firefox新一代内核Servo主要开发者
	TiKV工程师	Nicholas Cameron	Rust编程语言核心团队成員，曾任Mozilla研究院
	云产品工程师	Greg Weber	Haskell编程语言社区知名开发者
	TiKV工程师	YiWu	曾任Facebook、Yahoo软件工程师，RocksDB维护者、BlobDB开发者
	数据库内核工程师	Liquan Pei	曾任Pinterest ads infra tech lead、Confluent、Pivotal工程师
	Realtime Analytics Team Manager	马晓宇	曾任QuantCast SQL tech lead、网易大数据infra tech lead
	Infra工程师	黄梦龙	曾任畅游、华清飞扬、光合互娱主程序工程师
技术部门	R&D Group Dept. 1	由专注实时分析及HTAP场景产品开发的Real-Time Analytics Team（原AP团队）、TiDB团队以及负责产品规划以及管理的PM团队共同组成，专注TiDB产品研发	
	R&D Group Dept. 2	由专注于分布式存储层构建的Storage Team（原TiKV团队）、专注于分布式数据库架构领域的Arch Team、专注于分布式系统整体性能的Scheduling Team以及专注效率工具与生态工具的Ecosystem Tools团队共同组成，专注提升工程效率，构建分布式系统生态	
	R&D Group Dept. 3	由Cloud团队、效率至上EE（Efficiency Engineering）团队、QA团队、前段团队、UI设计团队共同组成，致力于构建高质量、易于使用的TiDB云产品，为在线生产系统的稳定性保驾护航，并不断提高产品质量	

来源：PingCAP、LinkedIn，头豹研究院

出身研发一线且具备国际视野的专业团队

三位创始人都是中国互联网公司研发一线出身的技术大咖，同时也从事过排名靠前的开源项目Codis Redis。创始人的技术水平和眼界始终在国内甚至在全球领先，对开源理念笃定并且有着开源领域一流的实践。

PingCAP的团队成員基本都有丰富的互联网基础架构大数据、以及开源项目的经验，同时作为专注新一代分布式和分布式存储的技术团队吸引了一众海外顶级工程师加入。Rust核心团队前成员Brian Anderson和Nick Cameron加入了以Rust作为开发编程语言的PingCAP TiKV，极大的提升了TiKV和PingCAP在Rust社区中的影响力，进一步推动中国Rust社区的发展。

产品与服务

- PingCAP是一家企业级开源软件服务供应商，提供包括开源分布式数据库产品、解决方案及相关的咨询、支持与培训服务

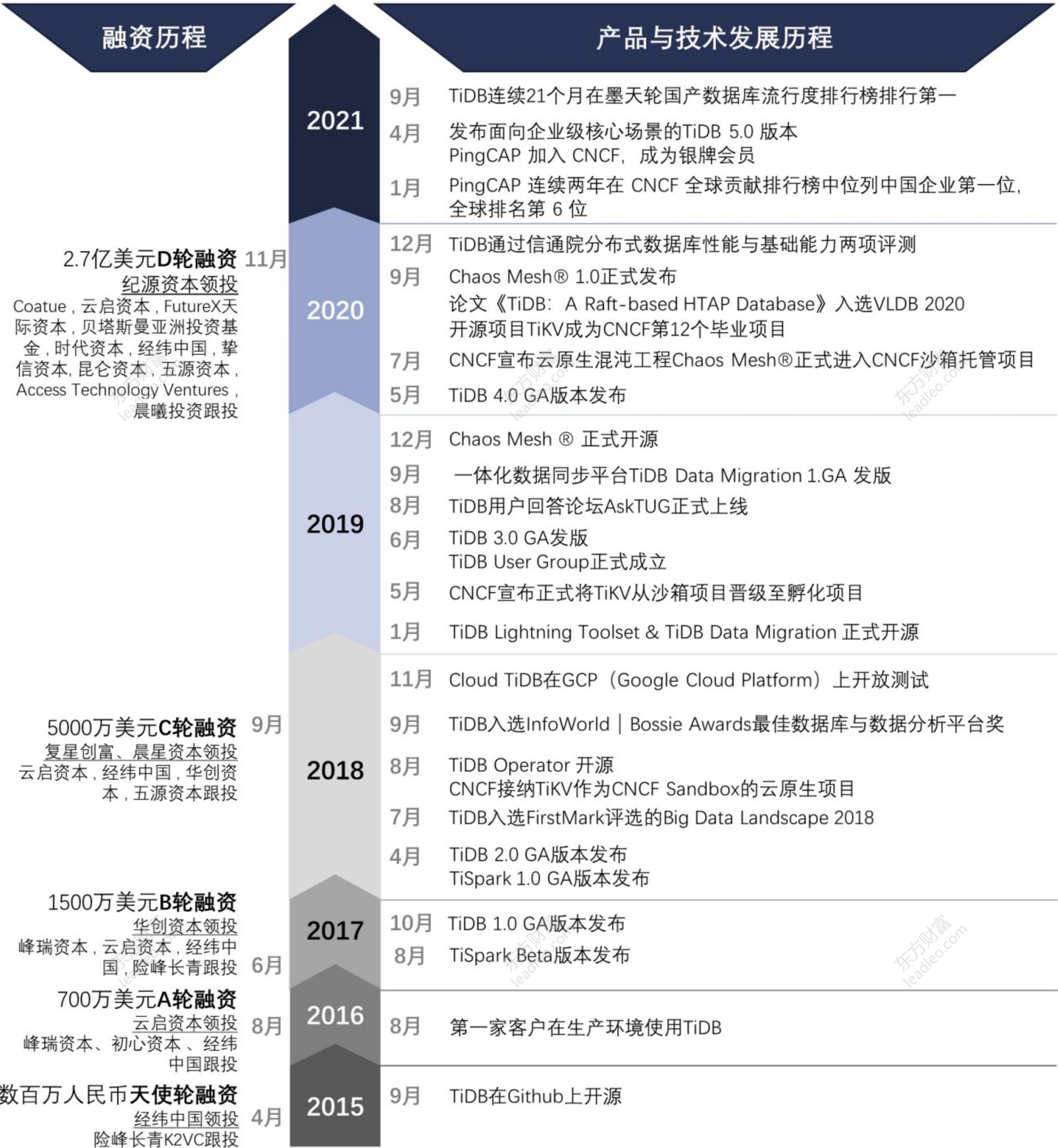
PingCAP的产品与服务



公司发展历程

- PingCAP保持对TiDB每年完成一次迭代进化的节奏，坐拥TiKV与Chaos Mesh®两个CNCf项目，具备全球先进的工业级分布式架构理论实践。不断获得资本加持，D轮以2.7亿美元的融资规模刷新全球数据库赛道的融资历史

