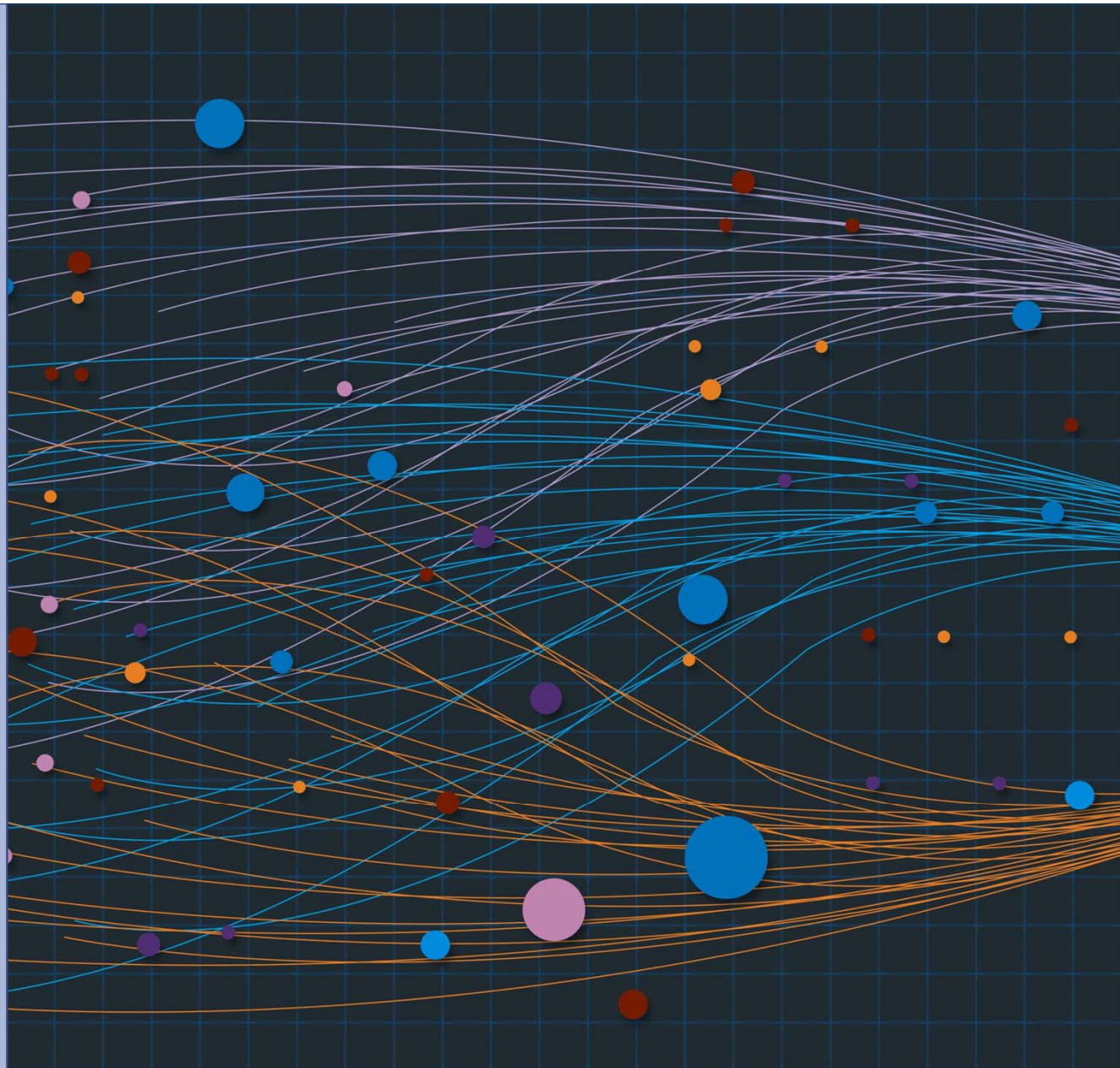


2021年 中国数据管理领域 发展白皮书

2021/10

报告提供的任何内容（包括但不限于数据、文字、图表、图像等）均系弗若斯特沙利文及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）。未经弗若斯特沙利文及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，弗若斯特沙利文及头豹研究院保留采取法律措施、追究相关人员责任的权利。弗若斯特沙利文及头豹研究院开展的所有商业活动均使用“弗若斯特沙利文”、“沙利文”、“头豹研究院”或“头豹”的商号、商标，弗若斯特沙利文及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弗若斯特沙利文或头豹研究院开展商业活动。



概要

数据管理定义

数据管理是利用计算机硬件和软件技术对数据进行有效的收集、存储、处理和应用的过程，旨在从原始数据中抽取、推导出有价值的信息，以支撑企业决策。数据管理平台最基础的功能，是支持各项数据管理流程线上化运行实现数据治理的常态化，此外数据管理平台助力企业用户更好的整合数据资产、提升数据质量。

综合型数据管理服务商依托自身技术优势，打造“采、存、算、管、用”全生命周期数据管理的一站式数据管理平台，助力企业降低构建大数据平台成本。

数据管理市场环境

中国数据管理市场具备全球最大、最复杂的应用场景，吸引了高热度的人才与资本。

数据管理对各行业收益均有明显的放大作用，其中在保险业和金融业的效果最为明显；更及时的决策/观察是数据管理的主要受益来源。需求端的用户则对数据价值有了更深的认识，通过改善传统绩效评价指标，发掘新的增长源和推出数据驱动的新业务，让数据管理创造收益。

伴随企业信息化进程的加快，提升企业决策效率、提高企业运营效率、规避企业管理风险成为下游企业用户向中游服务商购买数据管理产品及服务的主要因素。

产业发展对大数据复合型人才的需求日益增加，但现有的人才培养数量和速度难以满足现实需求，大数据服务行业面临人才短缺风险。企业对数据管理人才的认知从专家型转变为领导者，随着企业对数据管理的认知发生变化，数据管理人才在企业的地位正在提升。

数据管理解决方案

数据湖和数据仓库是实现正式数据管理解决方案的两种主流架构，数据仓库更重视大数据处理的效率，偏重于企业成长性；数据湖可处理非结构化数据，偏重于增加企业数据管理灵活性。

数据湖和数据仓库的边界正在慢慢模糊，数据湖自身的治理能力、数据仓库延伸到外部存储的能力都在加强，湖仓一体的出现让数据管理的灵活性与成长性得到了统一。

数据湖和数据仓库通过连结五大数据管理要素和提供数据高级分析的基础，实时支持企业决策，加速了企业数据的价值创造。

数据库

数据库是按照数据结构来组织、存储和管理数据的“仓库”，按照所使用模型所描述数据在某一时刻的状态或内容的不同，可分为关系型数据库和非关系型数据库。其中不同的架构模式又可分为集中式数据库和分布式数据库。

以数据分析、数据流通及事务处理技术为核心的大数据技术体系日渐完善，数据库技术和数据服务融合程度加深，助力企业用户数据价值释放。

数据治理

为了适应新时代基于数据的生产、分析、决策的需求，包括数据管理和数据价值变现的数据治理成为了关键。中国数据资产管理 and 应用尚处于摸索阶段，企业及政府部门尚未建立及完善有效的管理和应用数据模式，影响大数据在各领域的共享和应用。

目录

CONTENTS

◆ 数据管理综述	-----	11
• 范畴与定义	-----	12
• 厂商类型	-----	15
• 行业环境	-----	16
• 商业实践分析	-----	17
• 用户需求分析	-----	19
• 下游应用概况	-----	20
• 人才需求与供应	-----	22
◆ 数据管理解决方案	-----	24
• 数据仓库	-----	25
• 数据湖	-----	26
• 部署方案	-----	27
• 湖仓一体	-----	28
• 需求分析	-----	30
• 商业实践分析	-----	31
• 需求分析	-----	34
• 政策分析	-----	35
• 市场规模	-----	36

目录

CONTENTS

◆ 数据库	-----	37
• 定义与分类	-----	38
• 发展趋势	-----	39
• 分布式架构	-----	41
• 技术洞察	-----	42
• 行业发展历程	-----	49
• 产业链分析	-----	50
• 政策分析	-----	55
• 产业驱动因素分析	-----	56
• 市场规模	-----	58
• 行业风险分析	-----	59
◆ 数据治理	-----	60
• 定义	-----	61
• 发展路径	-----	62
• 数据资产管理措施	-----	63
• 应用现状	-----	66
• 市场规模	-----	67
• 企业对比分析	-----	68

目录

CONTENTS

◆ 工业领域的管理	-----	69
• 数据资源	-----	70
• 产业链	-----	71
• 中游业务模式	-----	72
• 需求分析	-----	73
• 应用价值	-----	74
◆ 附录	-----	75
• 数据管理厂商图谱	-----	76
◆ 名词解释	-----	77
◆ 方法论	-----	80
◆ 法律声明	-----	81

CONTENTS

◆ Data management overview	-----	11
• Scope and definition	-----	12
• Vendor Type	-----	15
• Industry environment	-----	16
• Business practice analysis	-----	17
• User needs analysis	-----	19
• Overview of downstream applications	-----	20
• Talent demand and supply	-----	22
◆ Data Management Solution	-----	24
• Data Warehouse	-----	25
• Data Lake	-----	26
• Deployment plan	-----	27
• Integration of Data Lake and Data Warehouse	-----	28
• Demand analysis	-----	30
• Business practice analysis	-----	31
• Demand analysis	-----	34
• Policy Analysis	-----	35
• Market Size	-----	36

CONTENTS

◆ Database	-----	37
• Definition and Classification	-----	38
• Development Trend	-----	39
• Distributed architecture	-----	41
• Technology Insights	-----	42
• Industry development history	-----	49
• Industry Chain Analysis	-----	50
• Policy Analysis	-----	55
• Analysis of industry driving factors	-----	56
• Market Size	-----	58
• Industry Risk Analysis	-----	59
◆ Data Governance	-----	60
• Definition	-----	61
• Development path	-----	62
• Data asset management measures	-----	63
• Application status	-----	66
• Market size	-----	67
• Comparative analysis of vendors	-----	68

CONTENTS

- ◆ **Data management in the industrial field** ----- 69
 - Data resources ----- 70
 - Industry Chain ----- 71
 - Midstream business model ----- 72
 - Demand analysis ----- 73
 - Application Value ----- 74
- ◆ **Appendix** ----- 75
 - Landscape of Data Management Vendors ----- 76
- ◆ **Terms** ----- 77
- ◆ **Methodology** ----- 80
- ◆ **Legal Statement** ----- 81

图表目录

List of Figures and Tables

图表1: 数据管理范畴	----	12	图表20: 数据管理解决方案发展阶段	----	28
图表2: 数据管理流程	----	13	图表21: 数据管理解决方案服务架构的变化	----	29
图表3: 数据管理平台的定义及构成	----	14	图表22: 数据管理解决方案在六大行业的需求分析	----	30
图表4: 中国数据管理平台服务商类型	----	15	图表23: 数据管理对行业核心价值节点带来的变化	----	31
图表5: 中国数据库行业环境	----	16	图表24: 五大数据管理要素	----	32
图表6: 各行业每10%数据管理的投入对资产收益率的提升	----	17	图表25: 数据湖和数据仓库作为加速器参与大数据管理	----	33
图表7: 正式数据管理解决方案的主要收益来源占比	----	18	图表26: 各领域数据管理解决方案的潜在需求	----	34
图表8: 2020年中国大数据服务企业用户需求动因	----	19	图表27: 数据管理解决方案相关政策整理	----	35
图表8: 2019年中国企业用户在大数据服务的投入意愿	----	19	图表28: 中国数据管理解决方案市场规模, 2015-2024年预测	----	36
图表9: 2020年部分地区对数据管理的重视程度	----	20	图表29: 数据库的分类	----	38
图表10: 2020年下游企业用户需求动因分析	----	20	图表30: 三类软件产品细分的关键差异	----	39
图表11: 中国数据管理在各个应用场景的成熟度	----	20	图表31: 数据库架构变革	----	40
图表13: 数据价值难以有效发挥的原因	----	21	图表32: 大数据技术融合发展	----	40
图表14: 2020年中国不同地域对大数据人才的需求	----	22	图表33: 数据属性、计算载体与数据库架构的演进	----	41
图表15: 2019年中国大数据行业人才需求与供给分布	----	22	图表34: 事务一致性属性及实现思路	----	42
图表16: 企业对数据管理人才认知的变革	----	23	图表35: 2PC vs. 3PC vs. Paxos vs. Raft	----	42
图表17: 数据仓库架构	----	25	图表36: 数据库集群架构分类	----	43
图表18: 数据湖架构	----	26	图表37: SN vs. SM vs. SD	----	43
图表19: 数据管理系统部署方案	----	227	图表38: NewSQL的演进路径	----	44

图表目录

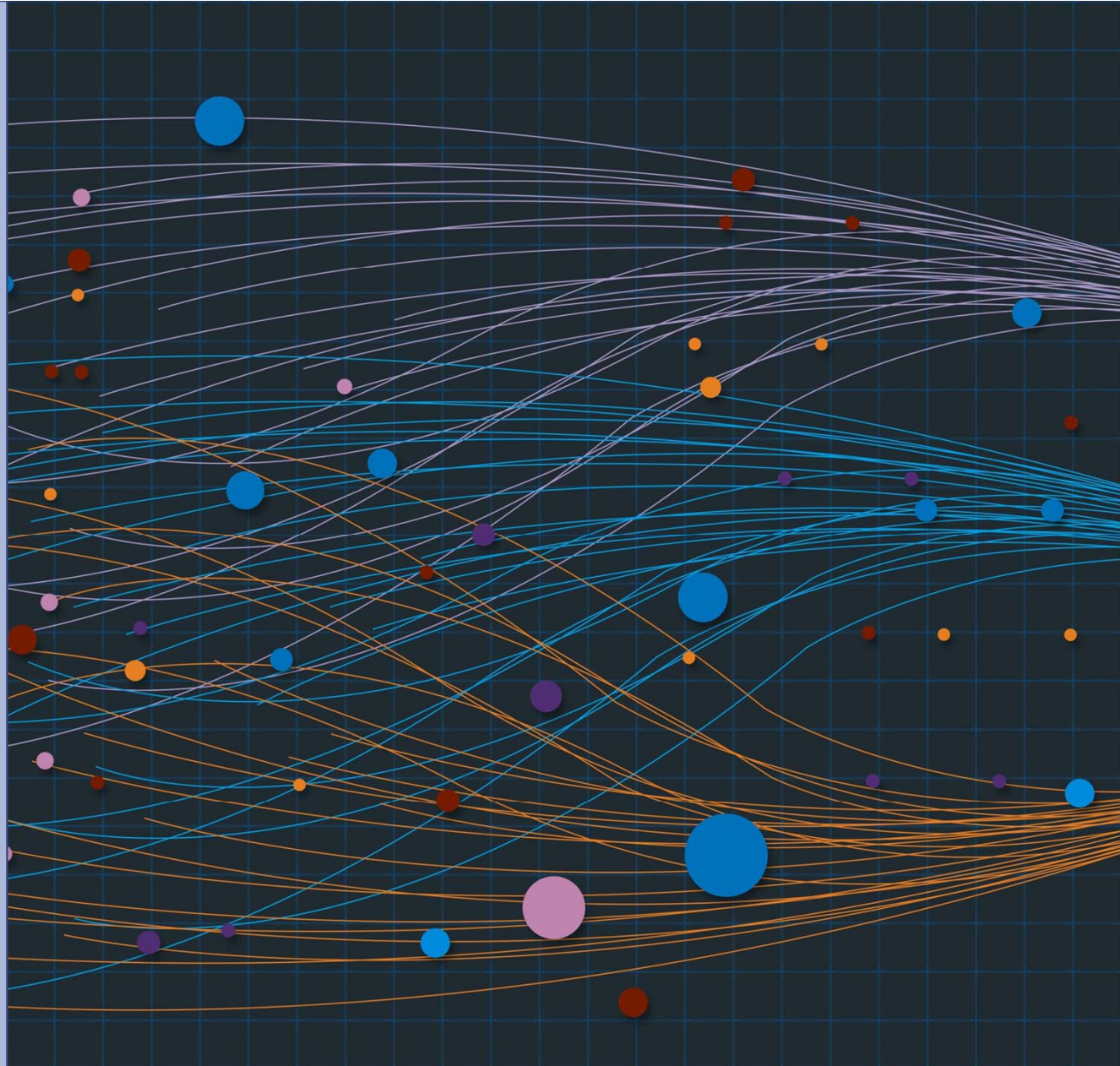
List of Figures and Tables

图表39: 传统数据库应用架构数据流	----	45	图表58: 2016-2025年中国数据库市场规模及预测	----	58
图表40: 分离架构HTAP简要原理	----	45	图表59: 数据库行业风险	----	59
图表41: OLTP vs. OLAP vs. HTAP	----	45	图表60: 数据治理宏观架构	----	61
图表42: 云化与微服务化的演进	----	46	图表61: 数据治理演变历程	----	62
图表43: 运维演变历程	----	47	图表62: 数据管理节点的业务影响与实施难度	----	62
图表44: 智能运维	----	47	图表63: 数据资产管理实施步骤	----	63
图表45: AI赋能数据库对比	----	48	图表64: 中国及海外国内外数据资产管理标准/模型及框架	----	64
图表46: 中国数据库行业发展历程	----	49	图表65: 国际数据管理协会DAMA-DMBOK2理论框架	----	64
图表47: 中国数据库产业链	----	50	图表66: 数据资产管理AIGOV五星模型内容及价值	----	65
图表48: 中国X86服务器提供商竞争格局, 2019-2020年	----	51	图表67: 数据治理相关规划	----	66
图表49: 中国X86服务器市场需求结构, 2019-2020年	----	51	图表68: 参与数据治理工作的部门	----	66
图表50: 供给端数据资源流	----	52	图表69: 数据治理参考标准或方法论	----	66
图表51: 开源社区构成	----	52	图表70: 大数据服务市场规模, 2018-2024年预测	----	67
图表52: 中游数据库厂商关键成功因素分析	----	53	图表71: 数据治理在大数据平台建设中的平均预算占比	----	67
图表53: 中国分析型数据库需求结构	----	54	图表72: 数据治理市场规模, 2018-2024年预测	----	67
图表54: 中国分析型数据库收入-数据乘数	----	54	图表73: 数据治理领域企业对比	----	68
图表55: 各国或地区数据库产业建设与发展方针	----	55	图表74: 中国工业大数据源特征	----	70
图表56: 数据库产业相关政策	----	55	图表75: 中国工业数据行业产业链	----	71
图表57: 数据库产业驱动因素	----	56	图表76: 工业大数据服务商的业务模式	----	72
			图表77: 中国工业大数据市场产品结构(按销售额计) 2020年	----	73
			图表78: 工业数据管理应用场景及应用价值	----	74
			图表79: 数据管理厂商图谱	----	76

Chapter 1

数据管理综述

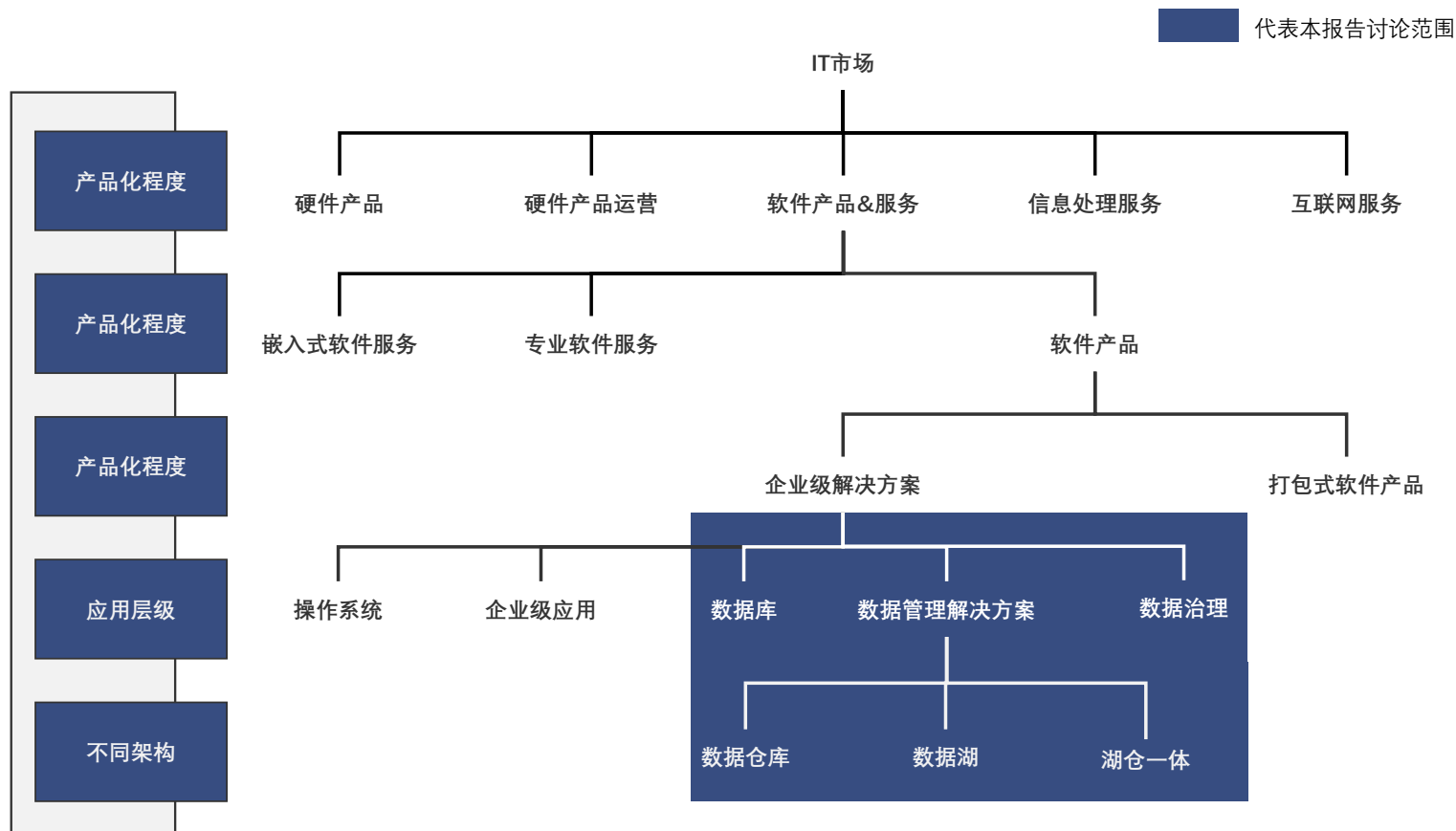
- 范畴与定义
- 厂商类型
- 行业环境
- 商业实践分析
- 用户需求分析
- 下游应用概况
- 人才需求与供应