PingCAP的发展历程



来源：PingCAP、企查查，头豹研究院

Chap 2

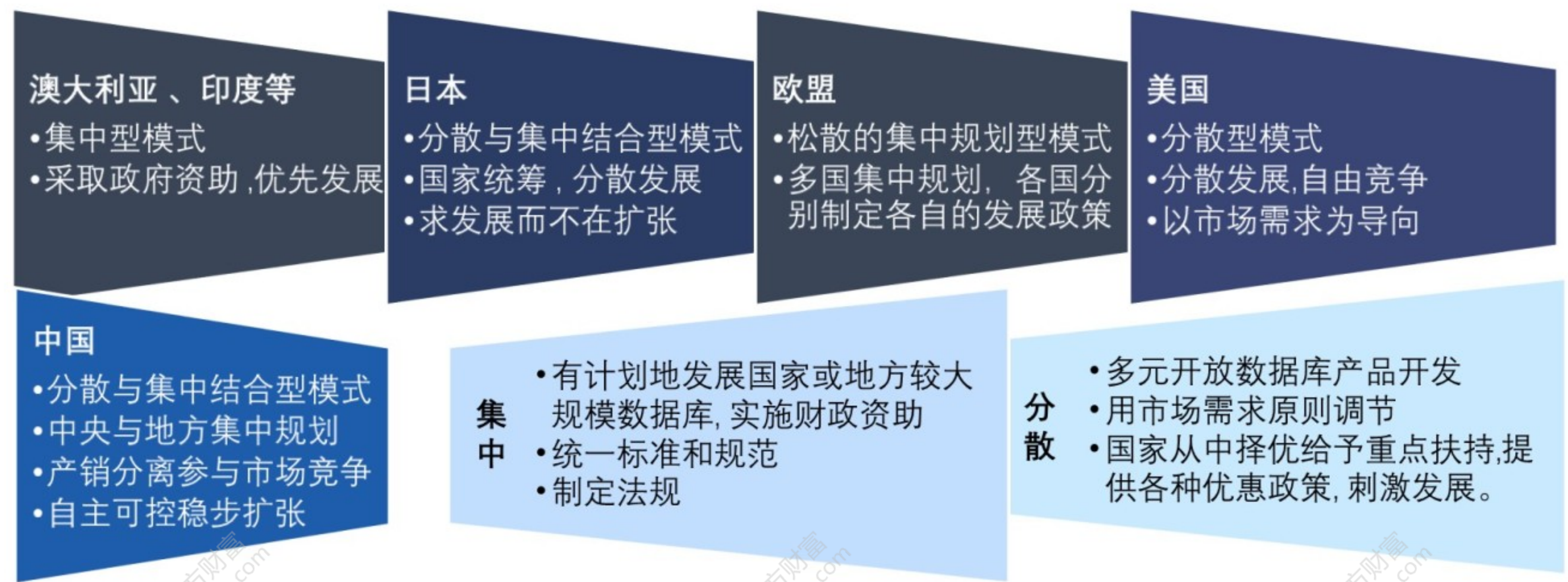
数据库行业分析

- 政策分析
- 数据库产业驱动因素分析
- 中国数据库市场环境分析
- 开源创新理念的发展

政策分析

- 中国数据库产业政策制定采取分散与集中结合型模式，具体政策是实现宏观政策目标的手段和措施,包括产权保护政策、需求引导政策、安全保密政策、鼓励开发政策、经营政策、国际合作交流政策、人才政策等

各国或地区数据库产业建设与发展方针



来源：中共哈尔滨市委党校、北京大学，头豹研究院

中国数据库产业部分相关政策

层面	发布机构/部门	时间	政策	关键内容
中央	全国人大	2021.06	《数据安全法》	规范数据处理活动，保障数据安全，促进数据开发利用，保护个人、组织的合法权益，维护国家主权、安全和发展利益
	国务院	2020.08	《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》	深化产业国际合作，推动国产替代进程
	发改委、科技部、财政部等	2020.09	《关于扩大战略性新兴产业投资培育壮大新增长点增长级的指导意见》	5G建设投资、关键芯片、关键软件等技术攻关
	工信部	2017.01	《大数据产业发展规划（2016-2020年）》	以强化大数据产业创新发展能力为核心，推动促进数据开放与共享、加强技术产品研发、深化应用创新
	国家发改委	2016.01	《关于组织实施促进大数据发展重大工程的通知》	以政府将重点支持大数据示范应用、共享开放、基础设施统筹发展，以及数据要素流通。国家发改委将择优推荐项目进入国家重大建设项目库审核区，并根据资金总体情况予以支持。
地方	天津市	2021.05	《天津市知识产权“十四五”规划》	率先提出提升信创产业知识产权质量
	广州市	2020.10	《广州市黄埔区、广州开发区促进信息技术应用创新产业发展》	全国首个区县级信创专项扶持项目，提出10条信创相关的奖励与补贴政策
	广西省	2020.10	《加快广西数字经济发展若干措施的通知》	对进入国家级信创产品库的企业落户在广西并在广西生产的予以分类奖补
	湖南省	2020.10	《湖南省区块链发展总体规划（2020-2025年）》	推动区块链与核高基相关产业协同发展 促进区块链和网络安全产业协同发展
	河北省	2020.07	《河北省电子信息产业重点攻坚行动计划（2020-2022年）》	每年组织实施10项重大科技攻关项目，部分技术达到国内领先国际领先
	贵州省	2020.06	《贵州省大数据融合创新发展工程专项行动方案》	打造大数据信息技术应用创新生态中心为相关数据库、中间件、应用软件、信息安全等企业提供服务
	浙江省	2020.03	《2020年浙江省软件与集成电路产业工作要点》	推动国产信息设备及适配软硬件快速迭代升级，全面提升信息技术自主可控核心能力

来源：全国人大常委会、发改委、国务院、工信部、各省市厅，头豹研究院

■ 数据库产业驱动因素分析

- 企业数据规模激增对分布式架构的需求、业务在线化对数据处理时效性的要求、高IT人工成本体现全托管价值、IT基础设施上云趋势、应用架构弹性化改造趋势、开源趋势是近期数据库产业的六大驱动因素

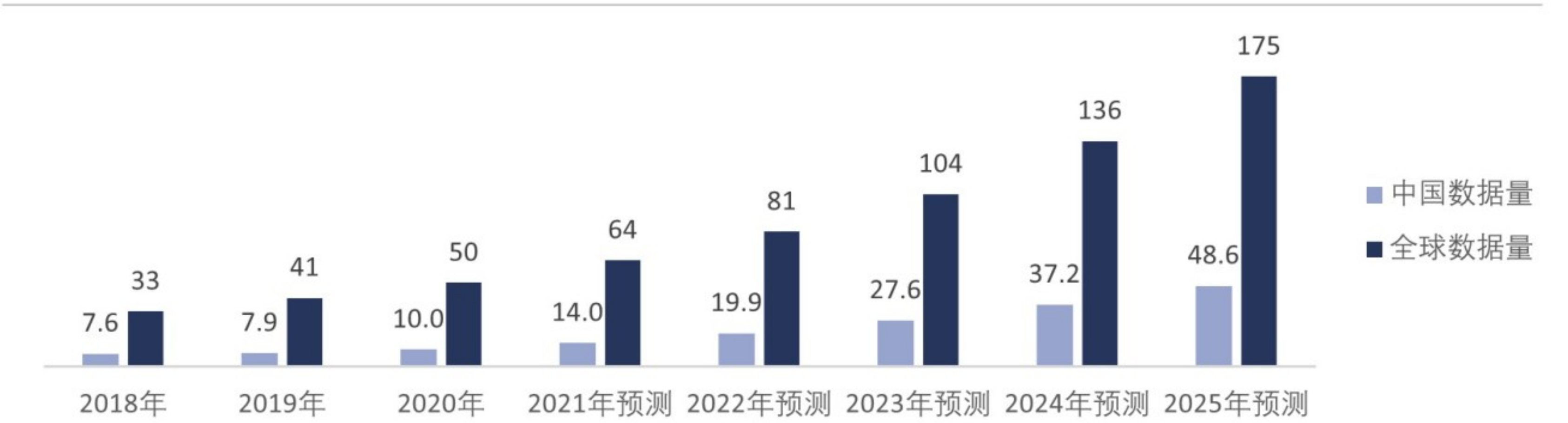
自上世纪70年代到本世纪初的30年间，数据库领域的产品形态与主流厂商几乎没有变革，停留在被Oracle、SQL Server、IBM DB2割据的形态。但是在移动互联网的普及之后，以下六点因素驱动了数据库产业发生巨大变革：

□ 因素一：企业数据规模激增，中心化架构不堪重负

我们正在目睹数据的爆炸式增长。中国“数据圈”将从 2018 年的7.6ZB增长到2025年的48.6ZB。这里的一个关键驱动因素将是物联网设备，预计到 2025 年，物联网设备创造的数据规模将占全球数据约一半。

伴随着这种数据量爆炸式增长的是数据种类的扩展，包括具有不同结构的数据以及新的数据模式，例如时间序列数据。于是，具备应对海量且非结构化数据的分布式架构NewSQL将逐渐走到数据库选型舞台的中心。

中国与全球年数据产生量，2018年-2025年预测 【泽字节ZB】



来源：William Blair、希捷，头豹研究院

□ 因素二：企业业务在线化，数据处理时效性要求大大提升

当今的经济在很大程度上依赖于数据，随着公司在其供应链的每个步骤中捕获、编目并尝试从数据中收集洞察力；企业收集大量客户数据以提供更高水平的个性化；政府为执法和情报目的收集和使用大量数据；消费者将社交媒体、娱乐、云存储和实时个性化服务融入他们的生活。

虽然这种数据繁荣带来了非凡的用户体验和商机，但它也带来了数据存储和访问方面前所未有的挑战，其中最为重要的是数据处理时效性。OLAP、HTAP的对分析效率的优势凸显。

因素三：IT人工成本逐年提升，全托管的价值日渐显现

随着支撑业务的IT系统越来越复杂，运维成本也水涨船高，普遍采用本地部署的公司运维成本中，软硬件和网络设备占30%，日常维护服务占30%，而运维人力则占了40%。人力成本包括现场维护、训练教育、人员流失、招聘成本等，而不完善的权责规范化管理、流程分级管理和量化绩效管理将把人力成本放大，将运维效率拖入泥潭。

而采用来自IT服务商的基础设施全托管服务比如云部署的IaaS服务模式，则可以发挥专业化分工的经济优势，解决IT运维人力成本困境。

因素四：IT基础设施上云，在云上处理数据成为必然

云数据库提供天然的灵活性、经济高效的部署方式、按需付费的支付模式。在存算分离理念的基础上，云化服务完美应对了弹性扩展、功能迭代、成本控制等特性需求，在资源需求差异化的场景中实现资源的合理配置，让基础设施管理就是代码调用(Infrastructure as a code)，将企业IT人员从繁复的基础设施运维中解放出来。

云计算进入工业化应用后，如今的容器技术与Serverless的融合跨代际得实现了计算密度的升级，提供了最佳的性能功耗比和性能价格比，大幅降低了用户使用门槛的同时吸引各行各业的IT团队对其上层应用微服务化以享受云计算带来的技术红利。

因素五：应用架构往弹性化发展，数据平台也需要相应的变化

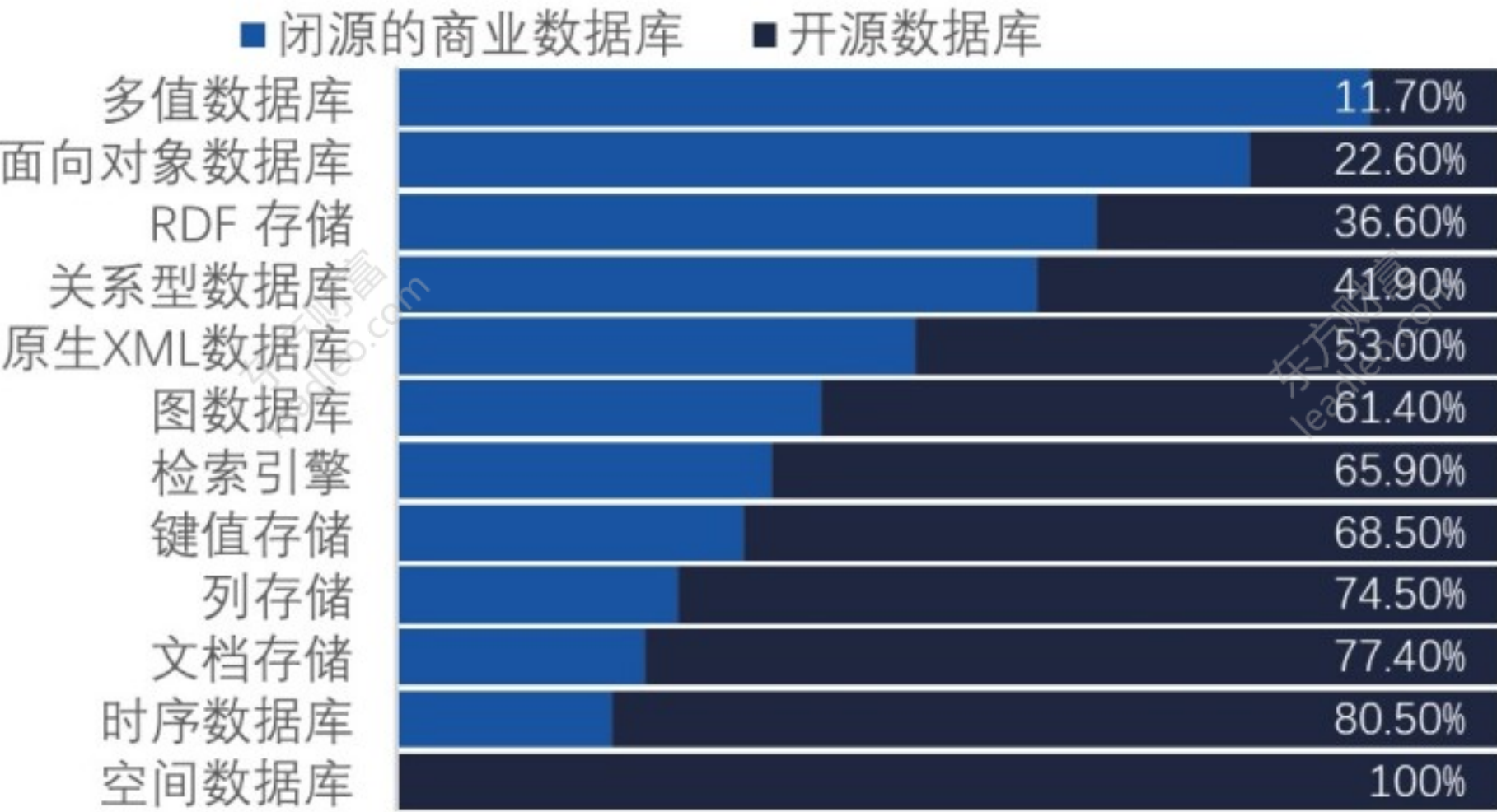
一方面，企业级SaaS服务和云端迁移的高速发展正是微服务架构普及的重要驱动因素。另一方面，微服务架构作为高效成熟的云计算相关技术，推动了企业上云的进程。

因素六：开源重塑产品迭代模式与市场推广模式

开源数据库是社区数据库，免费对外开放源代码，开发人员可以在开源代码的基础上修改或使用。相较于闭源的商业数据库，开源数据库以较低的成本、丰富的产品类型和活跃的社区支持，在包括文档存储、键值存储、图数据库等领域上已经成为了市场的主流产品。

随着开源数据库产品及配套工具和社区的流行度逐年上升，开源数据库将持续迭代，加速升级，吸引更多企业参与，帮助企业释放数字潜力。

不同数据库类型的开源数据库流行度占比

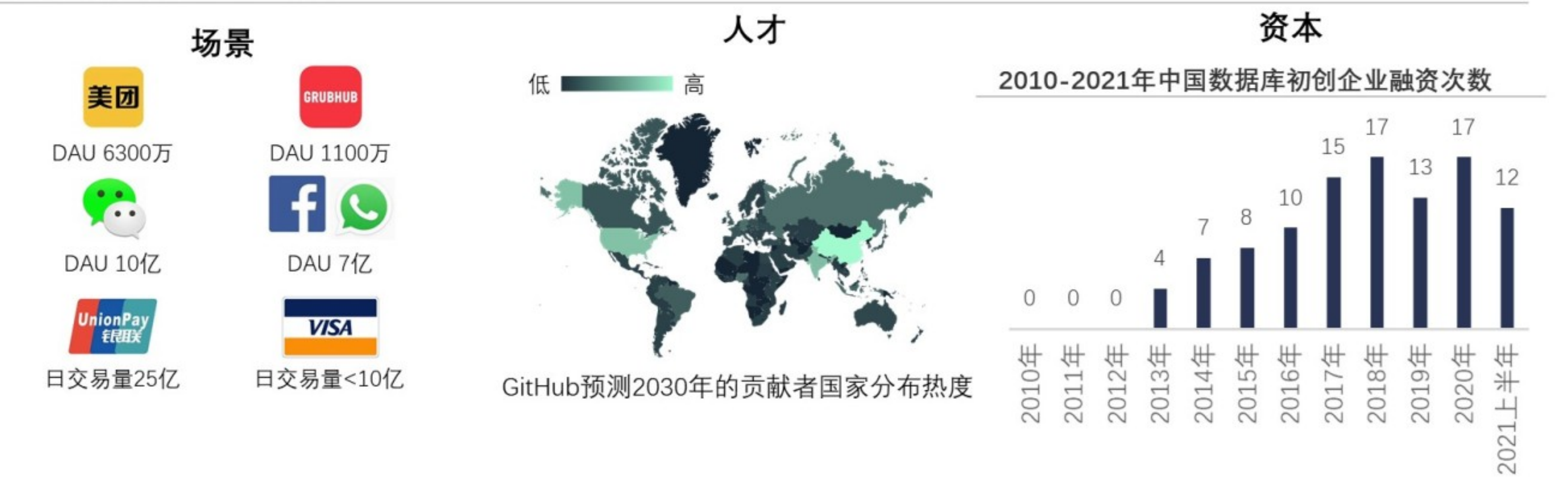


来源：DB-Engines, 头豹研究院

中国数据库市场环境分析

- 中国数据库市场具备全球最大、最复杂的应用场景，同时吸引了高热度的人才与资本。2020年中国数据库市场规模达188.43亿元，并将以21.1%的增速在2025年达到682.3亿元