数据管理综述 范畴与定义

数据管理是利用计算机硬件和软件技术对数据进行有效的收集、存储、处理和应用的过程，其目的在于充分有效地发挥数据的作用

数据管理范畴



数字化转型趋势

- ❑ 在数字化转型的浪潮中，随着数据量的增加和数据应用场景的丰富，数据间的关系变得更加复杂
- ❑ 数据管理的重要性凸显，企业应当拥抱技术更迭的节奏，以适应时代变化

数据管理

- ❑ 数据管理是利用计算机硬件和软件技术对数据进行有效的收集、存储、处理和应用的过程
- ❑ 其目的在于充分有效地发挥数据的作用

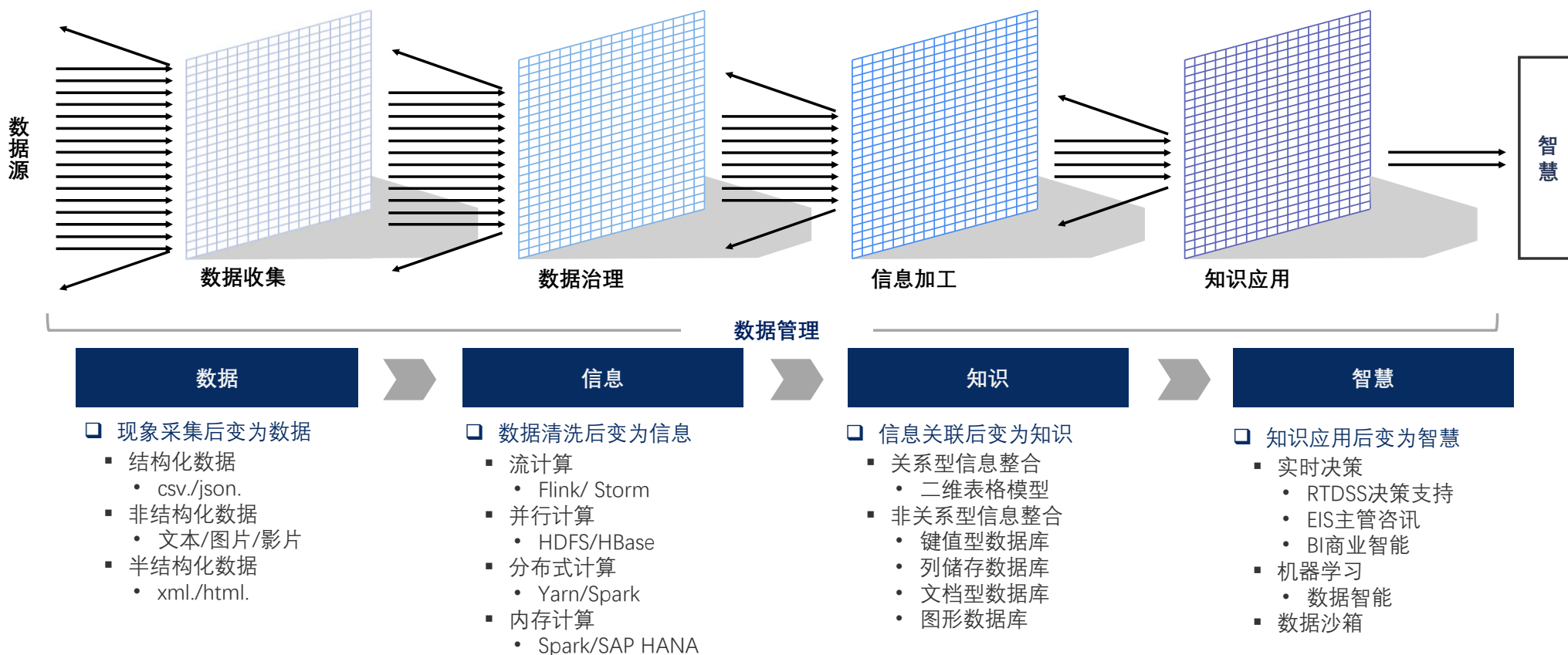
数据管理产品

- ❑ 数据管理产品包括数据库、数据集市、数据仓库、数据湖、湖仓一体、数据治理

数据管理综述 范畴与定义

数据管理是利用计算机硬件和软件技术对海量数据进行有效的收集、存储、计算、分析和应用的过程，旨在从原始数据中抽取、推导出有价值的信息，以支撑企业决策

数据管理流程



数据管理综述 范畴与定义

数据管理平台最基础的功能，是支持各项数据管理流程线上化运行实现数据治理的常态化，此外数据管理平台助力企业用户更好的整合数据资产、提升数据质量

数据管理平台的定义及构成



基础功能

定义：数据管理平台是对数据进行全链路、全生命周期管理的平台，帮助用户管理海量的多源异构数据，对用户数据进行统一的存储、清洗、加工。数据管理平台依托以数据标准、元数据管理、数据质量稽核为核心的数据治理体系帮助用户更好的整合数据资产、提升数据质量。

基础功能



各项管理流程的支持：实现数据管理各领域中的流程常态化运行



数据标准管理：实现对全行级数据标准的维护和落标分析



数据质量管理：实现数据质量检核定义、质量检核分析、质量检核执行，形成数据质量知识库



元数据管理：实现对元数据采集、维护、版本管理，展现血缘分析、影响分析与数据地图



数据需求管理：提供面向业务用户的数据需求沟通渠道，跟踪数据需求、统计数据需求



其他功能：数据模型管理、数据安全治理、主从数据管理等功能

来源：中国电子信息产业发展研究院，摩尔元数，头豹研究院

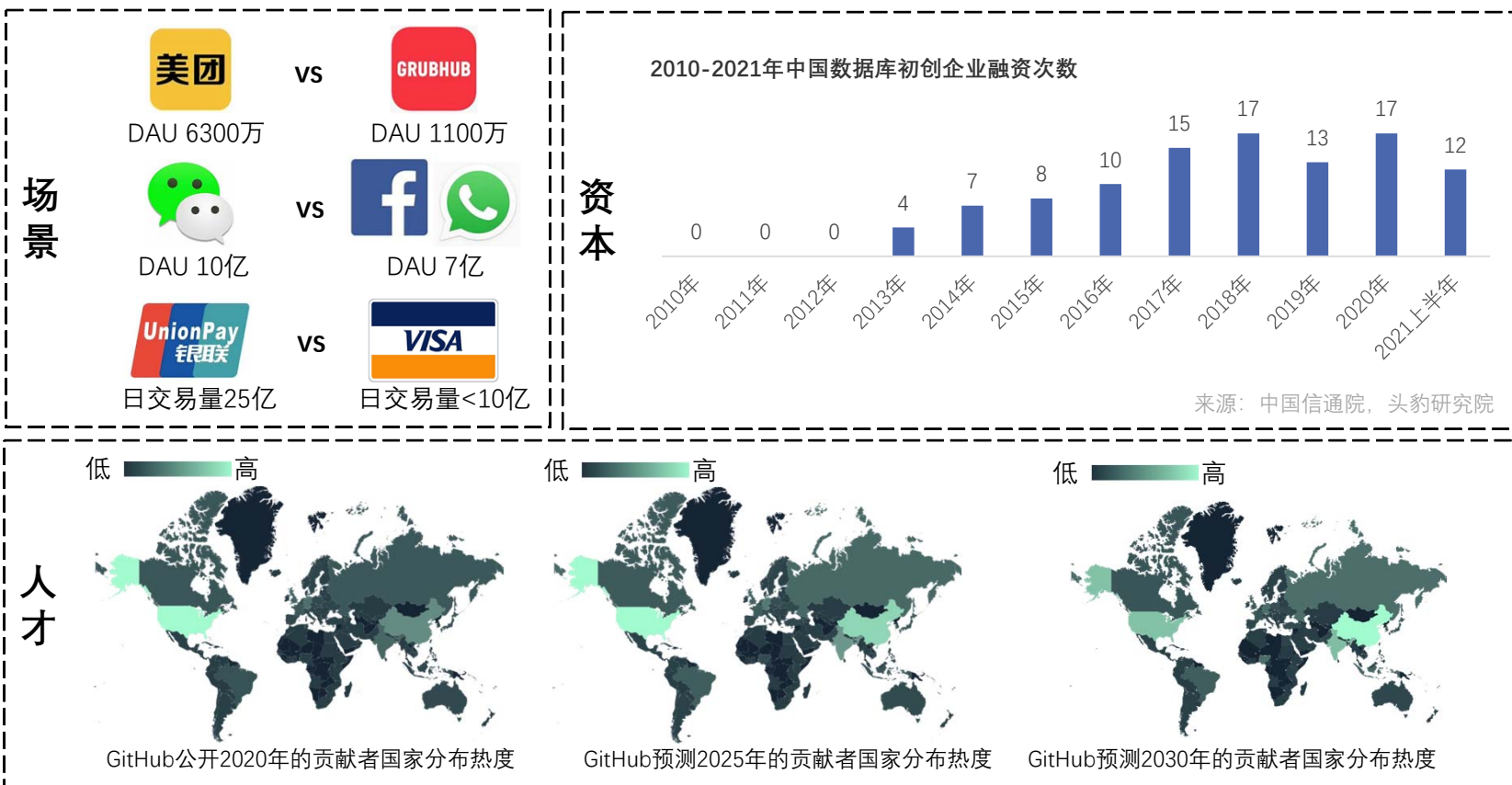
综合型数据管理服务商依托自身技术优势，打造“采、存、算、管、用”全生命周期数据管理的一站式数据管理平台，助力企业降低构建大数据平台成本

中国数据管理平台服务商类型



中国数据管理市场具备全球最大、最复杂的应用场景，吸引了高热度的人才与资本

中国数据库行业环境



来源：明势资本、Github、中国信通院，头豹研究院

最大、最复杂的应用场景

中国市场具备全球最大的数据流量规模，以外卖平台、即时通讯和支付举例，中国的数据场景代表了全球领先的实践场景，对数据库技术有着高热度的迭代需求，倒逼孵化更强大的数据库提供商。

IT相关人才热度将领先全球

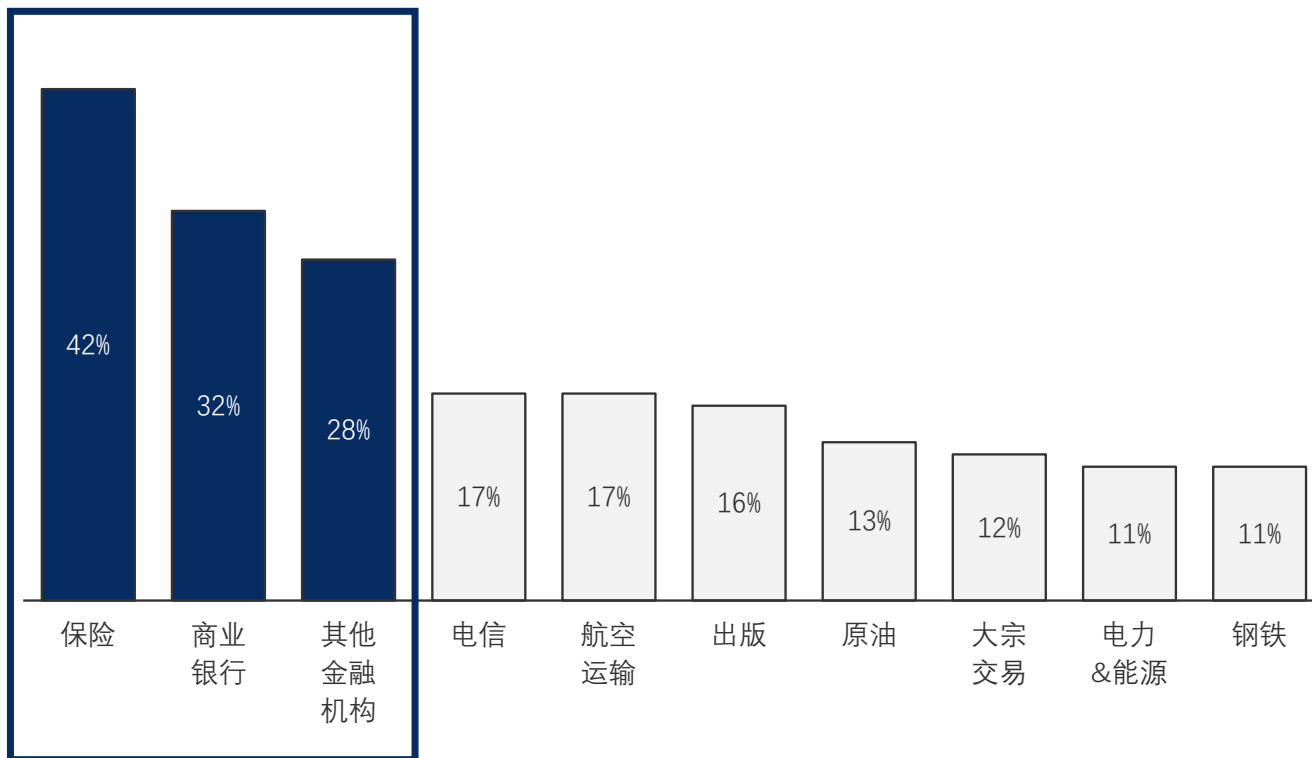
中国互联网行业的蓬勃发展与数字化转型趋势对IT相关人才有着巨大的需求，吸引着越来越多的人加入行业。

数据领域的融资热潮持续

中国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管理市场的高速增长。

数据管理对各行业收益均有明显的放大作用，其中在保险业和金融业的效果最为明显；更及时的决策/观察是数据管理的主要受益来源

各行业每10%数据管理的投入对资产收益率的提升 [百分比]



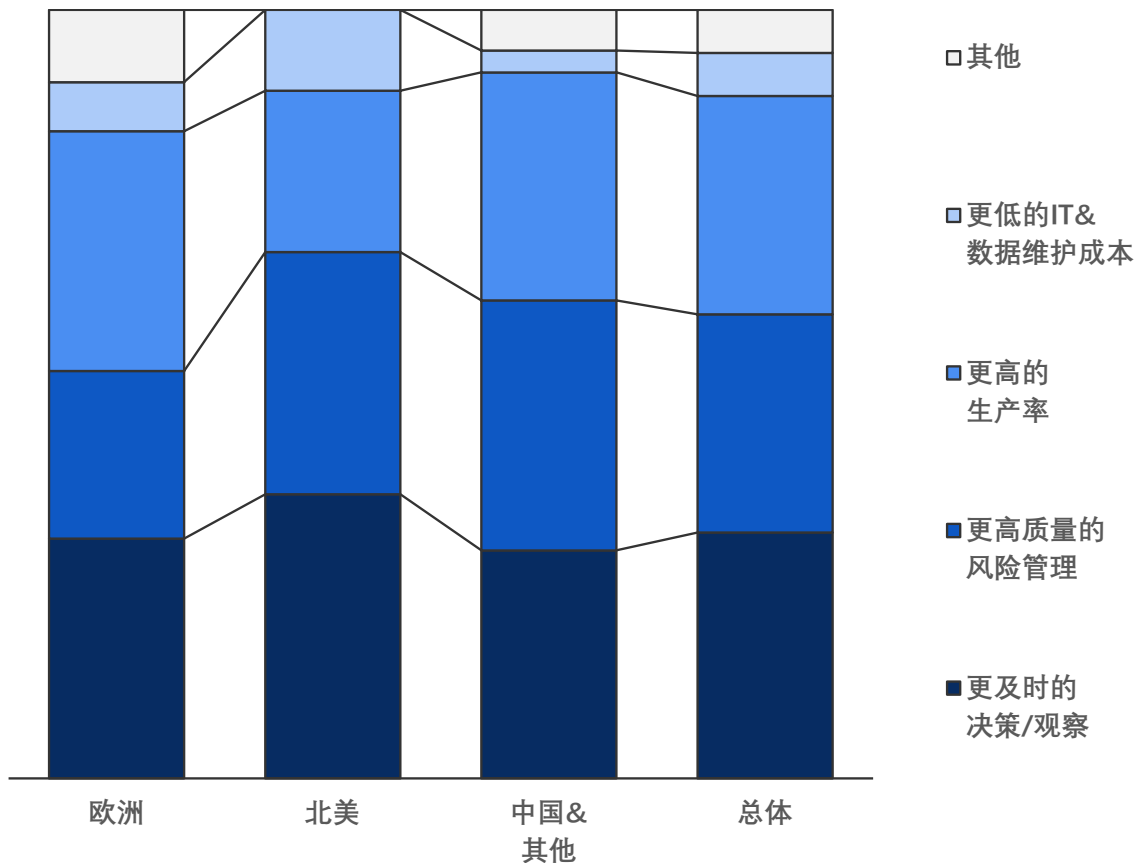
保险、商业银行、金融机构的行业共性：

- 劳动力密集型行业
- 数据壁垒
 - 营销环节，各子公司各自为政
 - 研发环节，金融衍生品的开发需要大量的数据，却不在体系内收集数据
 - 自身业务的连续性与跨部门协作性弱
- 大比例的对客业务
- 量大复杂的数据来源
 - 数据维度不统一（非结构化/半结构化）
 - 个人征信数据
 - 个人消费行为数据
 - 数据量大
 - 一个储户信用报告数据量高达10GB

数据来源：德克萨斯大学奥斯汀分校，沙利文

需求端的用户则对数据价值有了更深的认识，通过改善传统绩效评价指标，发掘新的增长源和推出数据驱动的新业务，让数据管理创造收益

正式数据管理解决方案的主要收益来源占比



数据管理产生收益的三种方式：

- 改善传统绩效评价指标
 - ✓ 加速增长
 - ✓ 提高生产力
 - ✓ 提高风控能力
- 发掘新的增长源
 - ✓ 从非结构化数据中发掘新的价值
 - ✓ 从结构化数据中更有效的提取价值
- 推出数据驱动的新业务
 - ✓ 实现数据的高级分析

无论何种增长方式都离不开大数据计算能力：

- 有效的大数据治理从冗杂海量的数据中发现关键的绩效管理节点
- 有效、快速、实时的大数据计算为高级分析提供直接创造价值的数据
- 高级分析将为企业提供更及时的决策和观察

数据来源：沙利文

企业用户将逐渐加大在数据管理解决方案的投入，希望实现决策和运营效率的提升以建立竞争优势，这种意愿将快速拉升数据管理解决方案市场需求

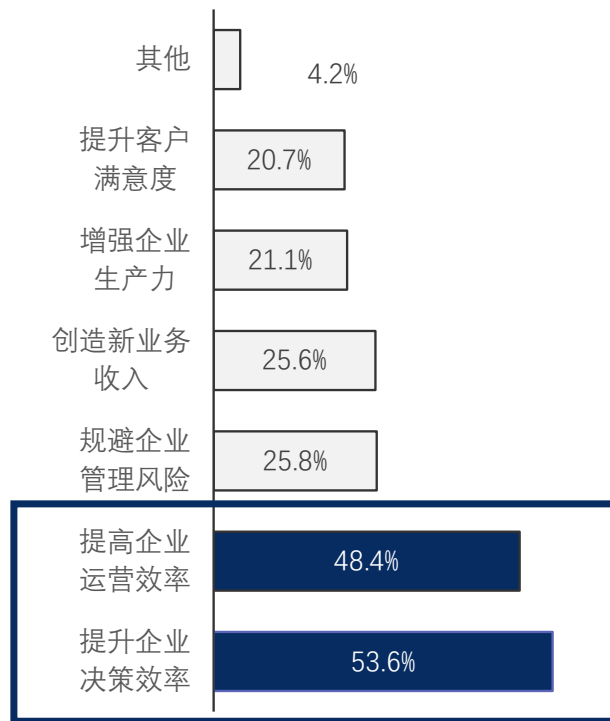
■ 数据管理解决方案驱动市场需求的逻辑

1. 有效提高财务表现：员工生产效率、净资产收益率、投资回报率、资产利润
2. 有效增进客户关系：增益创新能力获得新产品线的营收增长，拓展客户基础
3. 有效增益经营管理：资源利用水平、预测与生产计划、交付周期、服务条款

■ 边际投入兑换数据红利

不同程度对数据管理解决方案的投入，都会首先获得基础的数据存取与数据质量的明显增益。随着投入程度的增加，数据的不同属性包括质量、易用性、智能度、可及性和灵活性都会得到相应比例的增益。企业通过调整不同比例的投入与高级分析的策略，建立在行业中的竞争优势。