中国数据库市场环境梳理



来源：明势资本、Github、中国信通院，头豹研究院

全球最大、最复杂的应用场景在中国

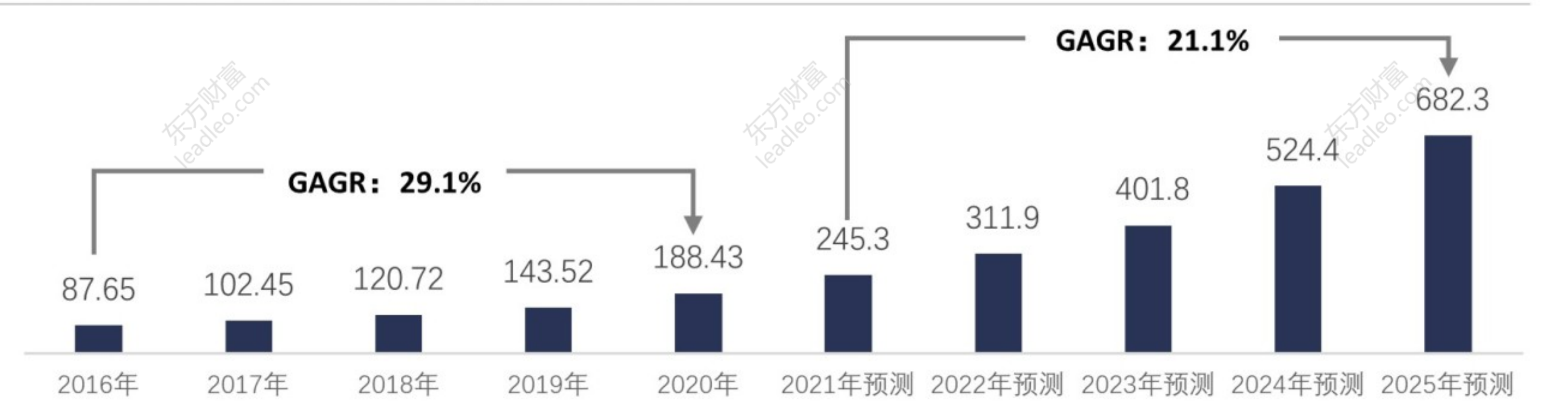
中国市场具备全球最大的数据流量规模，中国的IT相关人才热度将领先全球，吸引着全球资本市场对中国软件市场持续注资，进一步催化中国数据库市场的高速增长。

中国数据库市场规模

市场增量将主要来源于1. 未开发部署的中小型的业客户； 2.云数据库的普及； 3. 分布式架构数据库替代； 4. 实时数据处理系统上线； 5. 全托管、自动伸缩的数据库即服务； 6. 开源数据库商业化

2020年中国数据库市场规模达188.43亿元，过去五年的复合年均增长率（CAGR 2016-2020年）为29.1%。2021-2025年中国数据库市场将维持快速增长的趋势，年均复合增长率（CAGR 2021e-2025e）达21.1%，将在2025年达到682.3亿元的市场规模。

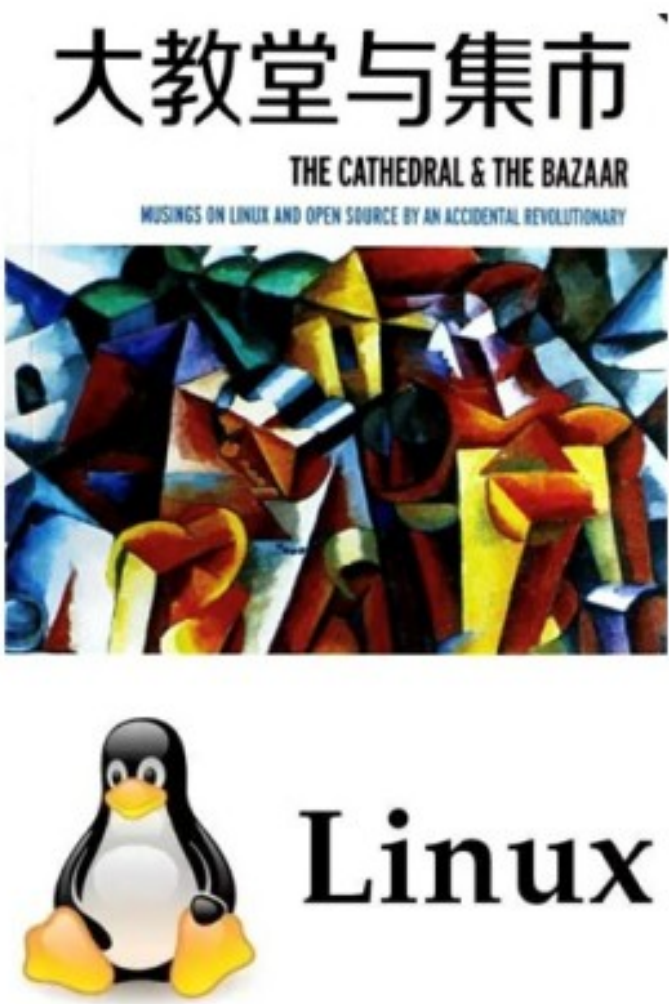
2016-2025年中国数据库市场规模及预测【亿元】



来源：中国信通院、国家统计局、Scale Grid，头豹研究院

开源创新理念的发展

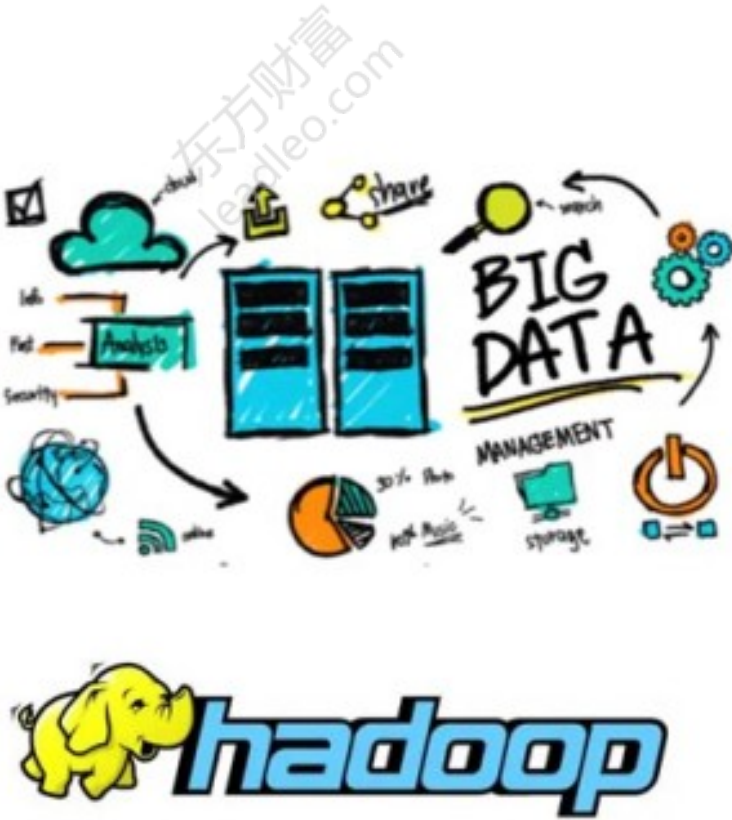
- 开源理念的发展经历了以Linux为代表的自由开发者运动，以Hadoop为代表的技术栈驱动社区，PingCAP以开源为坚定的公司战略，推动极致场景驱动社区协作为第三代开源



第一代开源：自由开发者运动

在PC时代，开源旗手Eric Raymond观察到传统软件开发采用的大教堂模式是封闭的、垂直的、集中式的开发模式，反映一种由权利关系所预先控制的层级制度；而与传统模式截然不同是集市模式，是并行的、点对点的、动态的多人协同开发模式，开发者之间通常仅仅靠互联网联系，在这种貌似混乱而无序的开发环境中，居然产生了质量极高和极具效率与生命力的软件，产生了诸如Linux这种世界级的操作系统。

Eric Raymond认为，未来会更多地属于用修建集市的方式造出一所大教堂的人们。开放式的文化会最终胜利，因为开放式合作可以在一个问题上投入多几个数量级的技术工时，封闭的世界无法赢得这样的竞争。



第二代开源：技术栈驱动社区

在移动互联网时代的业务需求对更新速度的要求提高了一个层次，在大数据领域尤其是分布式系统的开发过程中，由于业务的拆分，被迫会衍生出更多的分布式需求以及应对这些需求的技术，逐渐形成了像Hadoop、Spark、Spring Cloud这样的技术栈社区，例如Hadoop旗下又包含了Hive、MapReduce、HDFS等技术以实现不同的职能。

相比于重复造轮子，从技术栈驱动社区中调用技术框架于开发进程，将集市模式的商品从零件进阶为模块，进一步解放了开源项目的创新潜力与效率。而公司通过分享开源贡献，让社区代码不断优化迭代的同时，也能保证上线业务得到领先行业的底层实现逻辑。



第三代开源：极致场景驱动社区协作

在云原生与多云时代，业务与技术相互更加紧密，迭代速度更快、增长率更好、更代表未来的方案将是竞争力王牌。“场景驱动开发”执行于一款流行的开源软件之上，产品“触感”源于对普遍需求的准确把握，高流行度代表着高用户基数与多场景触达，经过一轮又一轮的用户反馈、bug修改、迭代建议后实现正向的迭代循环，覆盖更广的应用场景。

PingCAP运营TiDB社区的理念正是如此，坚定地认为开源是当前全球化背景下基础软件领域成功的有效路径。

来源：PingCAP，头豹研究院

Chap 3

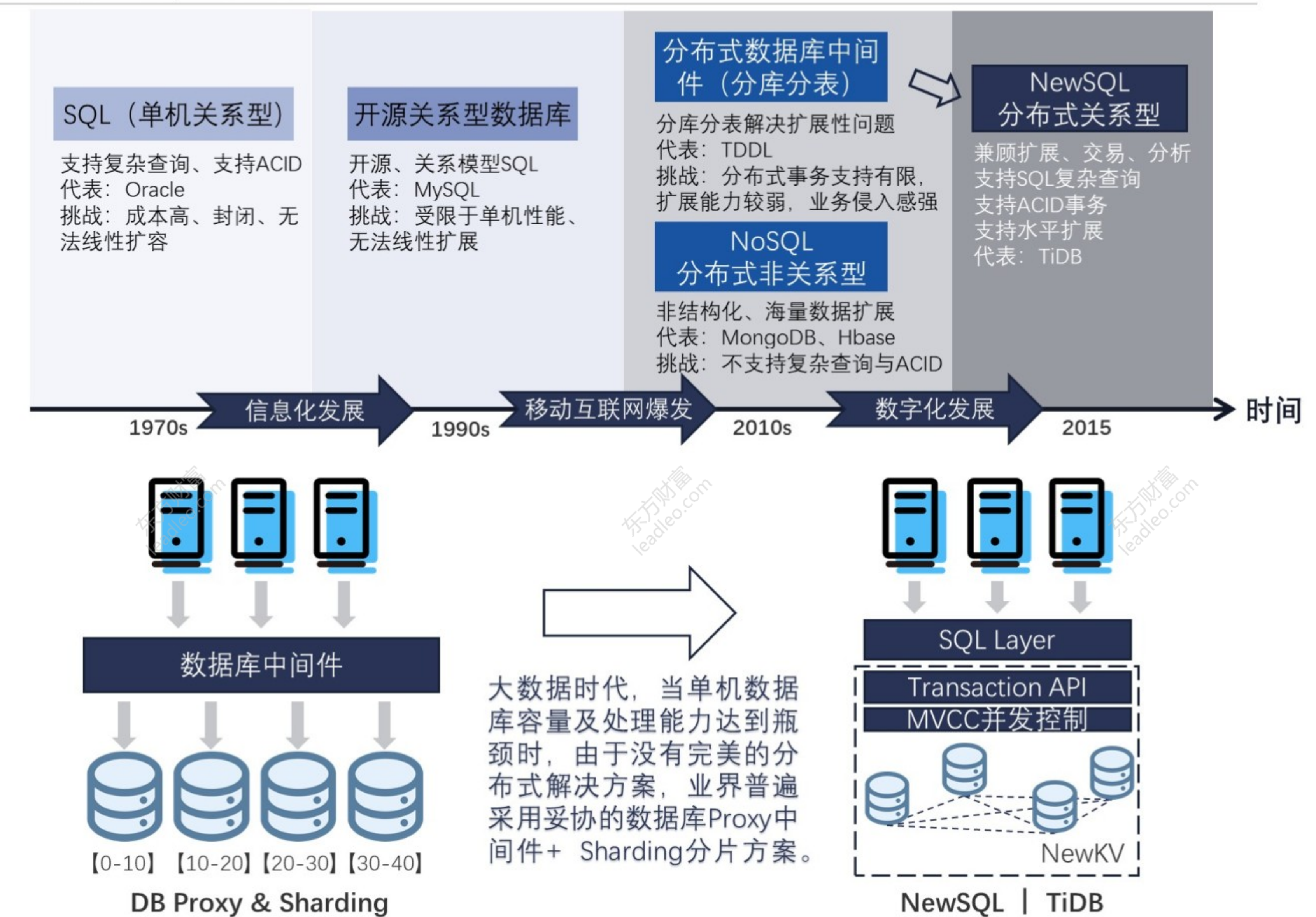
PingCAP技术路径与技术趋势洞察

- ❑ 分布式架构：NewSQL与Shared-nothing
- ❑ HTAP混合负载：TiKV与TiFlash
- ❑ 云原生：K8s与TiDB Cloud

分布式架构：NewSQL与Shared-Nothing

- NewSQL完美解决了MySQL的扩展问题，TiDB是NewSQL时代中第一个开源项目，采用Shared-Nothing架构，具备高可用性、带宽要求低、平滑扩展能力与远程部署能力

TiDB的NewSQL路径选择



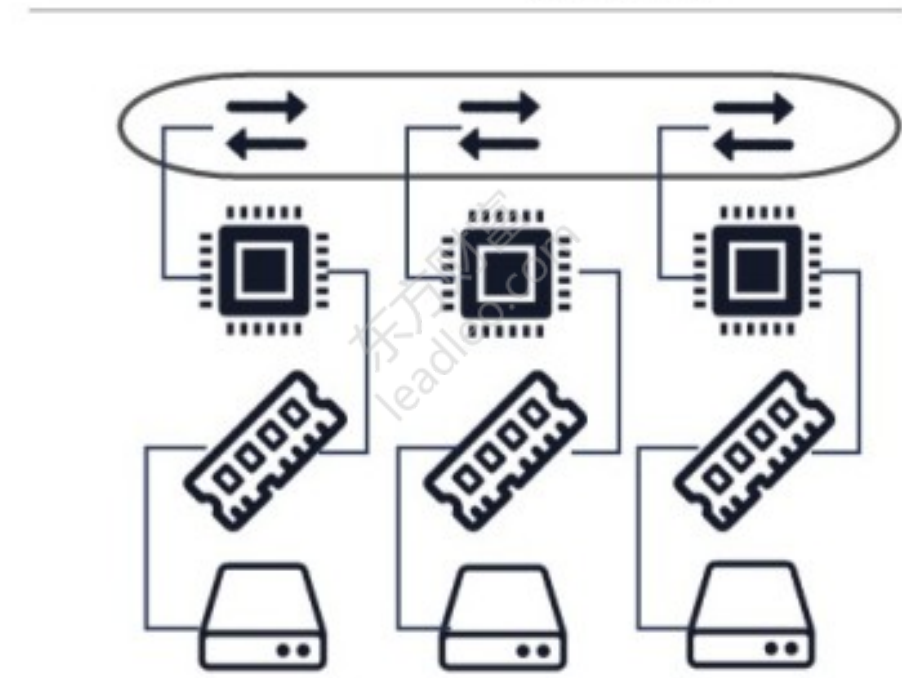
来源: PingCAP, 头豹研究院

NewSQL与Shared-Nothing

过去的系统大量依赖了 MySQL 分库分表 的技术, 运维和使用受限已不能满足时代需要。2012年, Google Spanner 和 Google F1 的论文创造性地提出了将事务型数据库的ACID保证与NoSQL的可扩展性与高性能结合的 NewSQL 原型。2015年, TiDB成为了NewSQL时代中第一个在社区中尝试开源实践的项目。

TiDB采用了 Shared-Nothing 架构, 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, 使每一个节点都是独立、自给的。可以通过简单地增加节点, 使数据库做到无限的扩展在TiDB 存算分离 的架构设计下, 可对资源分别在线扩缩容, 大大降低了运营成本。