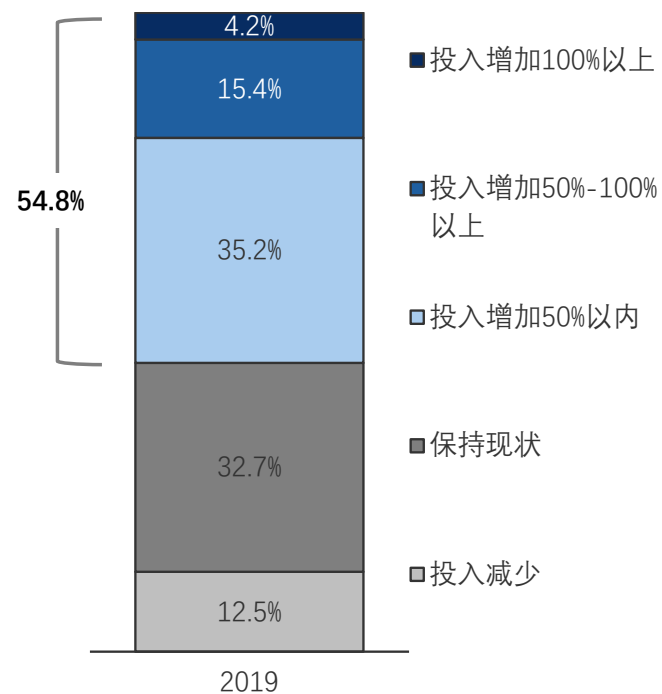
2020年中国大数据服务企业用户需求动因



提升企业决策效率和提升企业的运营效率是企业用户的首要需求

- 企业在运营、决策、风险管理、业务创新、生产及客户服务等效率提升意愿明显

2019年中国企业用户在大数据服务的投入意愿



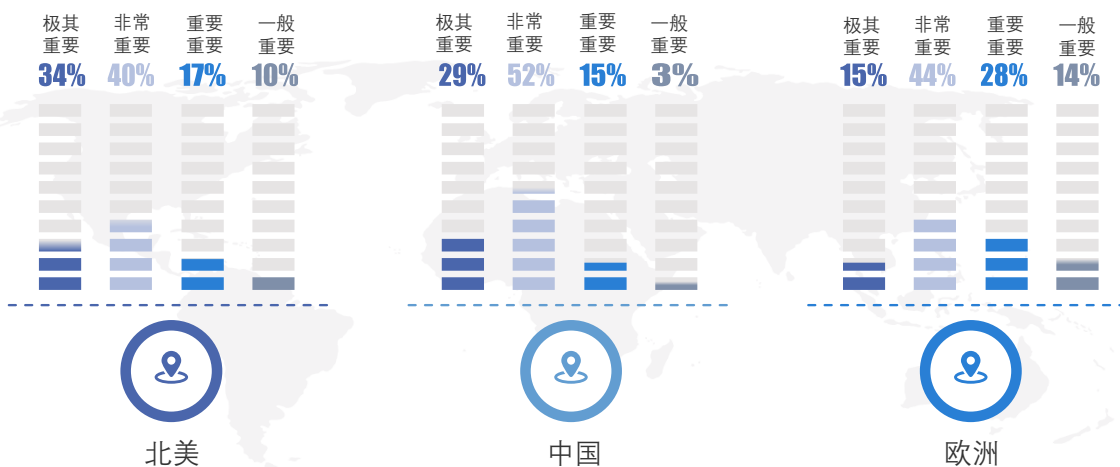
企业用户在数据管理解决方案的投入意愿较高

- 近55%的企业用户计划增加在数据管理解决方案的投入，将为大数据服务商提供较大发展空间。

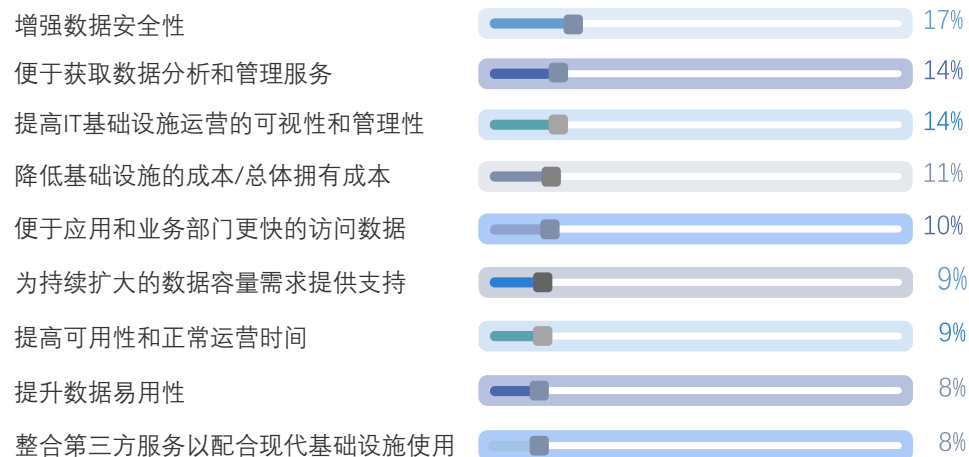
下游应用概况

伴随企业信息化进程的加快，提升企业决策效率、提高企业运营效率,规避企业管理风险成为下游企业用户向中游服务商购买数据管理产品及服务的主要因素

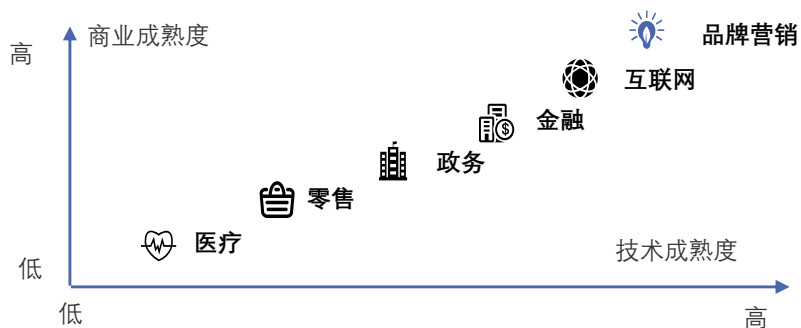
2020年部分地区对数据管理的重视程度



2020年下游企业用户需求动因分析



中国数据管理在各个应用场景的成熟度



数据管理产品在金融、品牌营销、政务领域应用较成熟

中国数据管理行业下游各领域基础设施成熟度及信息化程度是影响其数据智能应用的重要因素，数据管理产品在金融、品牌营销、政务领域应用较成熟。

- ✓ **金融领域**：金融行业于20世纪初开始信息化建设，基础信息化建设现已初步完成，此外，金融大部分数据均为结构化数据，数据标准化程度高，因此数据资产化程度高，数据管理平台在金融领域的渗透率较高。
- ✓ **品牌营销**：品牌营销是数据管理发展较早，较成熟的领域，在该领域，广告是品牌营销中最先落地的业务。在商业应用方面，品牌营销在广告场景已实现完全自动化，数据管理平台基于社交数据的营销自动化，基于海量数据形成消费者画像等业务也趋于成熟。

由于数据共享体系及数据整合标准化系统缺失，中国市场中的大数据利用率低下，信息共享和业务协同在地市和区县进展缓慢。

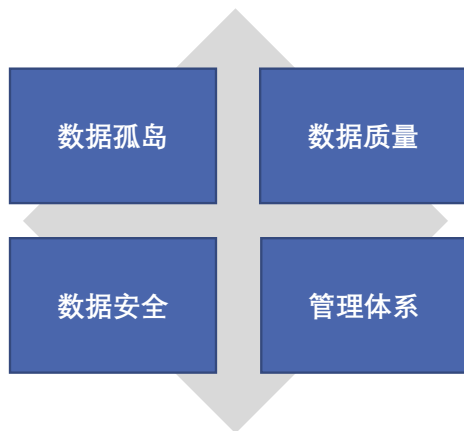
数据价值难以有效发挥的原因

缺乏统一的数据的视图

由于技术或管理制度的原因，企业的存在数据孤岛，散落在各个业务系统中。

缺乏安全的数据环境

数据安全风险包括了数据泄露与数据滥用。一旦发生数据安全事件，会对企业经营造成损失和对用户的隐私造成侵犯，束缚数据价值的释放。



数据质量低下

数据质量也决定了业务决策的分析，数据资产的价值建立在不断提升的数据质量的前提下。

缺乏数据价值管理体系

企业没有建立起一个有效管理和应用数据的模式，包括数据价值评估、数据成本管理等，对数据的服务和应用缺乏合规性的指导。

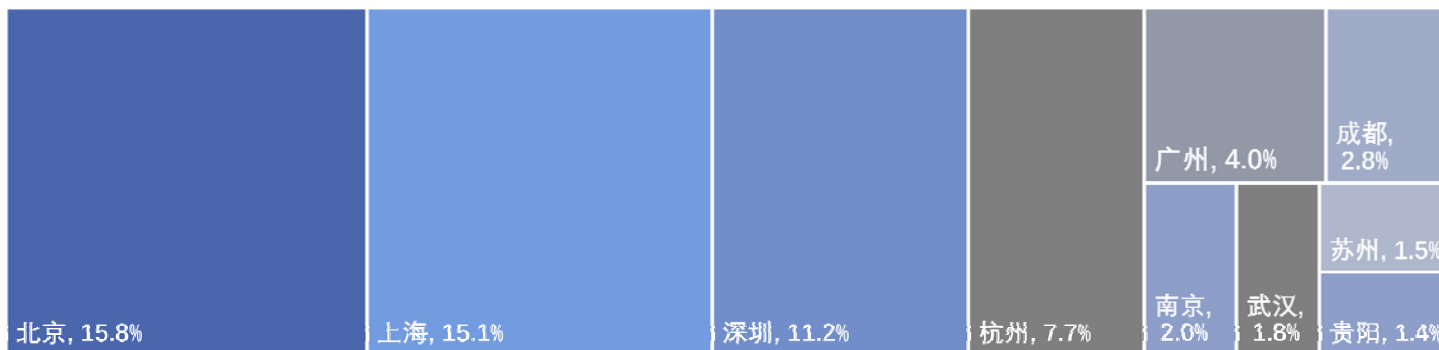
来源：阿里云，头豹研究院

数据割裂、数据共享体系的缺失导致数据利用率低

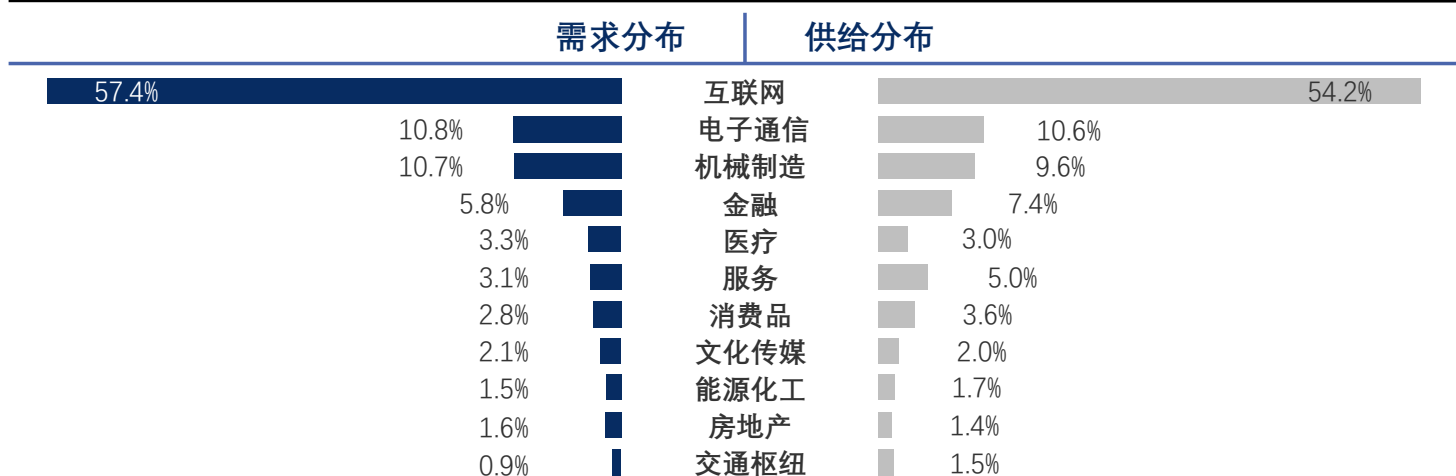
- 全球数据产生量呈现指数级增长，据Statista数据统计，中国是世界大数据存储量最高的国家，比重约占全球数存储量的22%，但数据利用率较低。
- 中国大数据服务行业数据源分布于政府、行业及企业三个主体中，其中80%以上的信息数据资源掌握在各级政府部门，但实现数据共享的省级政府部门仅占13%，实现少量数据共享的地市和区县仅占32%和28%，信息共享和业务协同在地市和区县进展缓慢。
- 大数据产业缺乏数据共享机制，不同行业数据整合缺少标准化的数据格式，数据割裂及数据共享体系的缺失影响数据利用率。大数据服务行业数据源的缺乏与信息孤岛的形成制约大数据产业的发展。

产业发展对大数据复合型人才的需求日益增加，但现有的人才培养数量和速度难以满足现实需求，大数据服务行业面临人才短缺风险。

2020年中国不同地域对大数据人才的需求



2019年中国大数据行业人才需求与供给分布



来源：猎聘网，Statista，中国信通院，头豹研究院

大数据复合型人才缺失，且行业及地域人才分布不均

- 中国（不含港、澳、台地区）大数据核心人才缺口达150万人，且存在地域和行业人才分布不均问题。
- 中国互联网、电子通信和金融三大领域人才需求占比近74%，其中互联网行业占据57.4%，但人才需求较高的三大领域人才供给占比不足，且人才需求过于集中，导致制造业等产业在转型升级过程中极度缺乏大数据人才。
- 目前，中国处于新旧动能转化提速期，大数据技术作为产业转型升级的底层基础工具，产业发展对大数据复合型人才的需求日益增加，但现有的人才培养数量和速度难以满足现实需求，大数据服务行业面临人才短缺风险。

企业对数据管理人才的认知从专家型转变为领导者，随着企业对数据管理的认知发生变化，数据管理人才在企业的地位正在提升

企业对数据管理人才认知的变革



■ IT时代进入DT时代

随着云服务和移动互联的普及，数据制造的能力向中小企业和消费者转移，各行各业竞争需要精细化运营，通过大数据分析洞察需求，生成情景化的预测性知识以建立差异化竞争力。

■ 大数据与物联网、人工智能将进一步融合

在新时代的竞争中，领导者应基于混合云、5G、物联网和边缘计算能力重新设计工作流程，提高企业适应力。开发基于数据的AI策略，将数据置于每一项业务决策的核心依据，制定清晰的商业计划，打造认知型企业。

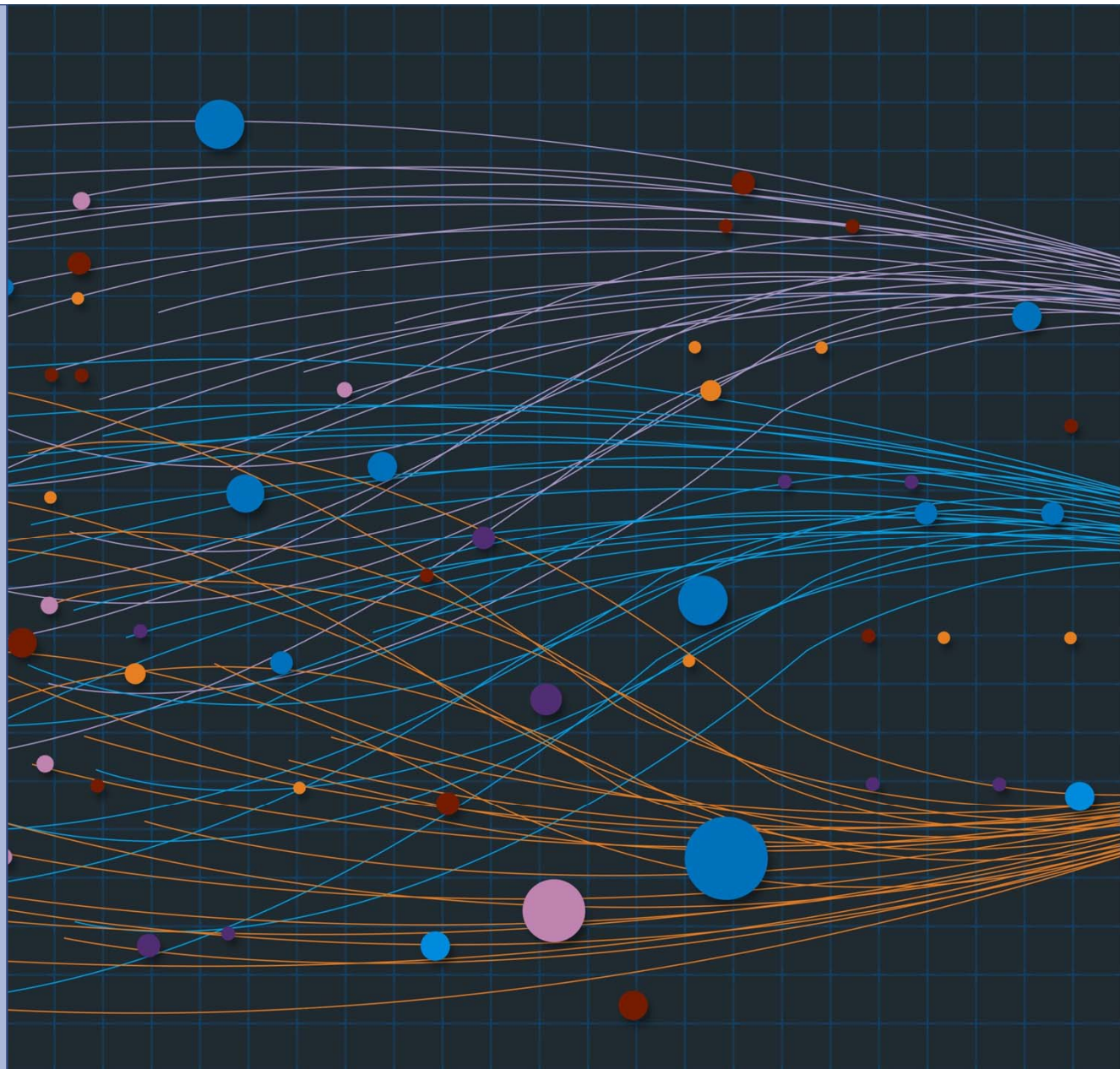
■ 数据决策需要领导者认知

数据决策：一个完整的预测支持决策的回路包括历史数据输入，模型训练，数据预测，决策，执行，结果收集和反馈。数据分析支撑管理决策，其前提是细化目的数据需要支撑哪部分的管理决策。比如：新功能上线后的用户活跃度，还需要进行哪些修改；新品的销售占比和区域表现，后期该如何倾斜；会员的购买力分析，如何针对性推送优惠信息等。只有明确了系列目的，数据才有可能针对性地被收集和使用，要求数据领导者做好全局规划，保证数据回路的有效推进。

Chapter 2

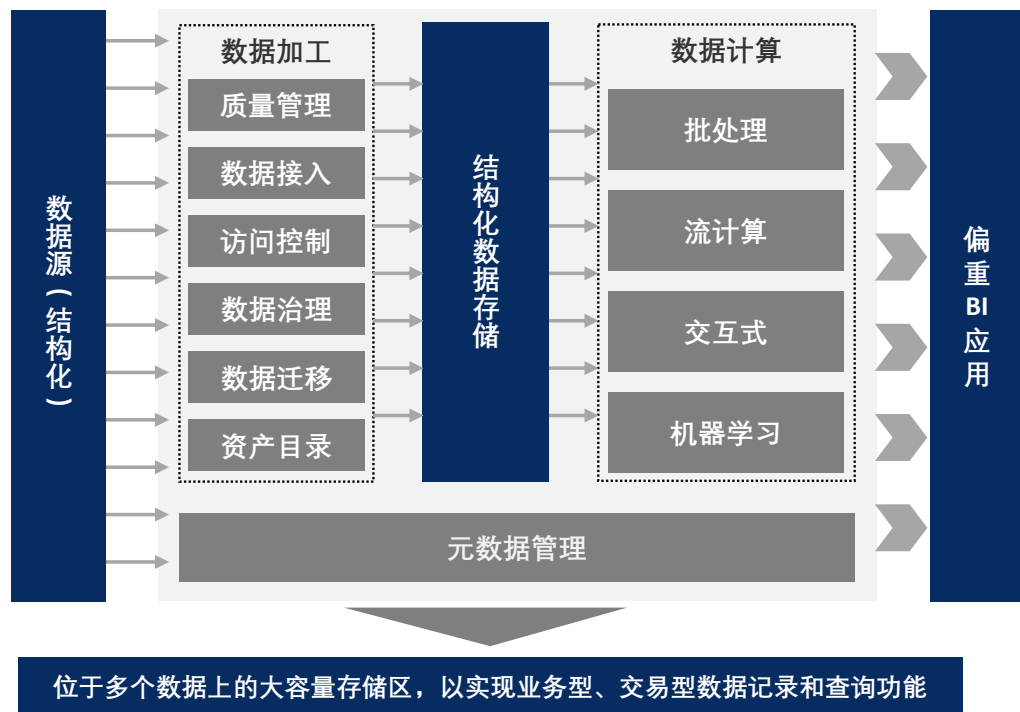
数据管理解决方案

- 数据仓库
- 数据湖
- 部署方案
- 湖仓一体
- 需求分析
- 商业实践分析
- 需求分析
- 政策分析
- 市场规模



数据湖和数据仓库是实现正式数据管理解决方案的两种主流架构，数据仓库更重视大数据处理的效率，偏重于企业成长性

数据仓库架构



本质

- 内置的存储系统，数据通过抽象的方式提供（例如采用Table或者View），不暴露文件系统
- 数据需要清洗和转化，通常采用ETL/ELT方式
- 强调建模和数据管理，供商业智能决策