Shared-Nothing架构

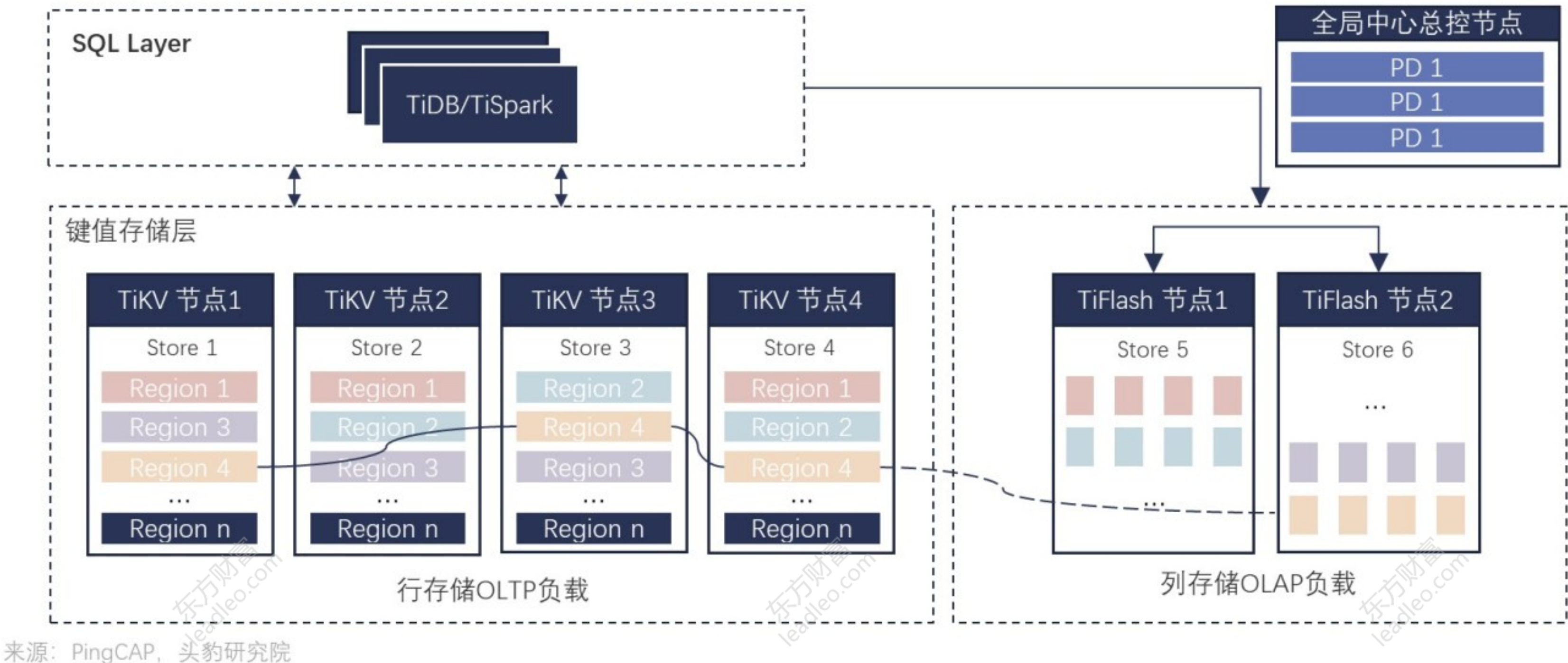


来源: 头豹研究院

HTAP混合负载：TiKV与TiFlash

- TiDB通过TiKV、TiFlash实现实时HTAP，解决了数据库面对海量数据库下的存储与并发处理的扩展性问题，实现在线高并发OLTP及OLAP海量数据分析

TiDB架构图



来源：PingCAP，头豹研究院

实时HTAP混合负载

PingCAP发布的论文《TiDB: A Raft-based HTAP Database》成为业界第一篇实时HTAP分布式数据库工业实现的顶级论文。

提供行存储引擎TiKV、列存储引擎TiFlash两款存储引擎，TiFlash通过Multi-Raft Learner协议实时从TiKV复制数据，确保行存储引擎TiKV和列存储引擎TiFlash之间的数据强一致。TiKV、TiFlash可按需部署在不同的机器，解决资源隔离的问题。

TiKV高性能键值存储引擎

TiKV 是一个分布式事务型的键值数据库，提供了满足 ACID 约束的分布式事务接口，并且通过 Raft 协议保证了多副本数据一致性以及高可用。在RocksDB的底层之上定制Titan引擎插件，将 value 从 LSM-tree 中分离出来单独存储，以降低写放大。通过协处理器 (Coprocessor) 可以为 TiDB 分担一部分计算，实现计算加速。

TiFlash高性能列式分析引擎

TiFlash是TiDB的扩展分析引擎，核心基于列存储引擎和向量化计算引擎，通过Raft Learner副本技术同步数据，不影响TiDB集群的OLTP交易，提供强一致数据读取。在v5.0版本中，TiFlash 引入MPP架构，可以将查询进一步下推到MPP集群，大幅缩短分析查询的执行实践。

云原生：K8s与TiDB Cloud

- TiDB支持多云部署，与K8s深度适配，提供TiDB Cloud全托管云数据库即服务，无缝衔接本地与云端体验，成为云原生的、自动驾驶的新一代云数据库

全面拥抱Kubernetes

随着 Kubernetes（K8s）的全面成熟，越来越多的组织开始大规模地基于 K8s 构建基础设施层。在 K8s 上运行 TiDB 不仅能实现企业技术栈的统一，降低维护成本，还能带来更高的可用性与安全性。TiDB与K8s深度耦合，适配Rancher、Openshift等主流K8s平台。

通过 TiDB Operator 可在公有云、私有云、混合云中实现部署工具化、自动化，使TiDB成为云原生技术栈数据库首选。

Chaos Mesh®云原生混沌工程平台

而在K8s上运行TiDB的稳定性是用户最重视的问题，TiDB 本身能够适应动态化的 K8S 环境。起源于TiDB的核心测试平台，PingCAP发布Chaos Mesh®解决分布式系统稳定性挑战，专注K8s平台，而不限于数据库领域。

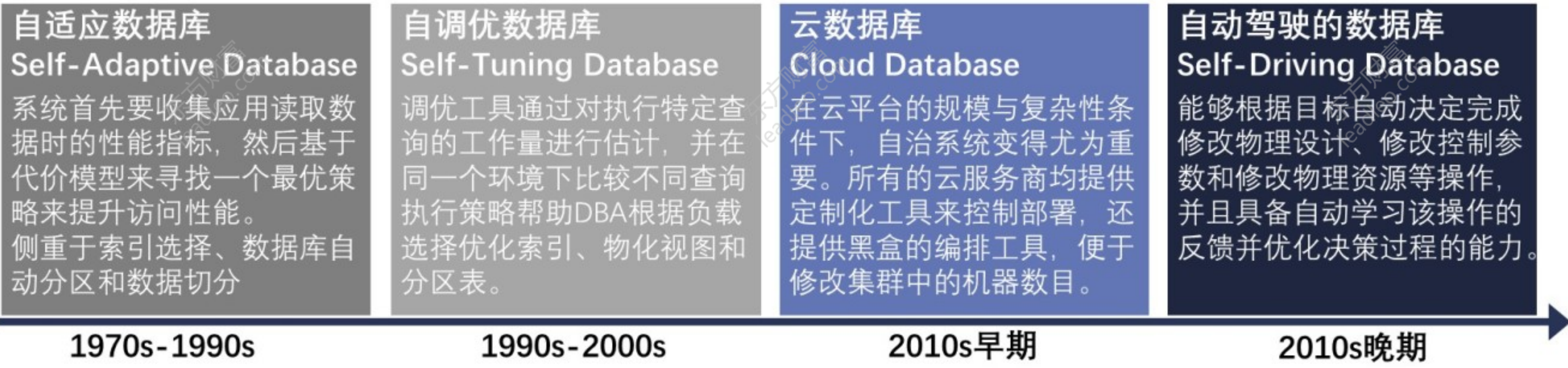
TiDB Cloud全托管的数据库即服务（DBaaS）

NewSQL数据库即服务平台，TiDB Cloud屏蔽TiDB数据库部署、运维和性能调优的复杂性，实现快速创建和管理TiDB实例。

结合云服务、云原生，将云作为基础设施，重构数据库API化、服务化，实现秒级资源供给、动态启动计算单元。全托管TiDB Cloud按需自动弹性伸缩，实现效益/成本的最优化，本质上是分布式的Self-Driving。

支持用户选择在不同云供应商上灵活迁移；借助VPC实现租户之间的安全隔离满足企业级安全；同时PingCAP原厂工程师提供专家级服务支持。

Self-driving数据库的演进路径



来源：CMU、Andy Pavlo，头豹研究院

Chap 4

PingCAP业务生态分析

- 云中立属性契合多云部署需求
- Ti-Star计划促进开放混合生态释放
- 开源商业化模式
- 行业案例梳理

■ 云中立属性契合多云部署需求

- 面对混合多云时代的发展大趋势，PingCAP作为纯数据库厂商，以云平台中立性构筑竞争优势。TiDB支持多云跨云部署策略，TiDB Cloud更是会成为PingCAP提升市场竞争力的主打产品