描述

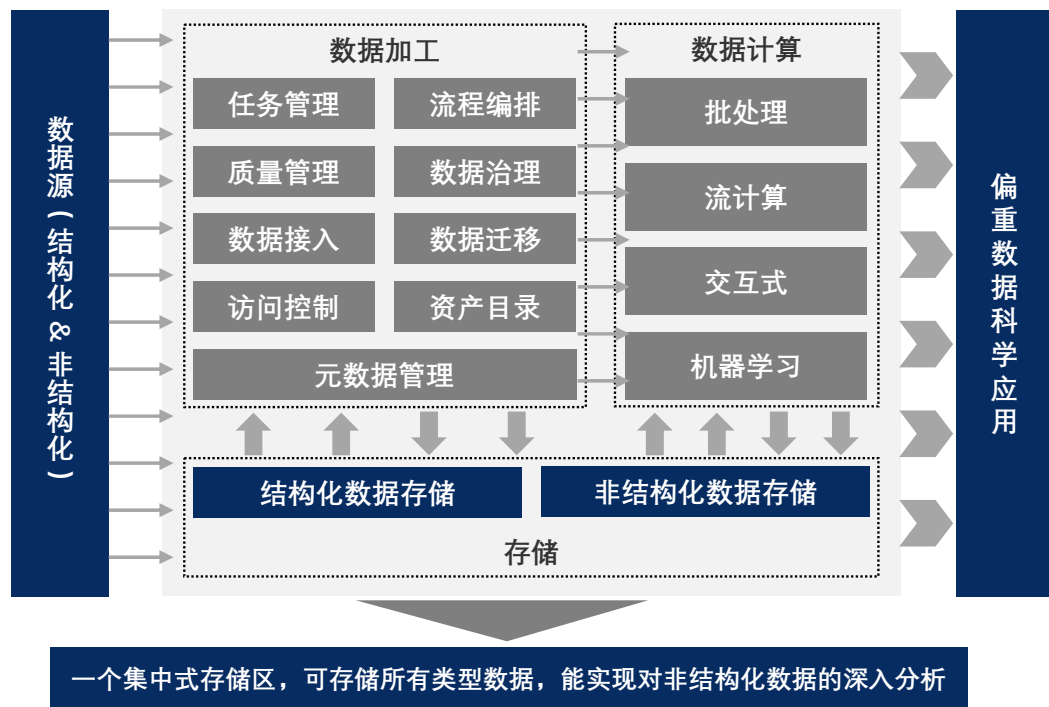
- 数据 来自事物系统、运营数据库和业务线应用程序的关系型数据
- 数据质量 可作为重要事实依据的高度监管数据
- Schema 设计在数据仓库实施之前（写入型Schema）
- 分析 批处理报告、BI和可视化
- 用户 业务分析师
- 性价比 更快的查询结果只需要较低的存储成本

优势

- 引擎深度理解数据，存储和计算可做深度优化
- 数据全生命周期管理，完善的血缘体系
- 细粒度的数据管理和治理
- 完善的元数据管理能力，易于构建企业级数据中台

数据湖和数据仓库是实现正式数据管理解决方案的两种主流架构，数据湖可处理非结构化数据，偏重于增加企业数据管理灵活性

数据湖架构



本质

- 统一的存储系统
- 存储原始数据
- 丰富的计算模型/范式
- 数据湖与上云无关

描述

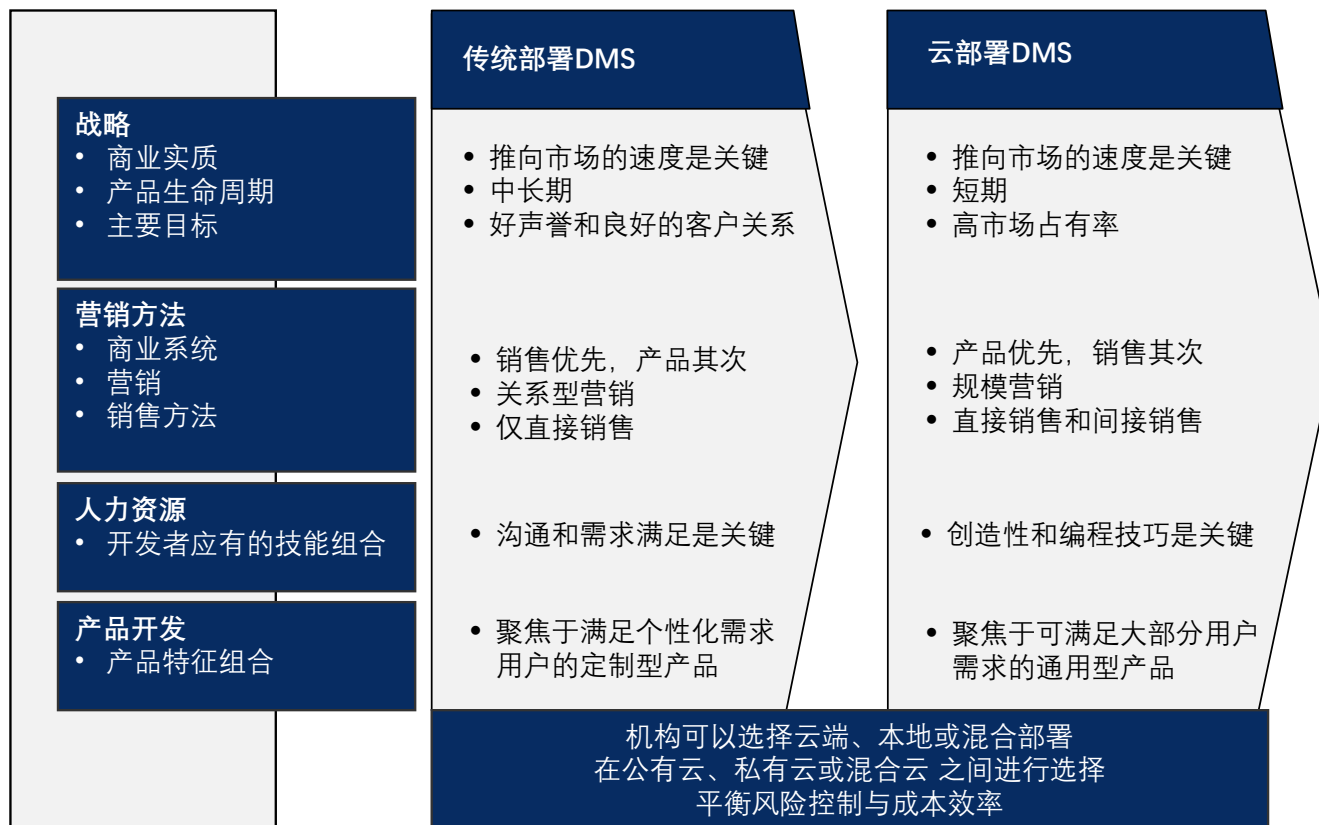
- 数据 来自设备、网站、应用和媒体等的关系型数据及非关系型数据
- 数据质量 任何无法进行监管的数据（例如原始数据）
- Schema 写入在分析时（读取型Schema）
- 分析 机器学习、预测分析、数据发现和数据分析
- 用户 数据科学家、数据开发人员和业务分析师
- 性价比 更快的查询结果只需要较低的存储成本

优势

- 收集和摄入所有数据源，获得整个无孤岛的数据集合
- 支持对实时和高速数据流执行ETL（抽取-转置-加载）功能
- 可扩展性和灵活性
- 具有人工智能的高级分析

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

数据管理系统部署方案



■ 具备数据储存计算处理分析能力的完整平台

以往，为了应对网络速度的不足、数据在各节点之间交换时间较长的问题，大数据分布式框架采用存储与计算耦合的形式，使数据在自身的储存点上完成计算，以降低交互。这是传统部署DMS。

■ 存储与计算耦合造成了额外成本

实际业务中，对于数据存储空间与计算能力的需求是各自变化的，使得两类资源的需求配比不可预见，当其一出现资源瓶颈时，资源的横向扩展必然导致储存或者是计算能力的冗余，并且对数据的迁移也造成了额外的成本。

■ 存算分离有效控制成本

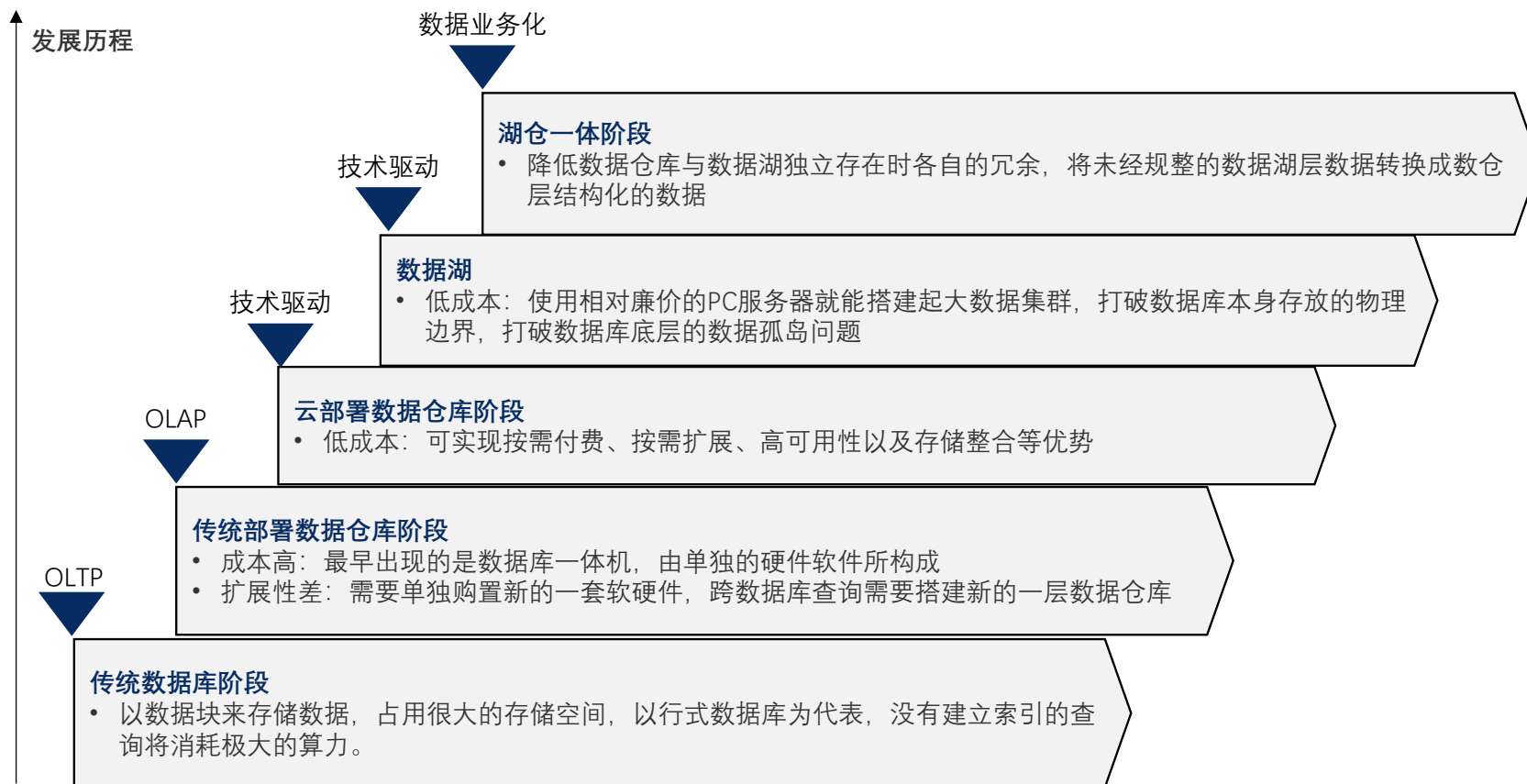
将储存和计算两个环节剥离形成两个独立的资源集合，互不干涉又通力合作，每个资源内部充分体现资源的规模聚集效应，使得单位资源的成本尽量减少，同时兼具充分的弹性以供横向扩展。当出现资源紧缺或富裕的时候，只需要对资源进行获取或回收，使用具备特定资源配比的专用节点进行，减少冗余，在资源需求差异化的场景中实现资源的合理配置。

■ 按需所取的云化服务优势显著

在存算分离的基础上，Serverless、云原生的概念的提出，使数据的处理分析能力摆脱了对于完整平台和工具的需求，大大降低了开发周期、节省开发成本，同时服务应用由提供方运维，实行按需付费，消除了复杂的运维过程和成本。

数据湖和数据仓库的边界正在慢慢模糊，数据湖自身的治理能力、数据仓库延伸到外部存储的能力都在加强，湖仓一体的出现让数据管理的灵活性与成长性得到了统一

数据管理解决方案发展阶段



数据管理面对的根本问题是对数据的存储与查询。

- 在传统数据库阶段，存储与查询都面临着巨大的成本与难度；
- 传统部署的数据仓库阶段，提高了对数据的治理能力降低查询的成本与难度，但可扩展性的局限决定了其降低查询成本的能力下兜界；
- 云部署数据仓库阶段，大大拓宽了可扩展性的边界，且免去托管、运维、软件投入等成本，按需收费实现高资源利用率，但无法解决数据仓库本身对非结构化数据不兼容的属性；
- 数据湖的实践，一举实现了存储性能的飞跃，兼容实时的、海量的、类型广泛的数据，且真正打破了数据库之间的物理壁垒；
- 湖仓一体的出现，吸收了数据湖对存储的优势和数据仓库对查询的优势，将大数据的使用门槛进一步降低。

对数据管理解决方案的需求逐渐聚焦于数据价值损失的尽量避免和数据对商业决策的更高支持，这两种需求的不断清晰促使数据管理解决方案产品“走向”湖仓一体

数据管理解决方案服务架构的变化

趋势：数据管理解决方案产品湖仓一体化

方式：架构重新被设计



提升企业的的数据管理解决方案效率

- 云原生的数据分析核计算架构提高了数据处理的流畅性

提升企业的的数据价值创造能力

- 统一的数据管理解决方案为企业提供数据资产化基础

对调用数据的灵活性和成长性需求，会随着企业用户的发展不断的演进。

■ 初创阶段的企业，数据从产生到消费的周期还很长，往往只需要联机事务处理（OLTP）系统记录业务事件，这是传统数据库的应用；而对不同业务的数据集中分析，才需要对数据清洗后储存在数据仓库中，提供联机分析处理（OLAP）分析，为企业提供成长性空间；

■ 当业务成长到一定的规模后，企业需要对海量且多样的数据进行储存与分析，强调对灵活性的需求，数据湖的优势凸显。分析方法则因为数据量的增加可以拓展到数据挖掘（Data Mining），进而支持决策支持系统（DSS）、主管资讯系统（EIS）分析出更具价值的信息与知识，帮助构建商业智能（BI）。

■ 而对于主流互联网企业的营销和运营等场景、电信行业的经营分析与金融行业的风控管理这些应用场景中，数据湖对海量数据存储能力与数据仓库处理高度结构化数据的提取能力都变得十分重要。在数据重力的作用下，巨量的数据传输成本已将实际的业务抵在重压之下，湖仓一体正是在数据业务化的需求中诞生。

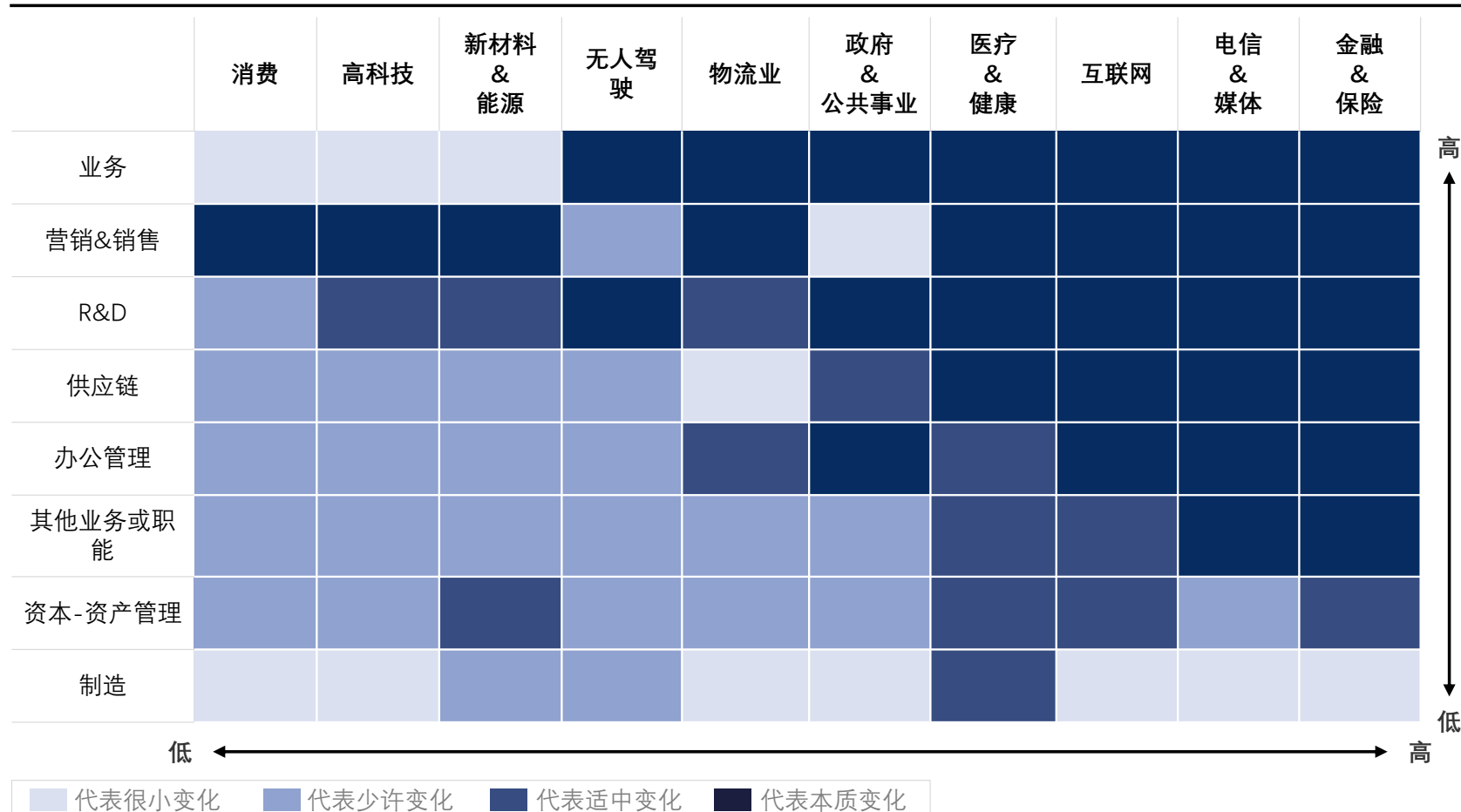
数据管理解决方案在金融、电信&媒体、交通、政府、医疗&健康与科技&能源六大行业有着不同的应用场景

数据管理解决方案在六大行业的需求分析



数据管理为各行业提供数据资产化的基础，数据的资产化将从营销、研发、供应链等关键价值节点重构企业价值链，构建数据生态圈良性循环

数据管理对行业核心价值节点带来的变化



价值节点变化

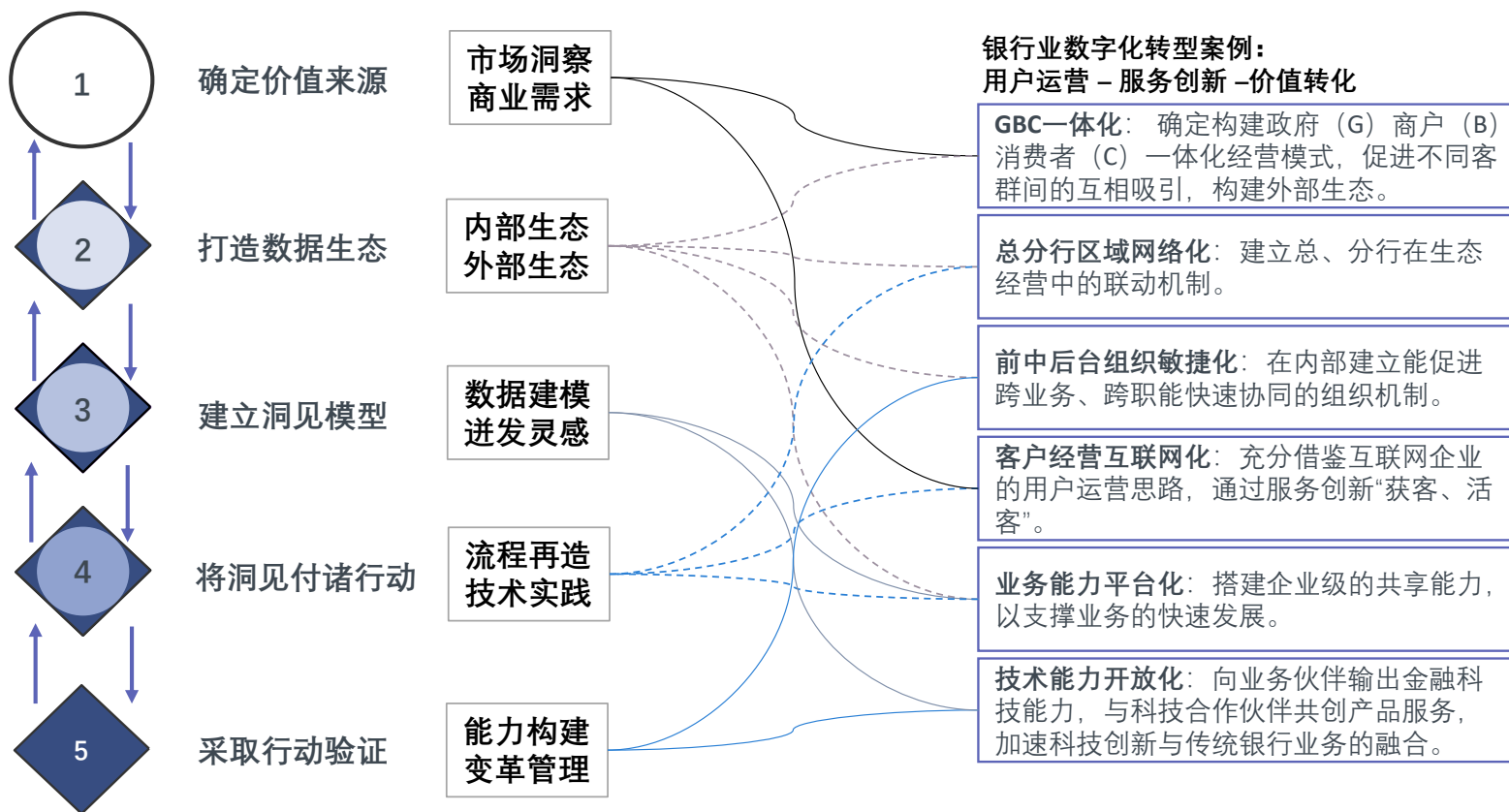
- 大数据管理本质改变了各行业对业务、营销和研发的价值节点实践。数据价值体现在展示现象的发展过程，描述现象的发展本质并对现象发展的趋势作出预测。
- 转化数据推导出相关方向上的预测、趋势分析、交叉销售策略推荐、客群画像匹配，进而掌握客户真正在意的产品和服务的要素。这种数据智能的应用将以数据为生产资料，驱动商业重塑。

价值链变化

- 数据的资产化不仅仅对营销与研发人员产生变化，下游对数据的监测、洞察能够在价值链条的其他功能构成比如仓储、生产计划、人员管理等提供优化建议。同时，数据的远程同步与实时可达性，能够让价值链的任何一个部分都能随时随地调整功能决策。

数据湖和数据仓库通过连结五大数据管理要素和提供数据高级分析的基础，实时支持企业决策，加速了企业数据的价值创造

五大数据管理要素



数据生态圈良性循环

- 数据的连通在横向维度上，一方面在外部能够集结不同行业的一系列客户，另一方面在内部能够促进价值链条上不同环节的紧密协作；
- 在纵向维度上，让生态圈的参与者在角色中增强体验，甚至在其中发挥主导作用。
- 通过网络效应，生态圈能够提供单个企业凭借自身无法提供的产品和服务，从而吸引新的客户，生成更多的数据。

数据湖和数据仓库通过连结五大数据管理要素和提供数据高级分析的基础，实时支持企业决策，加速了企业数据的价值创造

数据湖和数据仓库作为加速器参与大数据管理

数据从各处被采集

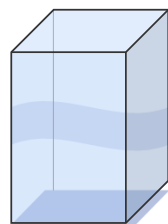
结构化数据

信用卡号码
日期
财务金额
电话号码
...