来源：PingCAP，头豹研究院

□ 多云部署需求

企业级用户在未来的多云场景中，选择多云策略可避免与某一个云厂商深度绑定，分散风险，保持议价能力，降低数据库的迁移粘性，充分释放云原生的灵活性。

□ 云中立的品牌方案

TiDB兼容MySQL协议生态，应用无需或者修改少量代码即可从MySQL迁移到TiDB，提供一系列数据迁移工具和迁移方案，具备高迁移便捷性。

几乎所有公有云厂商都以K8s作为一个公有云服务标准，TiDB与K8s深度耦合，具备高度统一的云原生设计，支持公有云、私有云和混合云部署，用户可以采用多云、跨云部署策略，让数据库在多个云上运行。

面对混合多云时代的发展大趋势，PingCAP作为纯数据库厂商，以云平台中立性构筑竞争优势，TiDB Cloud更是会成为PingCAP提升市场竞争力的主打产品，TiDB Cloud现已支持亚马逊AWS、谷歌GCP，且计划未来支持更多云平台。

Ti-Star计划促进开放混合生态释放

- TiDB开放混合生态桥接着社区和企业用户，Ti-Star计划覆盖了完整的企业客户生命周期，以更接地气的场景驱动推动着TiDB紧贴业务需求持续演进，催化极致场景驱动社区协作

桥接社区和企业用户的开放混合生态



面向企业用户的Ti-Star计划



混合生态

TiDB独特的开放混合生态包含了开源社区的Contributors、应用开发者、第三方生态伙伴以及企业用户，其中企业用户的体验在混合生态中占据着战略地位。

早期，PingCAP以首批的中国互联网用户作为启动社区的场景；第二步，以互联网客户经验的产品外推到企业级用户的数字化场景中打磨；第三步，将中国的互联网场景外推向海外互联网用户。初步实现场景驱动的社区协作。

Ti-Star合作共创计划

Ti-Star计划覆盖了完整的客户生命周期，PingCAP在接触、应用、团队建设和实践推广的每一步都提供原厂深度支持和持续的互动。端到端打造TiDB 用户成功案例及开源合作创新明星。同时，更接地气的场景驱动推动着TiDB紧贴业务需求持续演进，催化极致场景驱动社区协作。

来源：PingCAP, 头豹研究院

■ 开源商业化模式

- PingCAP以先进的产品技术为基座和核心驱动力，以开源为长期战略，以混合生态和开源运营连接移动互联网时代红利。TiDB以领先的开源表现处于开源产品的结晶期发展阶段，采用商业开源订阅模式和DBaaS的开源商业模式

PingCAP的开源商业化底层逻辑



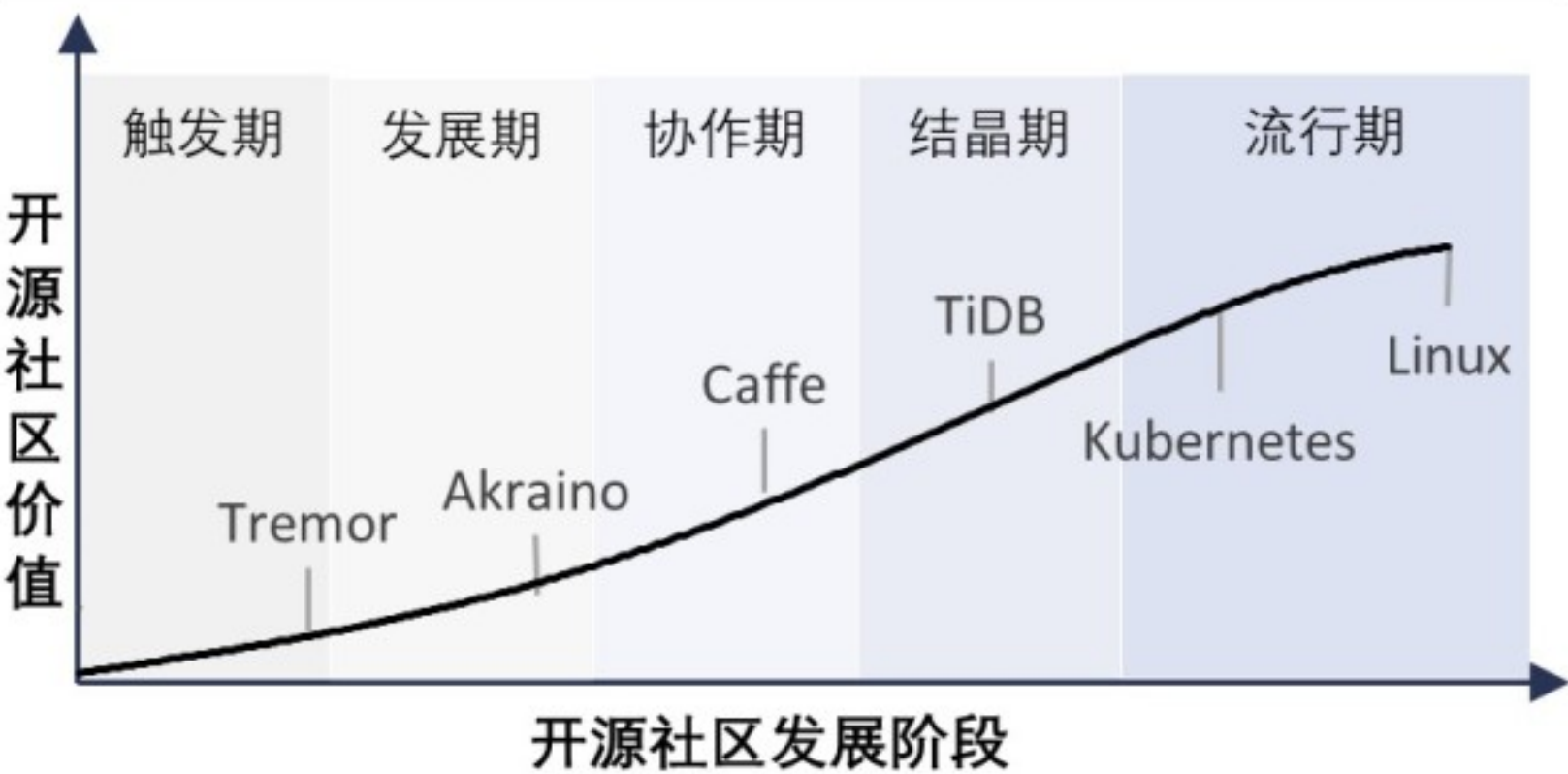
来源：PingCAP，头豹研究院

2020年中国开源项目活跃度排名

排名	项目名称	活跃度	排名	公司名称	活跃度
3	pingcap/tidb	4339.4	1	Alibaba	58353.73
11	tikv/tikv	2184.8	2	Baidu	24868.17
15	pingcap/docs-on	1915.3	3	Pingcap	22113.01
23	pingcap/pd	1481.4	4	Tencent	11285.7
26	pingcap/docs	1262.6			
31	pingcap/tidb-operator	1139.6			

来源：Github、开源社，头豹研究院

开源社区成熟度曲线图



来源：中国信通院、PingCAP，头豹研究院

TiDB社区版与商业版的区别

功能	参数名称	社区版	商业版
核心组件	SQL引擎、存储引擎、调度器	✓	✓
分析引擎	TiSpark、TiFlash	✓	✓
高可用	异地灾备（主从异步）	✓	✓
	多中心强一致多活	✗	✓
安全	日志审计、存储加密、通信加密	✓	✓
	安全审计扩展模块、白名单控制	✗	✓
运维管理	自动化部署、监控告警、备份/恢复	✓	✓
数据交换	数据CDC迁移同步工具等	✓	✓
支持服务	7*24产品故障响应	✗	✓
	规划/实施/主动式巡检/故障协查/知识转移/重要时期保障	✗	✓

来源：PingCAP，头豹研究院

TiDB Cloud 收费参考（AWS - 新加坡）

档位	节点	CPU	存储	每小时	每月
T3.Standard	TiKV	8 vCPU	1024 GiB	\$2.183/hr	\$1571.76
	TiDB	8 vCPU	-	\$0.98	\$705.6
H3.Standard	TiKV	8 vCPU	1024 GiB	\$2.183	\$1571.76
	TiDB	8 vCPU	-	\$0.98	\$705.6
	TiFlash	8 vCPU	1024 GiB	\$2.26	\$1627.2

来源：PingCAP，头豹研究院

□ PingCAP的开源表现领先

根据开源社统计，2020年PingCAP的分布式关系型数据库 pingcap/tidb，分布式事务型的键值数据库 tikv/tikv，文档项目 pingcap/docs-on和pingcap/docs 等六个项目进入Top50。