非结构化数据

网页
电子邮件
社交媒体平台内容
音、视频
...

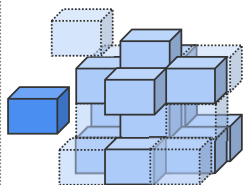
数据进入数据湖以被处理和调用



数据湖

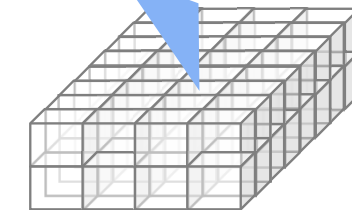
存储是多来源的大量结构或非结构化数据的地点
企业按需处理和分析数据

处理好的数据供高级分析所用



数据应用

数据科学家/数据分析师对处理好的数据进行高级分析



集中式数据仓库

有些数据被移至高性能内存，这些数据申请量大，通常需要后期处理



微服务

数据配置成微小的模块化组件，向数据使用者快速传输实时信息



内部报告

支持内部信息使用者决策



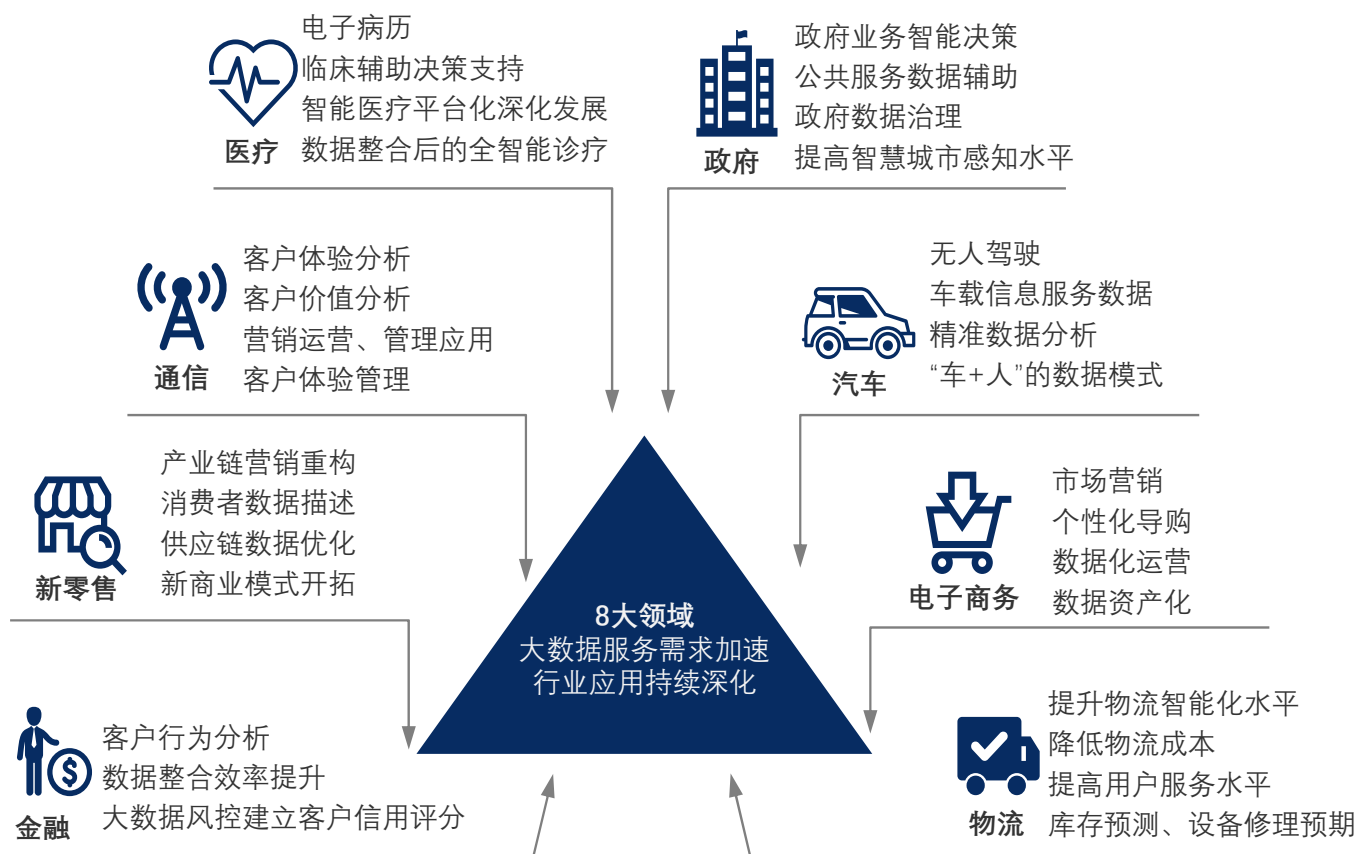
外部报告

直接转化为商业价值

数据湖和数据仓直接加速了数据的采集、存储、计算、分析和调用

伴随数据管理解决方案的底层设施逐渐成熟，技术升级和服务应用场景拓展的加速及深化，数据管理解决方案行业应用逐渐向各领域的核心业务延伸

各领域数据管理解决方案的潜在需求



■ 大数据产业

大数据产业是以数据及数据所蕴含的信息价值为核心生产要素，通过数据技术、数据产品、数据服务等形式，使数据与信息价值在各行业经济活动中得到充分释放的赋能型产业。

■ 衍生态态为各行业提供大数据融合应用

大数据融合应用在各行业、各领域，提供与行业紧密相关的软件应用和整体解决方案。需求逐渐由感知型应用转向预测型和决策型应用。

■ 工业大数据应用深化，推进“中国制造2025”

5G、物联网、边缘计算、区块链等技术逐步融入工业领域，支持企业敏捷管理和精细化运营的大数据管理解决方案推动制造业转型升级提速，更高质量地支撑工业大数据采集、汇聚、流通、分析、应用的价值闭环，助力更高效的产品工艺研发，实现数据价值变现。

■ 金融大数据应用拥有广阔前景

金融行业同时储存了大量的关于客户、账户、产品、交易等结构化数据和反映客户偏好、社会关系、消费习惯等信息的包括语音、图像、视频等非结构化数据。数据管理解决方案在金融业的深度应用，在交易欺诈识别、精准营销、黑产防范、信贷风险评估、供应链金融、投资行情预测等多领域的具体业务中不断释放数据价值。打破金融机构之间数据壁垒，实现跨机构的数据融合将成为趋势。

伴随数字经济建设速度的加快，政府对大数据行业发展重视程度逐渐提升，大数据产业多层次政策体系逐渐完善，数据管理解决方案等细分产业持续受益

数据管理解决方案相关政策整理

颁布主体	政策名称	颁布日期	政策要点
深圳市人大常委会	《深圳经济特区数据条例（草案）》	2020-12	公共数据属于新型国有资产，其数据权归国家所有。深圳市政府代为行使区域内公共数据的数据权，授权市数据统筹部门制定公共数据资产管理实施办法并组织实施。
全国人大常委会	《中华人民共和国个人信息保护法（草案）》	2020-10	确立了个人信息处理应遵循的原则，即强调处理个人信息应当采用合法、正当的方式，具有明确、合理的目的，处理信息应遵循公开、透明的原则。
全国人大常委会	《数据安全法（草案）》	2020-07	第十条 国家积极开展数据领域国际交流与合作，参与数据安全相关国际规则和标准的制定，促进数据跨境安全、自由流动。
工信部	《关于工业大数据发展的指导意见》	2020-04	针对工业大数据现阶段，在加快数据汇聚、推动数据共享、深化数据应用、完善数据治理、强化数据安全、促进产业发展六个方面全盘布局，系统推进。
国家互联网信息办公室	《数据安全管理办法（征求意见稿）》	2019-05	维护国家安全、社会公共利益，保护公民、法人和其他组织在网络空间的合法权益，保障个人信息和重要数据安全。
国家标准化管理委员会	《信息安全技术大数据交易服务安全要求》	2017-12	有助于理清数据交易安全界限，促进数据交易行为合法合规，推动中国数据交易机构的安全建设，促进数据交易行为合法合规，促进全国数据要素有序流通，充分释放数据红利，助力“数字中国”建设。
工信部	《大数据产业发展规划（2016-2020年）》	2017-01	以大数据产业发展中的关键问题为出发点和落脚点，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，打造数据、技术、应用与安全协同发展的自主产业生态体系。
国家发改委	《关于组织实施促进大数据发展重大工程的通知》	2016-01	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。

强调个人信息保护，明确权利属性，数据权立法探索初现

- 数据作为一种新型生产要素首次被写入其中，与土地、劳动力、资本、技术等其他生产要素并驾齐驱。
- 自然人对个人数据依法享有数据权；公共数据属于新型国有资产，其数据权归国家所有；要素市场主体也有数据权，任何组织和个人不得侵犯。奠定了发展数字经济的法理基础，直接要求数据流通市场规范化。
- 未来，可预见面向数据监管、隐私安全、权利归属的解决方案提供商将建立起一条新的数字经济赛道。

加强数据跨境流动的立法，保障国家信息安全

- 中国在建立数据跨境流动管理体系，平衡国家安全、个人隐私保护、产业竞争等多方利益，既要满足全球化经济所要求的数据流动也要满足安全所需对数据的监察控制。
- 国际关系或政治敏感度将成为所有相关企业要重点追踪的维度，直接影响了外资数据管理方案提供商在中国市场上所面临的数据跨境流动监管力度的加大。
- 另外，消费者或企业对于数据认知的提升，也将直接影响其选择同类解决方案的偏好，国产数据管理方案提供商将通过占领部分外企撤出中国市场的份额实现增长。

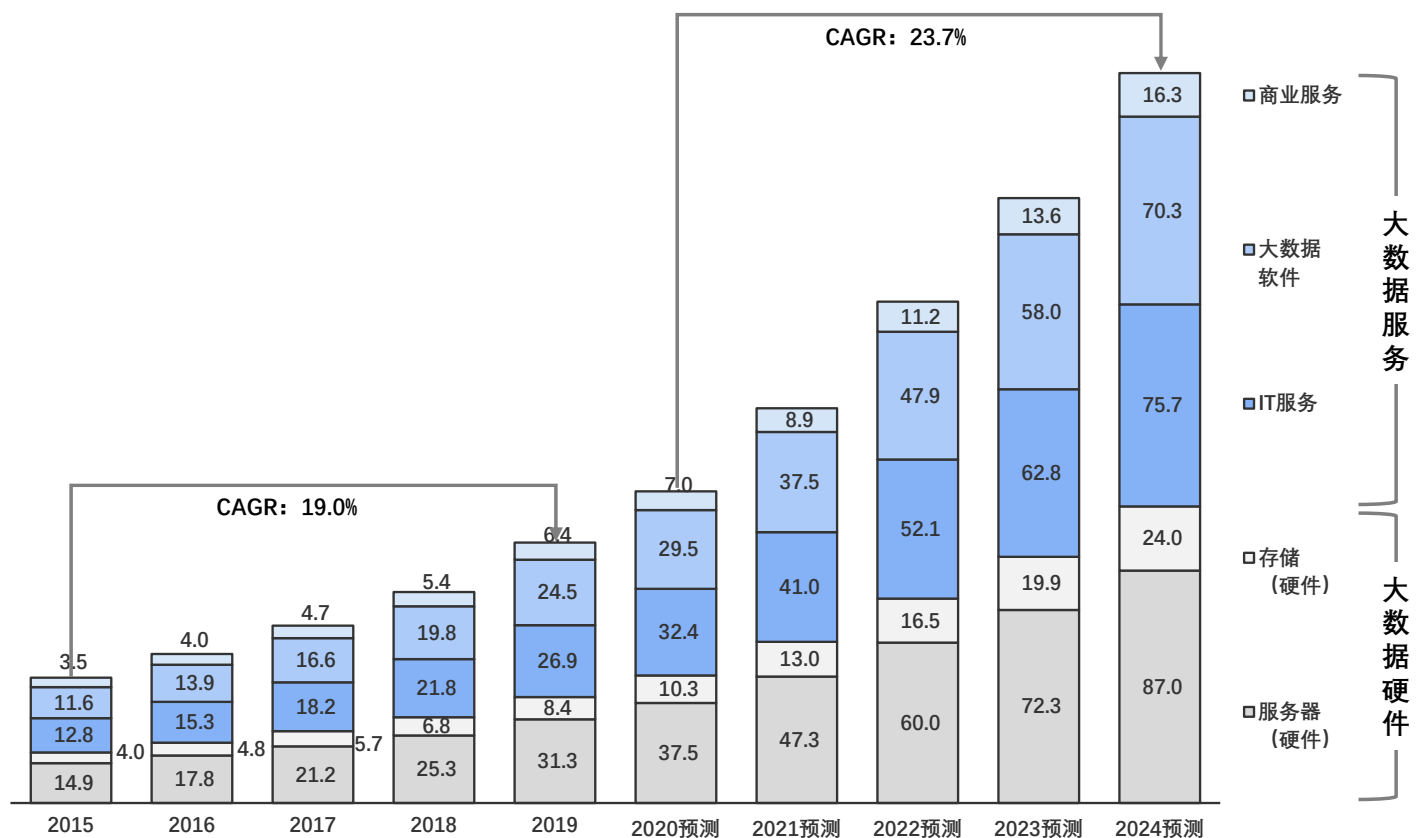
大数据逐渐向各细分应用领域延伸，强调大数据技术与具体应用场景的结合

- “大数据”作为热词早已深入消费者、企业和各地政府的心，回归理性，找到适合企业当前时期且行之有效的数据管理解决方案才是数字经济发展的厦根基。
- 政企合作，服务民生问题，建立跨部门的数据互通，加速审批流程，实现惠民便民的大数据应用将成为服务提供商与开发商的热点，在各个方向重现抗疫期间大数据迸发的应用潜能。

来源：人大常委会、国务院，发改委，工信部，国家标准委，国家互联网办公室，头豹研究院

利好政策的持续颁布以及大数据技术的创新融合推动中国的大数据产业快速发展，未来伴随更多的数据应用场景逐渐落地，数据管理解决方案市场有望持续扩容

中国数据管理解决方案市场规模，2015-2024年预测 【亿美元】



市场需求扩大

- 企业用户将逐渐加大在数据管理解决方案的投入，数据希望实现决策和运营效率的提升以建立竞争优势，这种意愿将快速拉升数据管理解决方案市场需求。
- 数据管理人才逐渐成为数字转型企业的领导者，进一步扩大并巩固对数据管理方案的需求。

数据管理产品趋于成熟，性价比高

- 数据管理市场竞品丰富，产品安全稳定功能齐全，能为各行业各企业的需求提供差异化定制。
- 湖仓一体的发展结合了两个成熟产品的优势，满足灵活性、成本、性能、安全、治理等企业级特性，同时进一步降低整体成本。

国家安全政策利好

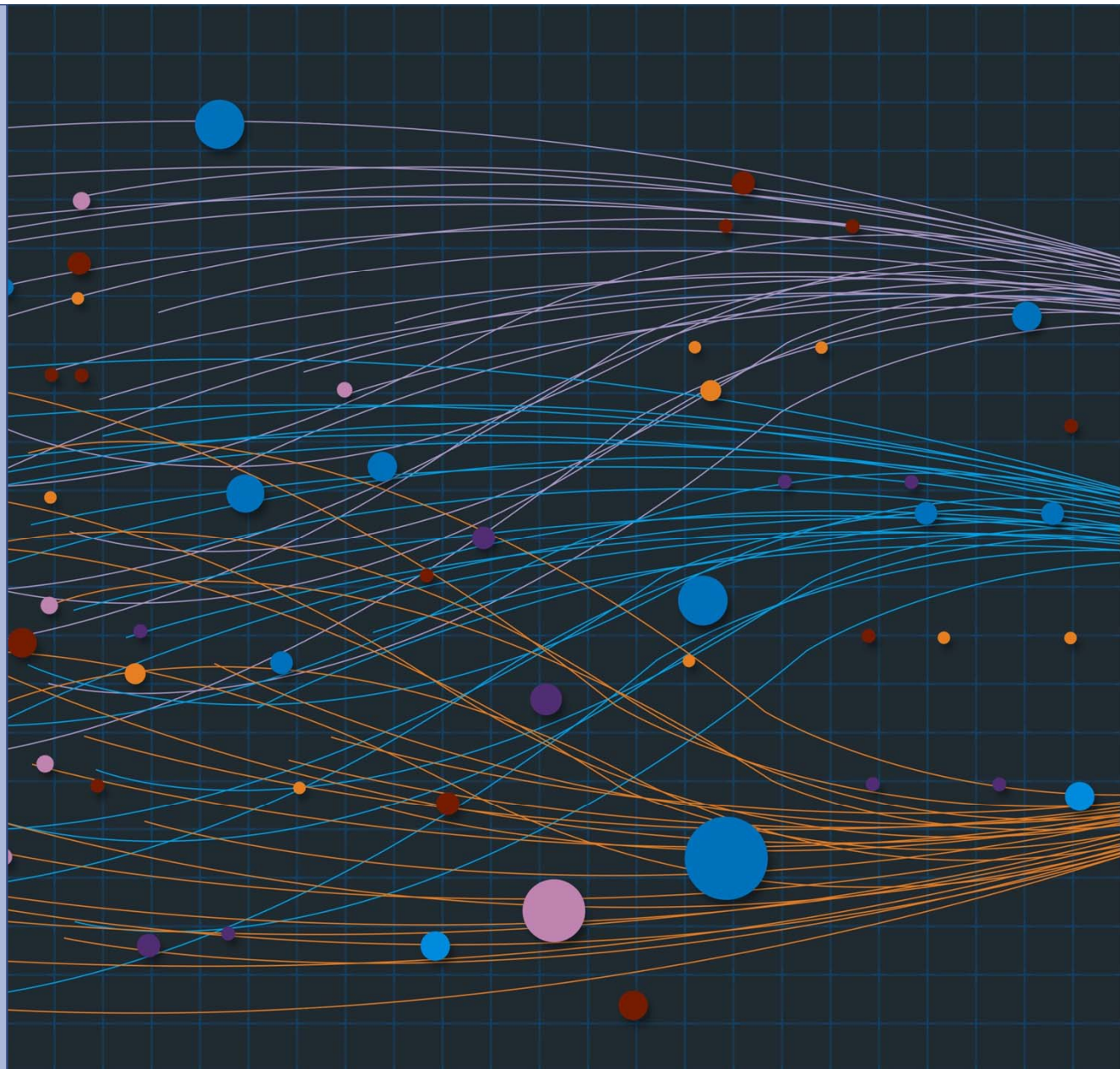
- 随着国家政策在个人信息保护、数据跨境流动、国家信息安全等方面的立法探索，企业与消费者对数据的重视度将进入新的阶段。
- 激活对数据权利、数据隐私、数据安全等话题的讨论，倒逼数据管理解决方案的提供商在数据的数量、种类、速度和价值的产品维度之上，新增安全性作为新的开发方向。
- 同时，对国家数据安全的考虑，中国厂商将得到远高于外资厂商将会得到的市场与政策支持。

数据来源：沙利文

Chapter 3

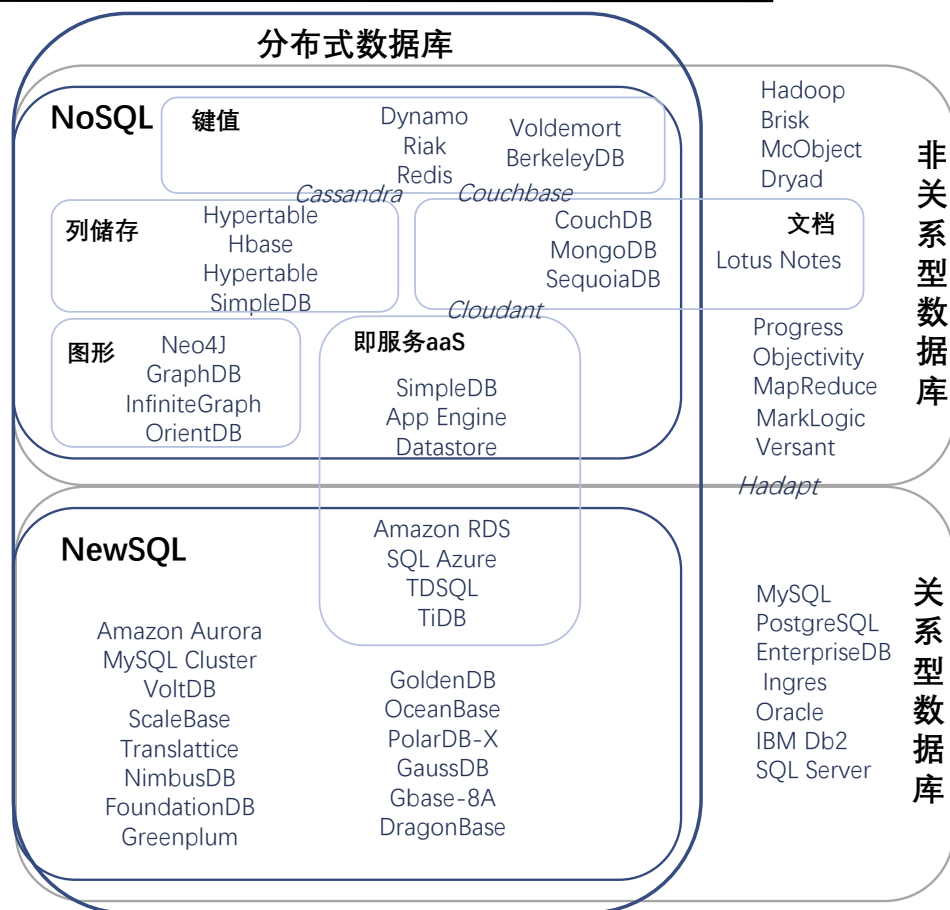
数据库

- 定义与分类
- 发展趋势
- 分布式架构
- 技术洞察
- 行业发展历程
- 产业链分析
- 政策分析
- 产业驱动因素分析
- 市场规模
- 行业风险分析



数据库是按照数据结构来组织、存储和管理数据的“仓库”，按照所使用模型所描述数据在某一时刻的状态或内容的不同，可分为关系型数据库和非关系型数据库。其中不同的架构模式又可分为集中式数据库和分布式数据库

数据库的分类



对应本质

优势

劣势

❑ 采用关系模型来组织数据

❑ 遵循ACID原则

❑ 分布式

❑ 一般不遵循ACID原则

❑ 一种数据结构化存储方法的集合

- **易理解**: 二维表结构是非常贴近逻辑世界一个概念
- **易用**: 通用的SQL语言管理数据库
- **易维护**: 丰富的完整性降低了数据冗余和不一致的概率

- **高读写性能**: 无需经过SQL层的解析
- **易扩展**: 基于键值对，数据无耦合
- **易部署**: 开源，价格优势
- **查询快**: 数据存储于缓存
- **支持基础存储格式**: Key-Value、文档、图片等

- 多表查询机制的限制导致扩展难
- 不擅长大量数据的写入处理、为有数据更新的表做索引或表结构 (schema) 变更、字段不固定时的应用、对简单查询需要快速返回结果的处理

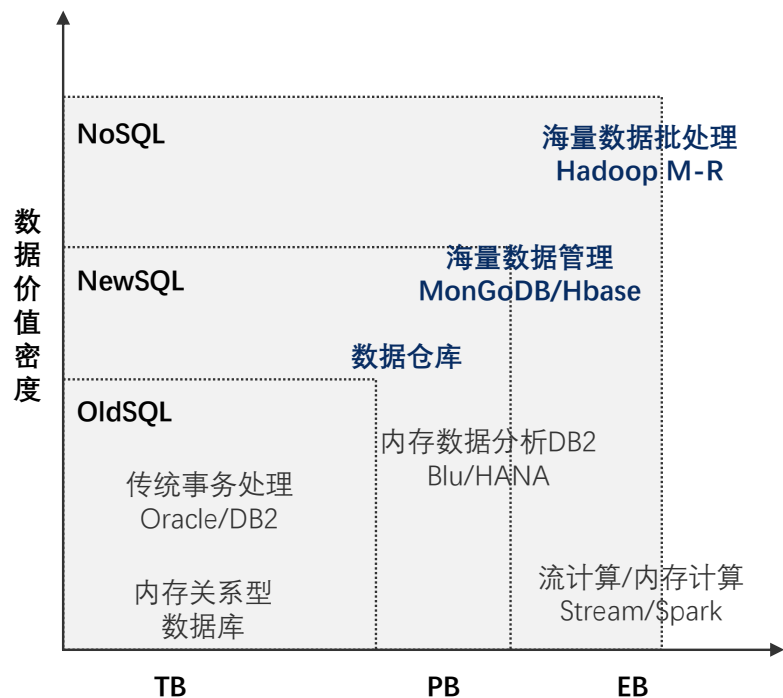
- 维护的工具和资料有限
- 不提供SQL支持，学习和使用成本较高
- 不提供关系型数据库对事物的处理

来源: 451 Group, 头豹研究院

来源: 头豹研究院

用于支持企业价值创造的数据管理在处理数据的种类、数量和速度上均呈现爆发式增长。数据量级和价值密度的增长将催生对新一代数据库的需求

三类软件产品细分的关键差异



数据上云成为趋势

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。

非关系数据库比重提升

数据增速集中于影音文件、社交信息等非结构化数据。

内存数据库更广泛应用

相较于磁盘储存，内存能够满足对信息互动性需求的高并发、低延时、快速读取的特点。

流数据库快速发展

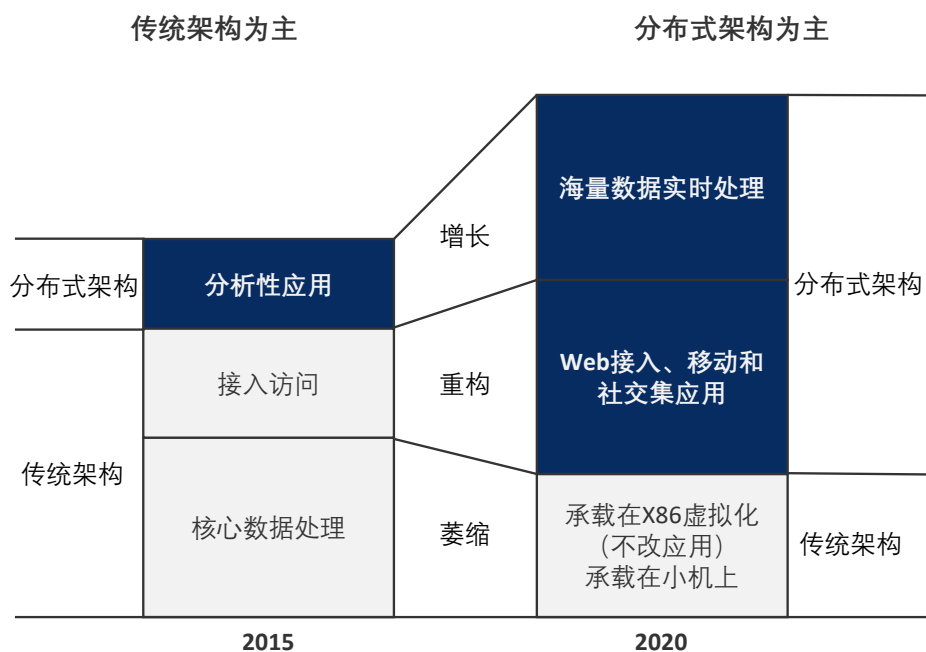
结合事务处理和实时分析，应对海量数据涌入时，仍保持实时响应和低延迟。

开源数据库生态不断完善

如MySQL、PostgreSQL、MongoDB等开源数据库，正在以价格低廉、性能相等、功能丰富的特点在中小型企业中占领商业数据库的市场。

以数据分析、数据流通及事务处理技术为核心的大数据技术体系日渐完善，数据库技术和数据服务融合程度加深，助力企业用户数据价值释放

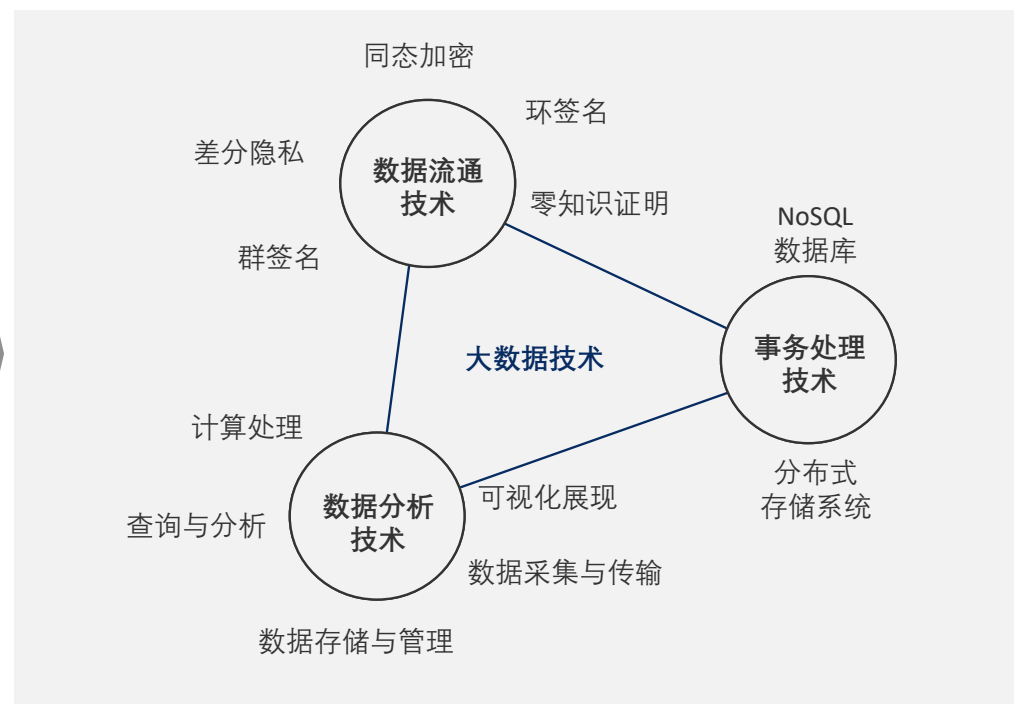
数据库架构变革



性能及可用性依赖于硬件

性能及可用性依赖于扩展能力及软件系统

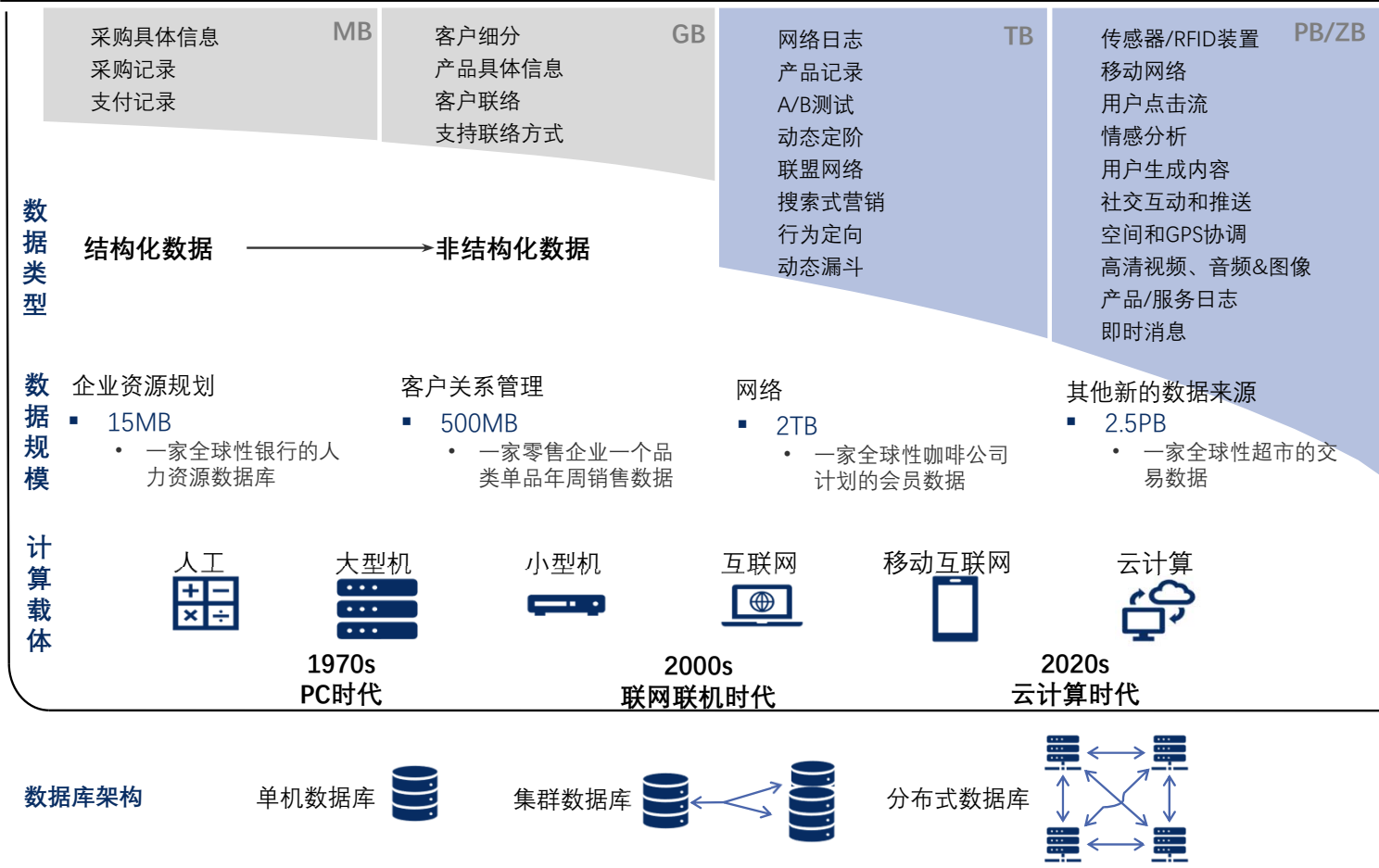
大数据技术融合发展



传统大数据技术难以应对数据量较大的非结构化数据，持续增长的数据处理需求推动大数据处理系统和分析技术的快速发展

随着计算载体的迭代与业务需求的相互促进，传统数据库已落后于计算场景的需求，让步于分布式数据库，分布式数据库凭借资源池化管理与兼容多类型接口的数据服务能力满足了云计算场景的需求

数据属性、计算载体与数据库架构的演进



■ 数据属性正在经历代际剧变

随着井喷式的需求进程，数据处理所面对的数据类型、数据规模、计算速度的要求趋于严苛。云已经领先传统终端成为了最重要的计算场景，而分布式架构也正在替代传统架构实现海量高效的数据处理。

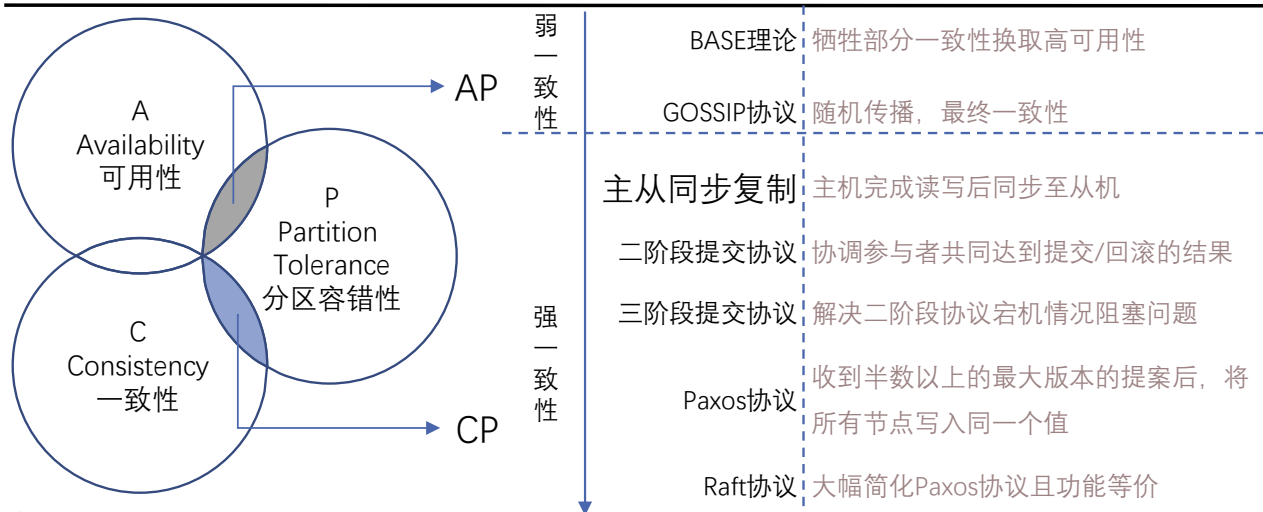
■ 云计算场景推动分布式数据库的发展

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

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

从以2PC为代表的写一致性协议过渡到以Paxos为代表的多数派一致性协议，可应用的分布式事务处理技术走向成熟；分布式数据库以多副本机制来保证系统的安全可靠，而多副本带来的一致性问题的解决则让以Raft为代表的一致性协议成为关键

事务一致性属性及实现思路



来源: 头豹研究院

2PC vs. 3PC vs. Paxos vs. Raft

	2PC	3PC	Paxos	Raft
数据一致性	2	2	3	3+
安全性	1	1+	2	3
容错性	1	2	3	3+
低复杂性	3	3	1	2
高并发性能	2	1	3	2
低维护成本	3	2	1	1

1-3 为推荐分, 更推荐得分高者

来源: 51CTO, 上海交通大学, 纽约大学, 沙利文

一致性的定义

在数据库系统中通常用事务来保证数据的一致性和完整性。在分布式系统中，数据一致性往往指的是由于数据的复制，不同数据节点中的数据内容是否完整并且相同。

- 弱一致性：异步复制，不能保证任何一次读都能读到最近一次写入的数据。
- 强一致性：同步复制，任何服务器上的读操作总是能看到最新写入的数据。

实现强一致性的底层路径

- 写一致性：在写入数据的同时将数据复制到其他服务器上；
- 读一致性：写入后不再复制，在读取时使用版本来协调复制；
- 多数派一致性：引入数据版本号判断并读取最新写入的多数派数据副本。

CAP理论

CAP理论证明了，任何分布式系统只可同时满足以上两点，无法完全兼顾三者。而分布式系统都需要满足分区容错性，必须在一致性和可用性之间进行权衡。

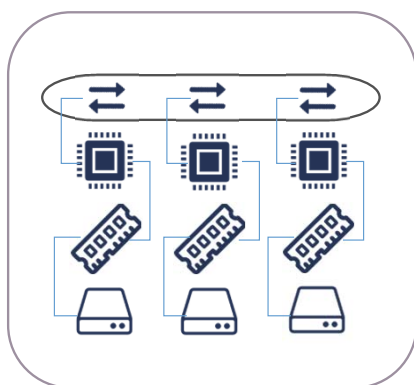
- AP系统：选择了高可用和分区容错性，此时，那个失去联系的节点依然可以向系统提供服务，不过它的数据就不能保证是实时同步的。
- CP系统：要求数据在各个服务器上强一致的，然而网络分区会导致同步时间延长，可用性得不到保障。适合对结果一致性非常敏感的金融级数据库，保障钱财安全是最大的应用场景。

在高可用性、带宽要求低、平滑扩展能力与远程部署能力上的优势决定了无共享架构将成为未来的存储应用主流

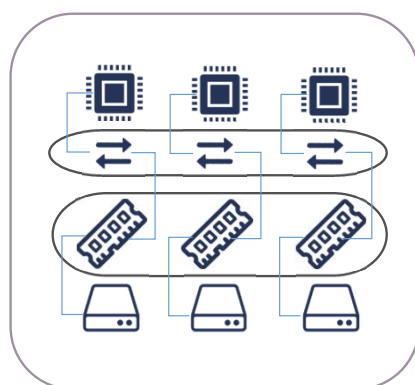
数据库集群架构分类

SN vs. SM vs. SD

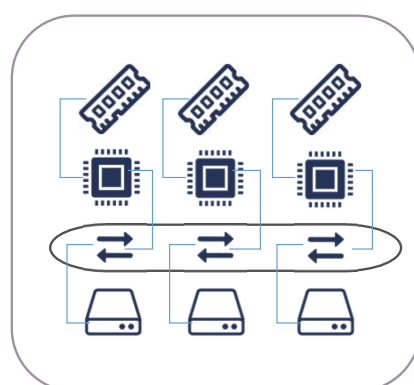
无共享架构
Shared nothing (SN)