PingCAP位居中国科技公司活跃度排名第三位。

□ TiDB社区处于结晶期发展阶段

根据中国信通院的开源社区成熟度曲线模型，TiDB已进入结晶期发展阶段。在该阶段，TiDB社区已具备充分的用户案例和完善的贡献机制。通过社区内开源治理与运营体系，和对潜在客户开发和销售拓展，进入可持续发展的“开放-连接-融合”的良性商业循环。

□ 商业版开源订阅模式

开源软件订阅模式，指核心功能完全开源，基础版本免费，但是对额外功能或升级版本以及部分支持服务需要通过订阅获得。采用这种模式的开源软件公司包括Confluent、Kafka和Elasticsearch。

TiDB的开源社区版本免费，核心功能与商业版保持一致，用户可以通过订阅获得PingCAP原厂的商业保障及服务及额外的安全、高可用组件。

□ DBaaS全托管租赁模式

DBaaS数据库即服务，将数据库软件部署在云端，客户向云厂商支付计算和存储资源费用，向DBaaS厂商支付产品订阅费。采用这种模式的公司包括AWS Aurora和Snowflake。

TiDB用户可以选择TiDB Cloud，获得开箱即用的TiDB服务，屏蔽TiDB数据库部署、运维和性能调优的复杂性。

名词解释

- ◆ **云原生：**区别于从本地环境移植到云上的大部分程序，云原生强调最初的开发就是为了最终部署到云环境上。在公有云、私有云和混合云等新型动态环境中，赋能组织或企业去构建和部署可弹性扩展的应用。
- ◆ **Infra：**Infrastructure基础设施层，包括服务器和网络。
- ◆ **异地容灾：**在不同的地域，构建一套或者多套相同的应用或者数据库，起到灾难后立刻接管的作用。异地容灾对企业应用及数据库起到了安全性、业务连续性等方面的作用，因此他与私有云或者公共云起到的容灾结果是完全不同的。
- ◆ **行存储：**Row-based storage, 数据是按照行数据为基础逻辑存储单元进行存储的，一行中的数据在存储介质中以连续存储形式存在。
- ◆ **列存储：**Column-based storage, 数据是按照列为基础的逻辑存储单元进行存储的，一列中的数据在存储介质中以连续存储形式存在。
- ◆ **泽字节：**计算机存储容量单位，英文ZettaByte，简称ZB，是EB的1024倍。
- ◆ **NoSQL：**Not only SQL, 泛指非关系型的数据库，区别于关系数据库，它们不保证关系数据的ACID特性。NoSQL具有非常高的读写性能，无关系性也使数据库的结构简单。
- ◆ **NewSQL：**NewSQL 是对各种新的可扩展/高性能数据库的简称，这类数据库不仅具有NoSQL对海量数据的存储管理能力，还保持了传统数据库支持ACID和SQL等特性。
- ◆ **IaaS：**Infrastructure as a Service，即基础设施即服务。指把IT基础设施作为一种服务通过网络对外提供，并根据用户对资源的实际使用量或占用量进行计费的一种服务模式。
- ◆ **SaaS：**Software as a Service，软件即服务，即通过网络提供软件服务。SaaS平台供应商将应用软件统一部署在自己的服务器上，客户可以根据工作实际需求，通过互联网向厂商定购所需的应用软件服务，按定购的服务多少和时间长短向厂商支付费用，并通过互联网获得SaaS平台供应商提供的服务。
- ◆ **Kubernetes：**K8s，Google于2014年开源容器编排调度管理平台。相比与Swarm、Mesos，K8S引入Pod、Replica、Label、Service等机制简化了容器调度与管理，提供可靠性，增加了功能特性。
- ◆ **Serverless无服务器架构：**在无需管理服务器等底层资源的情况下完成应用的开发和运行，是云原生架构的核心组成部分。

方法论

- ◆ 头豹研究院布局中国市场，深入研究10大行业，54个垂直行业的市场变化，已经积累了近50万行业研究样本，完成近10,000多个独立的研究咨询项目。
- ◆ 研究院依托中国活跃的经济环境，从社会保险、人工智能、大数据等领域着手，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业的视野解读行业的沿革。
- ◆ 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- ◆ 研究院密切关注行业发展最新动向，报告内容及数据会随着行业发展、技术革新、竞争格局变化、政策法规颁布、市场调研深入，保持不断更新与优化。
- ◆ 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。

法律声明

- ◆ 本报告著作权归头豹所有，未经书面许可，任何机构或个人不得以任何形式翻版、复刻、发表或引用。若征得头豹同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“头豹研究院”，且不得对本报告进行任何有悖原意的引用、删节或修改。
- ◆ 本报告分析师具有专业研究能力，保证报告数据均来自合法合规渠道，观点产出及数据分析基于分析师对行业的客观理解，本报告不受任何第三方授意或影响。
- ◆ 本报告所涉及的观点或信息仅供参考，不构成任何证券或基金投资建议。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告或证券研究报告。在法律许可的情况下，头豹可能会为报告中提及的企业提供或争取提供投融资或咨询等相关服务。
- ◆ 本报告的部分信息来源于公开资料，头豹对该等信息的准确性、完整性或可靠性不做任何保证。本报告所载的资料、意见及推测仅反映头豹于发布本报告当日的判断，过往报告中的描述不应作为日后的表现依据。在不同时期，头豹可发出与本报告所载资料、意见及推测不一致的报告或文章。头豹均不保证本报告所含信息保持在最新状态。同时，头豹对本报告所含信息可在不发出通知的情形下做出修改，读者应当自行关注相应的更新或修改。任何机构或个人应对其利用本报告的数据、分析、研究、部分或者全部内容所进行的一切活动负责并承担该等活动所导致的任何损失或伤害。

“拒绝苟活 瞬间精彩”

我不想在北上广苟活，我要的是生活。
每一天工资在活期里真的活起来，
我感觉我活着！

—— 发财宝（货币基金）

过往业绩不预示未来表现，市场有风险，投资需谨慎



中正达广基金
ZHONGZHENG DAGUANG FUND

价值 | 平衡 | 快乐 | 爱❤️

证监会核准的独立基金销售机构（沪证监许可[2015]85号）



详情咨询



头豹报告库账户

- 全行业覆盖、近5000本报告展现、支持100万+数据搜索、每年持续更新1000+行企研究报告
- 解决细分行业知识空白
- 价值研究体系助力投资决策
- 月卡、季卡、年卡灵活订阅

报告找不到，马上上头豹

让专业 更专业

头豹定制报告

- 轻量化咨询：低价（5万起） 高质（深度） 高效（2周起）
- 对口行业资深分析师执笔
- 满足企业及机构：品宣、业务发展、信息获取等诉求

详情咨询



■ 头豹研究院简介

- ◆ 头豹研究院是中国大陆地区首家B2B模式人工智能技术的互联网商业咨询平台，已形成集行业研究、政企咨询、产业规划、会展会议行业服务等业务为一体的一站式行业服务体系，整合多方资源，致力于为用户提供最专业、最完整、最省时的行业和企业数据库服务，帮助用户实现知识共建，产权共享
- ◆ 公司致力于以优质商业资源共享为基础，利用大数据、区块链和人工智能等技术，围绕产业焦点、热点问题，基于丰富案例和海量数据，通过开放合作的研究平台，汇集各界智慧，推动产业健康、有序、可持续发展



四大核心服务

企业服务

为企业提供定制化报告服务、管理咨询、战略调整等服务

云研究院服务

提供行业分析师外派驻场服务，平台数据库、报告库及内部研究团队提供技术支持服务

行业排名、展会宣传

行业峰会策划、奖项评选、行业白皮书等服务

园区规划、产业规划

地方产业规划，园区企业孵化服务

报告阅读渠道

头豹官网 —— www.leadleo.com 阅读更多报告

头豹小程序 —— 微信小程序搜索“头豹”、手机扫上方二维码阅读研报



添加右侧头豹分析师微信，身份认证后邀您进入行研报告分享交流微信群



详情咨询



客服电话

400-072-5588



上海

王先生： 13611634866
李女士： 13061967127



深圳

李先生： 18916233114
李女士： 18049912451



南京

杨先生： 13120628075
唐先生： 18014813521

头豹 Project Navigator 领航者计划介绍



备注：活动解释权均归头豹所有，活动细则将根据实际情况作出调整。

读完报告有问题？

快，问头豹！你的智能随身专家



扫描二维码
即刻联系你的智能随身专家

千元预算的
高效率轻咨询服务

STEP04 专业高效解答
书面反馈、分析师专访、
专家专访等多元化反馈方式

STEP03 解答方案生成
大数据×定制调研
迅速生成解答方案

STEP02 云研究院后援
云研究院7×24待命
随时评估解答方案

STEP01 智能拆解提问
人工智能NLP技术
精准拆解用户提问