共享内存架构
Shared memory (SM)



共享磁盘架构
Shared disk (SD)



- 定义**
- 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**
- 无限的水平扩展
 - 无单点，可用性高
 - 业务侵入性小

- 劣势**
- 实现较为复杂
 - 在某些场景下需要重新思考业务设计

- 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

- 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

- 内存访问干涉与网络带宽竞争问题
- 扩展能力有限

- 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

- 快速适应变化的负载
- 高可用
- 数据不需要分区

- 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

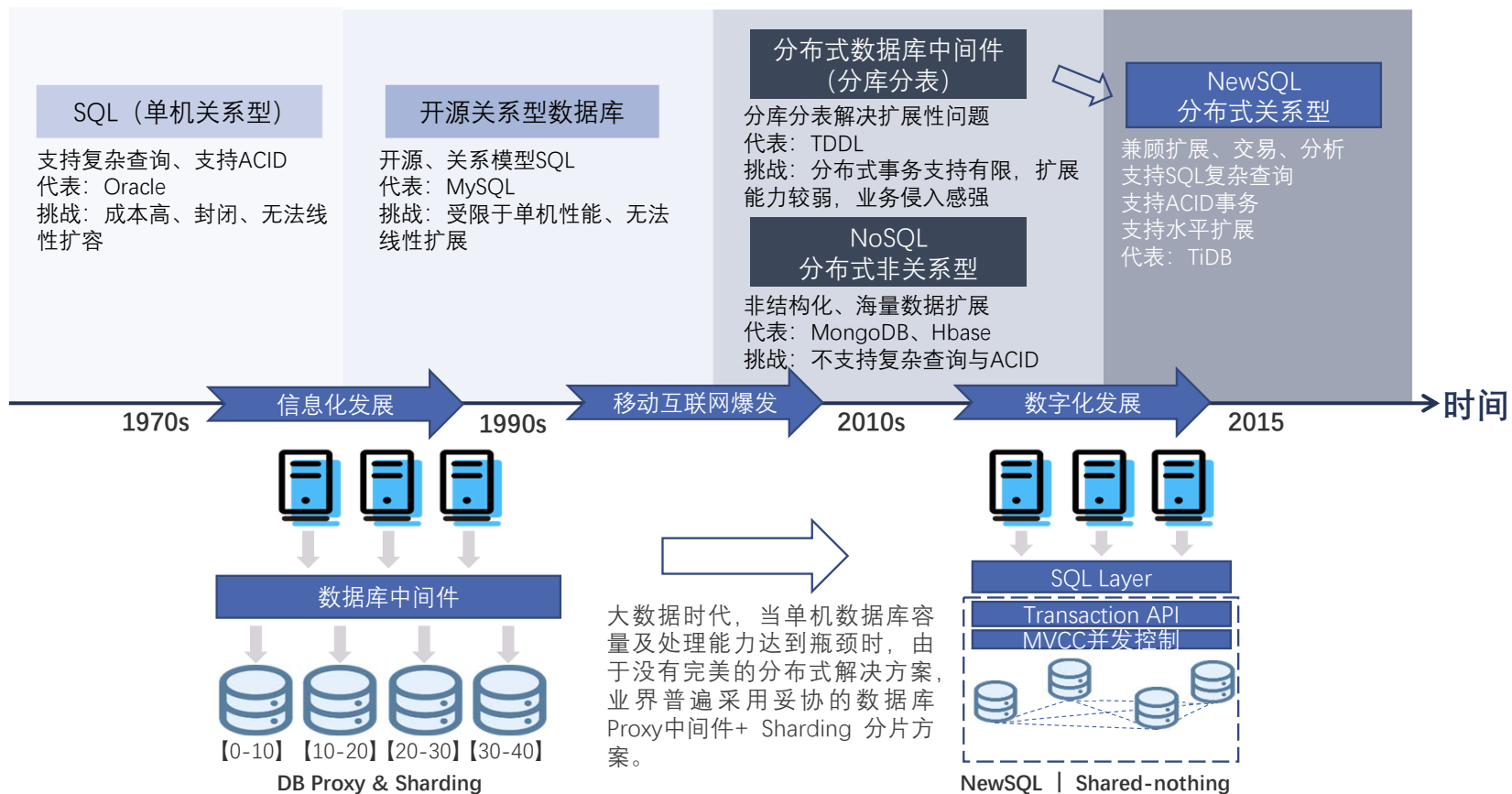
	SN	SM	SD
并行控制能力	2	2	1
灾难恢复能力	2	3	1
数据库设计易度	1	2	2
负载均衡能力	1	3	2
高可用性	3	1	2
消息数量处理	1	3	2
带宽要求	3	1	2
横向扩展能力	3	1	2
远程部署能力	3	1	2
关键会话敏感性	3	1	2
系统镜像数量	1	3	1
热点敏感度	1	1	1

来源: Oracle, ScaleDB, 韩国工业科学院RIST, 加州大学伯克利分校, 沙利文

1-3 为推荐分, 更推荐得分高者
来源: 加州大学伯克利分校, 沙利文

NewSQL将事务型数据库的ACID保证与NoSQL的可扩展性与高性能相结合，完美解决了MySQL的扩展问题

NewSQL的演进路径



■ NewSQL

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。

来源: PingCAP, 头豹研究院

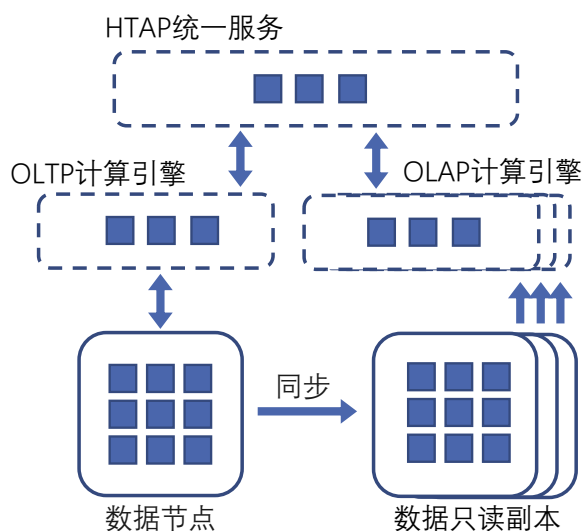
HTAP解决了数据库面对海量数据库下的存储与并发处理的扩展性问题，实现在线高并发OLTP及OLAP海量数据分析

传统数据库应用架构数据流



来源：卡内基梅隆大学，PingCAP，头豹研究院

分离架构HTAP简要原理



来源：GBASE, PolarDB-X, 头豹研究院

OLTP vs. OLAP vs. HTAP

	OLTP	OLAP	HTAP
响应速度	3	2	3
吞吐量	1	2	3
并发访问量	2	1	3
事务大小	1	2	3
数据访问量	1	2	3
索引规模	2	3	3
实时性	3	1	2
物化视图	1	3	3

1-3 为推荐分，更推荐得分高者

来源：卡内基梅隆大学，头豹研究院

■ 传统数据库应用架构存储属性

OLTP：将不同属性连续存储，即行存储，一条数据的所有属性是连续存储的，可以使得插入/更新/删除更快。

OLAP：将不同数据的同一属性连续存储，即列存储，使得查询操作只读关心的数据属性，而不是一整条数据，减少浪费；可以更好地支持复杂查询。

■ 传统数据库应用架构的问题

OLAP 和 OLTP 系统间会有分钟/小时级的时延，且OLAP和OLTP数据库之间的一致性无法保证，难以满足对分析的实时性要求很高的业务场景。

高可用冗余的副本数量大且分散，ETL数据异步传输链路运维复杂度高，副本同步和运维的难度和成本高。

软件开发团队需要为不同的数据库编写查询语句，且有可能需要将不同系统的数据进行聚合，开发成本高。

■ 分离架构的HTAP实现收敛技术栈

低成本，底层数据可快速复制，并且同时满足高并发的实时更新。

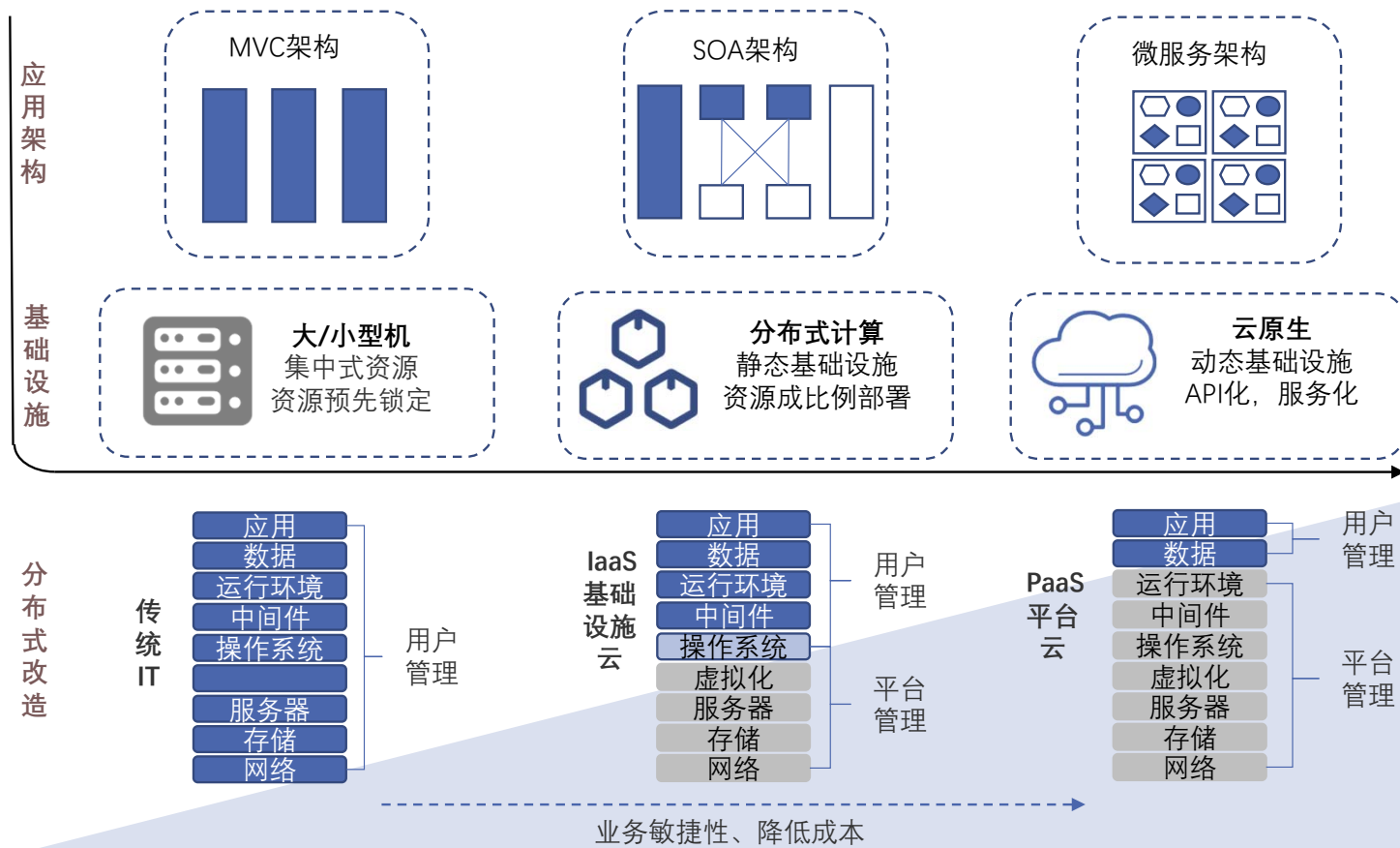
块连续存储，行存储频繁被修改的热数据，列存储需要查询和分析的冷数据。

大规模多级并行处理能力MPP，以无共享架构集群具备线性扩展能力。

资源隔离，提供AP、TP资源链隔离机制，避免相互影响。

分布式数据库满足云原生需求，在微服务化应用开发及云化平台的趋势下，混合云与私有云部署将成为长期暂态

云化与微服务化的演进



■ 云化基础设施改造

现阶段数据库基于X86环境的分布式架构替代大/小型机环境的集中式架构，并开始逐步迁移各类应用到云平台上，实现数据与业务分离，在保持接口兼容性和数据一致性前提下提供分布式存储与服务能力。

■ 云化与微服务

微服务架构是一种将单应用程序作为一套小型服务开发的方法，以使每个应用可独立地进行开发、管理和加速，小团队敏捷交付缩短周期，运营成本大幅下降。随着敏捷开发、持续支付、DevOps理论的实践，以及Docker等LXC（轻量级容器）部署的成熟，微服务架构开始流行。

在微服务化应用开发以及云化平台的趋势下，云分布式数据库调用云基础设施资源将不存在限制，满足了上层应用的弹性扩展、高并发、高吞吐量与灵活敏捷的要求。

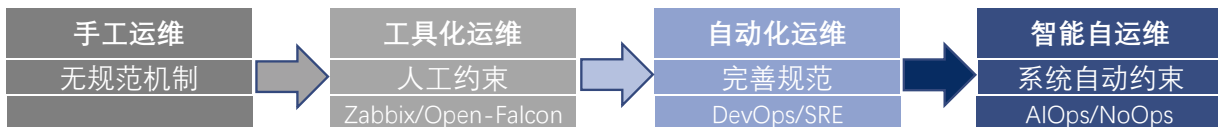
■ 云化趋势

在中国金融级需求中，处于核心业务对于数据主权、安全隐私的考量，混合云与私有云部署已成为金融级需求上云的长期暂态。

来源：华为云、PingCAP、中兴通讯，头豹研究院

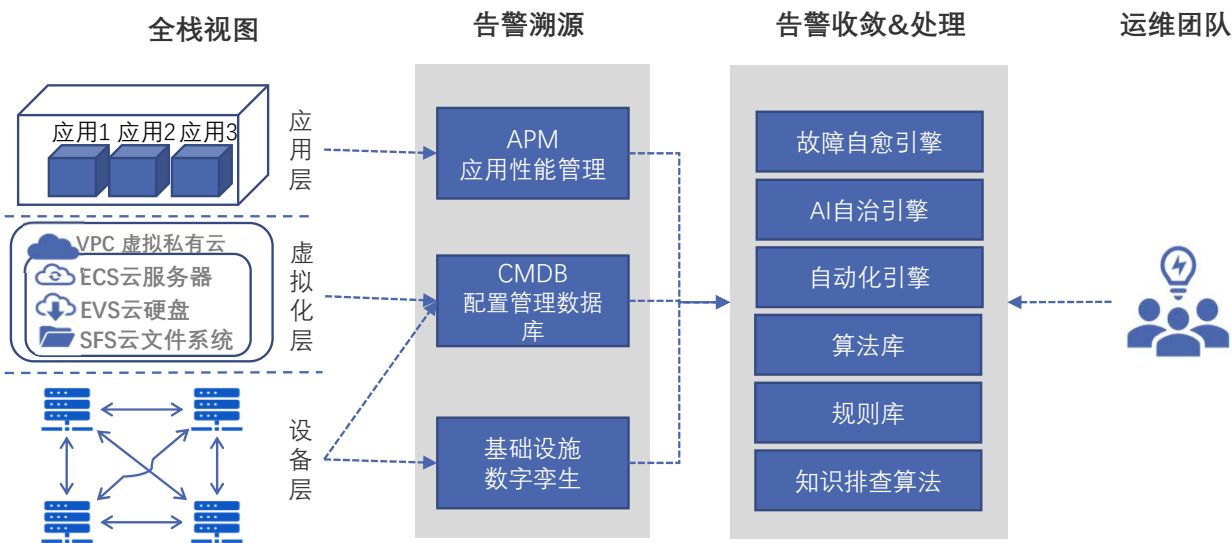
运维云化与智能化是适应海量计算环境中保障数据库安全高效得运行的必然发展路线

运维演变历程



来源：运维派、阿里云，头豹研究院

智能运维



来源：华为云、阿里云，沙利文

■ 传统IT运维

在手工运维和工具化运维的阶段中，运维服务是由技术驱动的，服务形式单一，多以机房或数据中心的IT基础架构为范围，实行响应式、救火式的被动工作模式，自动化程度低，人员成本高。

■ 云运维

运维上云是由客户需求驱动服务，服务形式从机房拓展到桌面端、移动端、云端，实行主动式、自动化的工作模式，人员成本低。对比传统IT运维，极大降低了对运维人员的依赖，从故障后处理进阶到故障前规避。

■ 生产力工具适配生产力发展

用户增长与企业投入系统建设维护的成本呈非线性关系，仅增加运维人员的数量无法满足系统发展需求。为了提高IT的整体效率与质量，总结重复、可追溯的现象，形成规则，完成自动运行维护成为发展的必要。

■ 技术欠账

“技术欠账(Technical Debt)”是IT运维团队在代码、技术文档、开发环境、第三方工具、应用冗余和开发实践方面，阻碍变革、降低效率的不足之处。技术欠账可以借助上云实现清晰化，在云上运维中把AI嵌入最关键的业务流程，AIOps通过自动收集和分析基础架构、应用程序和云服务等指标，以及日志和事件等原始数据，通过机器学习分析并从技术角度识别问题。

AI算法替代启发式算法，通过学习历史数据，根据现状在运行时进行动态调整，解决传统数据库的痛点

AI赋能数据库对比

	调优时长	训练时间成本	工作质量	适用性	硬件适用性
人工调优	数日-数周	数月 (训练DBA)	低	X	X
基于搜索	数小时	数小时	低	X	X
传统机器学习	数秒	数小时	中+	✓	X
强化学习	毫秒	数小时-数日	高	✓	✓
深度学习	毫秒	数小时-数日	中	✓	✓

	数据场景	质量	适用性
线下索引选择	静态	低	低
线上索引选择	动态	中	中
半自动索引选择	动态	高	中
基于机器学习的索引选择	动态	高	高

■ 传统数据库

在传统数据库中，大量地使用启发式算法通过传统经验去优化数据库，已然无法满足更高的执行要求，比如针对众多用户的实际场景进行定制化开发。过去采用系统预定义参数组合或可调节参数开关等方式，由DBA根据经验进行调整。

■ AI for DB (AI4DB)

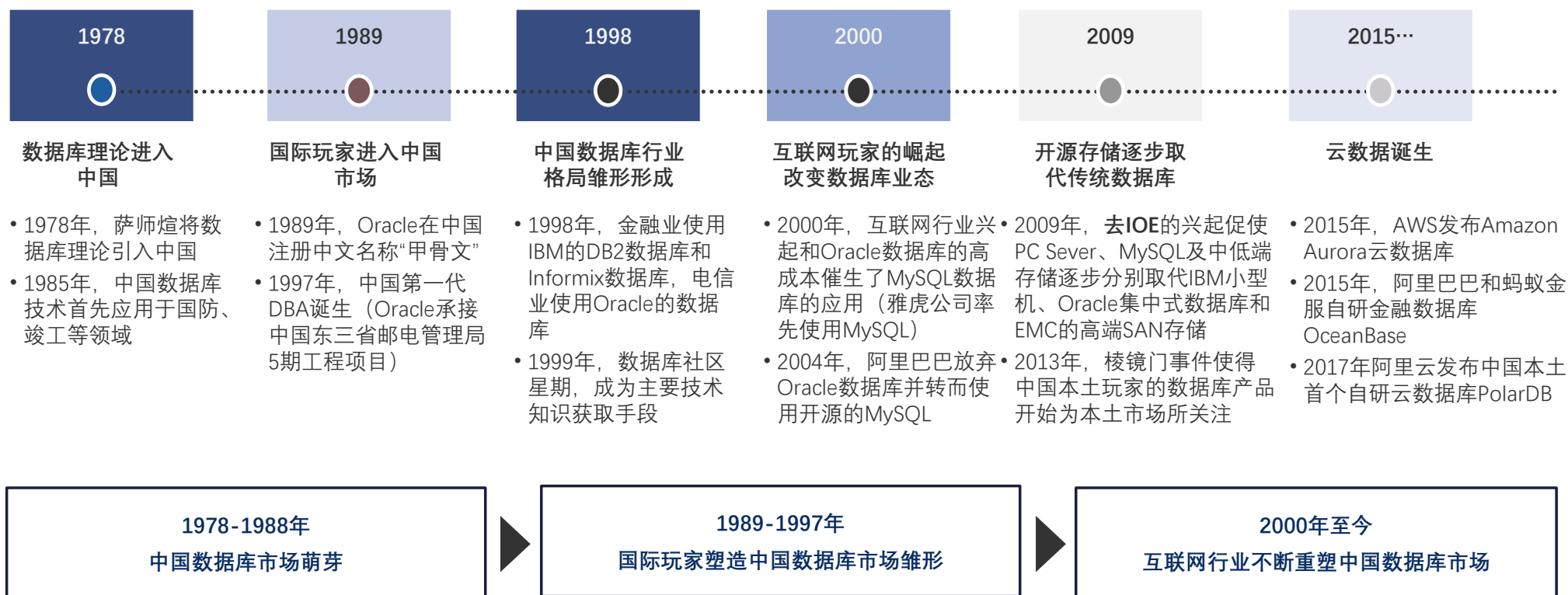
利用AI算法替代启发式算法，解决传统数据库的痛点，典型方向有：

- **优化器**：传统代价优化基于采样统计信息进行表连接规划，存在统计信息不准、启发是连接规划等问题。
- **参数调优**：数据库有数十至上百个可调节参数，其中很多参数时连续值调节空间，依靠人工经验无法找到最优参数组合。
- **自动化索引推荐和视图推荐**：在数据库的众多表与列中自动构建索引和视图，来提升数据库的性能。
- **事务智能调度**：事务的并发冲突时OLTP的难点，可以通过人工智能进行智能调度来提升数据库的并发性能。

来源：清华大学、头豹研究院

中国数据库市场发展至今经历了三个阶段：从中国数据库市场萌芽，到国际玩家塑造中国数据库市场雏形，再到互联网行业不断重塑中国数据库市场

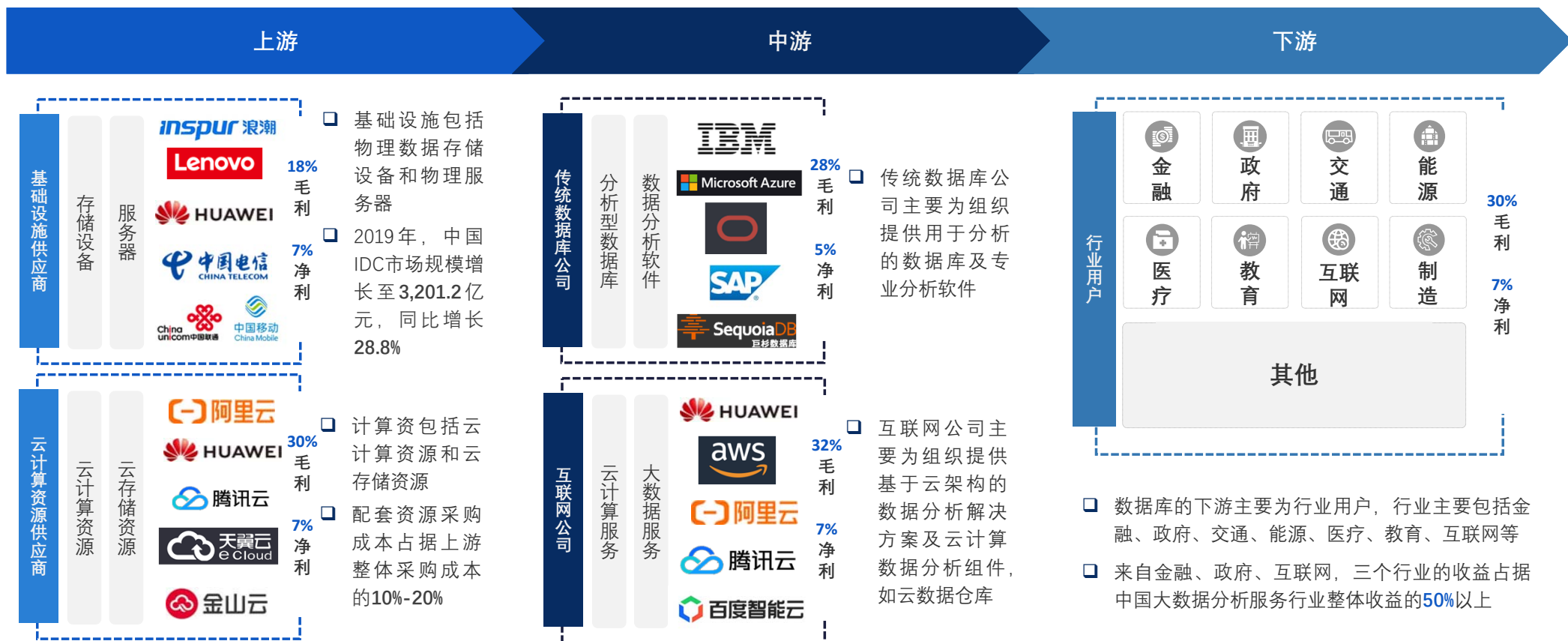
中国数据库行业发展历程



来源：头豹研究院

中国数据库产业链上游主要市场参与者为基础设施提供商和云计算资源提供商，产业链中游参与主体为数据库提供商，下游端则主要为各行业用户

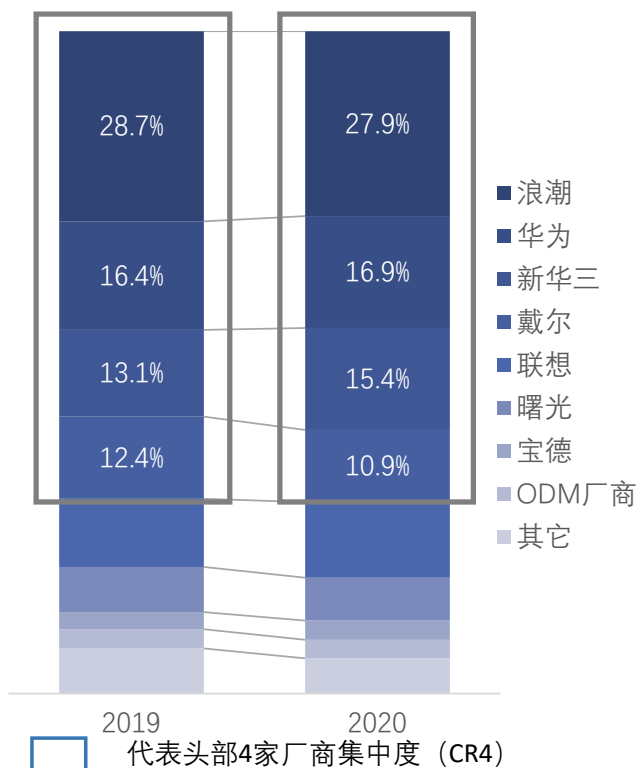
中国数据库产业链



中国数据库行业上游供应商的议价能力较高，尤其X86服务器厂商具有较强的议价能力

中国X86服务器提供商竞争格局，2019-2020年

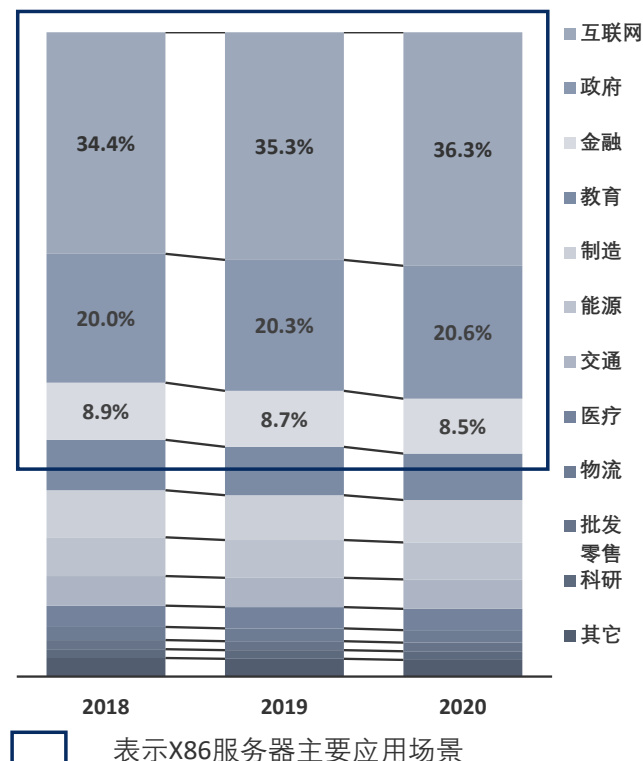
[百分比]



- 供应商的集中程度较高：头部4家厂商集中度CR4由2018年的70%上升到2019年的70.6%（均大于30%），已经形成寡头垄断市场
- 供应者前向一体化的可能性较高：数据库硬件与软件虽然有本质区别，但软件的高利润对上游硬件供应商有着较大吸引力

中国X86服务器市场需求结构，2018-2020年

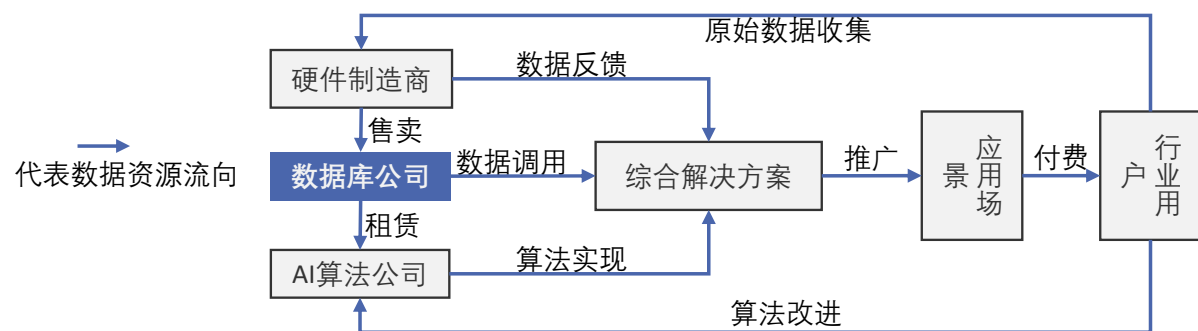
[百分比]



- 供应品的可替代程度较低：目前市场中尚未出现数据库所需原材料的替代产品
- 数据库产品对于原料供应商重要性较小：数据库的需求占用户需求比重约为1.3%，对原料供应商来说业务比重较小
- 供应品的特色和转变费用较低：数据库产品依照不同规格标准进行生产，同一标准规格产品同质化程度较高

数据库运营商正通过联合硬件和AI算法公司增强数据资源的流动性。开源是发展开发者群体和用户群体以及放大技术影响力的最佳选择，是生态推进的重要手段

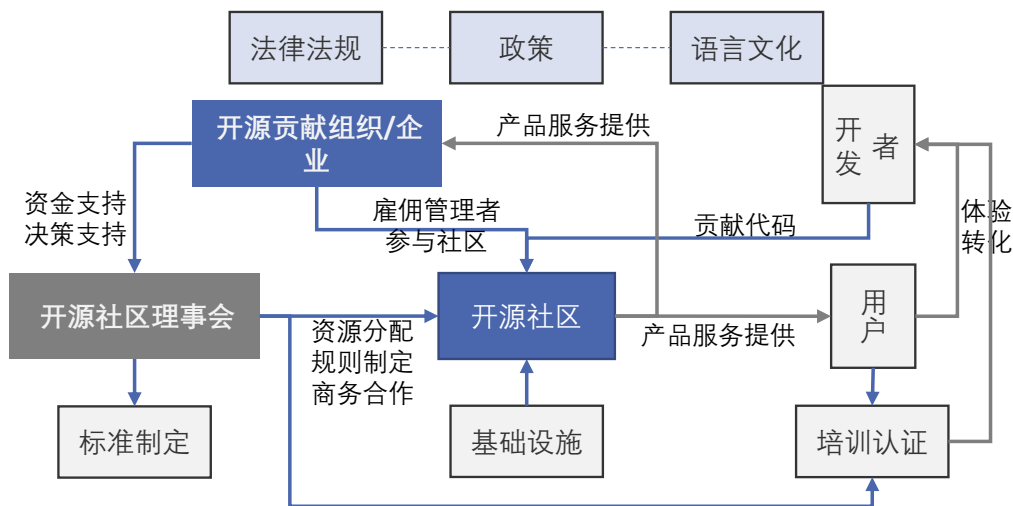
供给端数据资源流



■ 数据库运营的商业模式

主要是来自于硬件厂商提供数据，而数据库公司与AI厂商联合，针对特定的应用场景而开发综合解决方案，从而提供给下游用户使用，而用户的付费则是数据库公司的主要收入来源。

开源社区构成



■ 开源生态吸引产业链上下游协作

开源数据库能够吸引包括ISV独立软件应用开发商、中间件开发商、OS操作系统开发商、服务器/底层硬件提供商的产业链上下游厂商参与，通过免费提供开源版本，企业获取了用户体验的验证，也形成了对外部开发者持续的吸引力。

■ 盈利模式

MySQL是主流的开源数据库架构，被Oracle收购并控股，Oracle同时作为全球领先的数据库厂商正控制着MySQL的竞争力，稳固自家付费产品和服务的地位。相似的，开源数据库厂商大都保留了具备企业级特性的分布式架构内容转而以付费项目根据具体需求提供更优质的服务，在开源社区中实现提供商和用户的双赢。

中游数据库厂商关键成功因素在于关注技术、产品质量和政策关系，加大此三方面的投入可以使数据库厂商迅速获得比较优势

中游数据库厂商关键成功因素分析

	总分	技术	销售	市场推广	品牌	物流	售后服务	采购	产品成本	产品质量	资金	政策关系	生产能力	人力资源
技术	20	1	2	2	1	2	2	2	1	1	1	2	1	2
产品质量	18	1	2	2	2	2	2	1	1	1	1	1	1	1
政策关系	17	0	2	2	1	2	2	1	1	1	1	1	1	2
资金	16	1	2	1	1	2	2	1	1	1	1	1	1	1
产品成本	16	1	1	1	2	2	2	1	1	1	1	1	1	1
人力资源	15	0	2	2	1	2	2	1	1	1	0	1	1	1
生产能力	14	1	1	2	1	1	1	1	1	1	1	1	1	1
采购	14	0	1	1	1	2	2	1	1	1	1	1	1	1
销售	12	0	1	2	2	2	2	1	1	0	0	0	1	0
品牌	11	1	0	1	1	1	2	1	0	0	1	1	1	1
市场推广	9	1	0	1	1	2	2	1	1	0	0	0	0	0
售后服务	6	0	0	1	0	2	1	0	1	0	0	0	1	0
物流	4	0	0	0	1	1	1	0	0	0	0	0	1	0

如果纵列因素比横行因素重要则打2分，同样重要打1分，不重要打0分

来源：沙利文

▣ 技术

由于数据库产品的制造和研发特点导致，数据库行业的技术壁垒明显，各数据库厂商也纷纷通过加大研发力度来获得更高的数据库产品竞争力。

▣ 产品质量

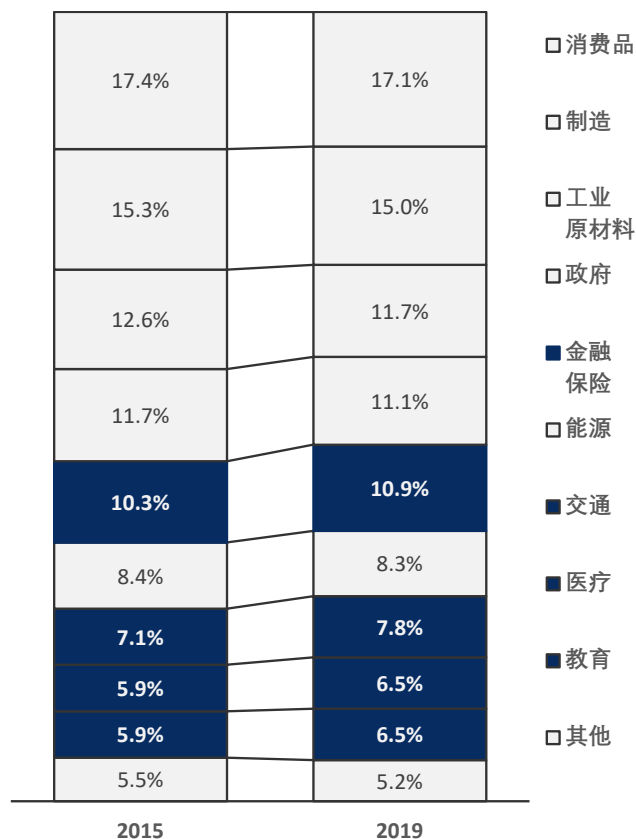
数据库产品质量将直接影响行业用户管理效率和价值创造，尤其在对于数据资源依赖度较高的互联网、金融保险、消费品等领域。

▣ 政策关系

中国政府对于计算机软件行业存在明显的政策指引，并且某些数据库的应用会涉及数据安全问题，因此数据库产品在监管上会受到明显的宏观政策影响。

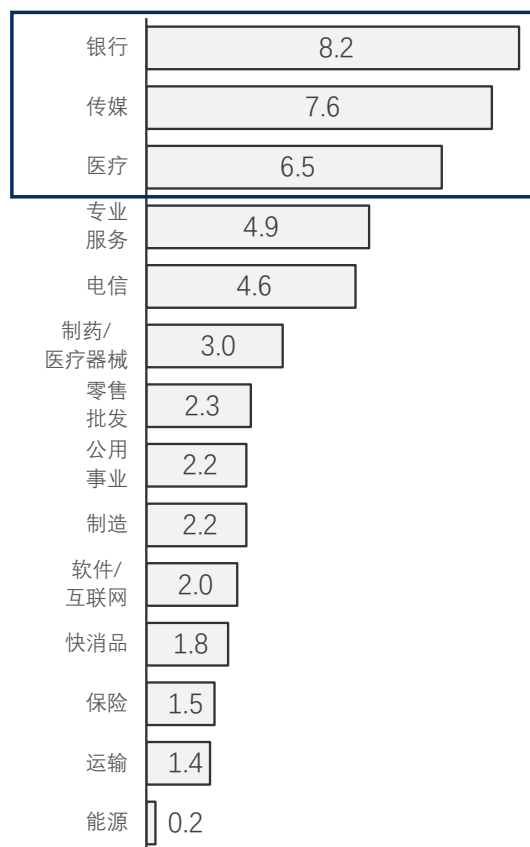
中国数据库市场下游行业用户的议价整体能力较弱，但金融保险、交通、医疗和教育等细分行业用户议的价能力正在提高

中国分析型数据库需求结构 [百分比]



来源：Wind，头豹研究院

中国分析型数据库收入-数据乘数 [GB/美元]



收入-数据乘数即每一美元收入所需数据量
来源：Wind，头豹研究院

下游各应用场景参与者的集中程度较低：

目前中国数据库产品应用较为分散，2019年占比最高的行业用户集中在消费品，为**17.1%**

对数据库的可替代程度适中：

云数据库对传统部署数据库的部分替代替代（较传统部署数据库在成本和拓展性方面更具优势）

云数据库在部分利基市场应用程度加深：

金融保险、交通、医疗和教育等细分行业用户比例在上升

数据库行业下游厂商转换成本较高：

数据库的重新部署需重新定义Schema及进行数据迁移，成本较高

数据库产品对行业用户影响较大：

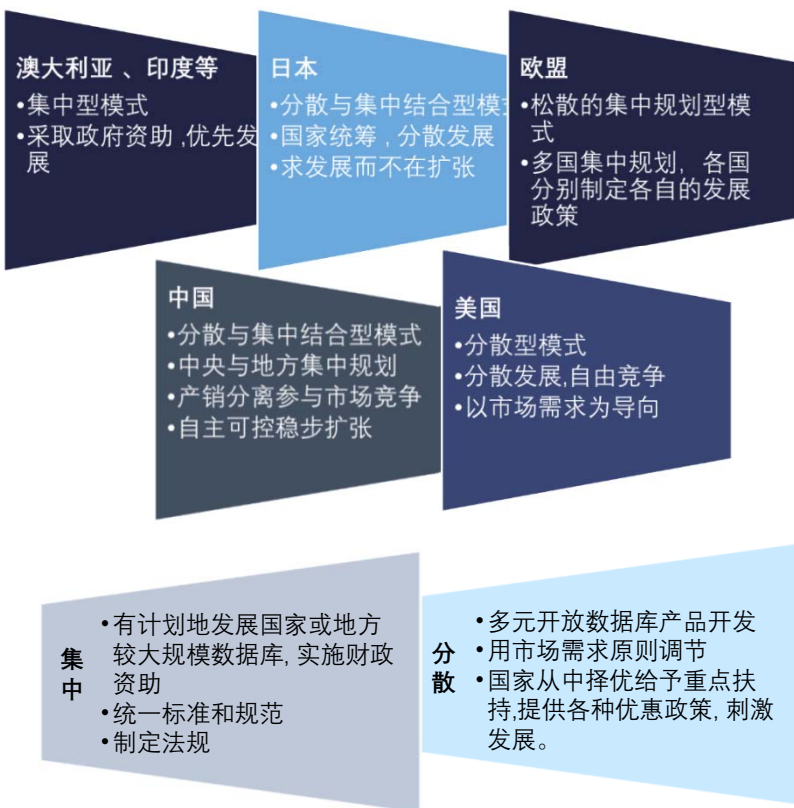
大数据时代背景下，数据库产品将直接影响数据依赖度高行业的价值实现（如银行、传媒和医疗，收入-数据成数超过均超**5GB/美元**）

下游厂商后向一体化的可能性较低：

下游应用厂商所涉及领域与数据库产品实属不同领域

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发展与建设方针



数据库产业相关政策

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: 中共哈尔滨市委党校、北京大学、头豹研究院

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

数据库产业驱动因素

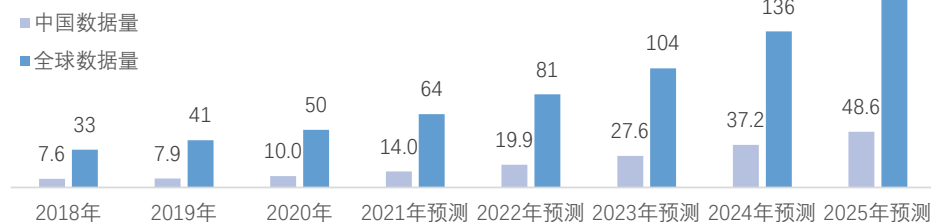
企业数据规模激增，中心化架构不堪重负

因素一

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

企业业务在线化，数据处理时效性要求大大提升

因素二

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。

IT人工成本逐年提升，全托管的价值日渐显现

因素三

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，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。

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

数据库产业驱动因素

IT基础设施上云，在云上处理数据成为必然

因素四

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。

应用架构往弹性化发展，数据平台也需要相应的变化

因素五

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。

开源重塑产品迭代模式与市场推广模式

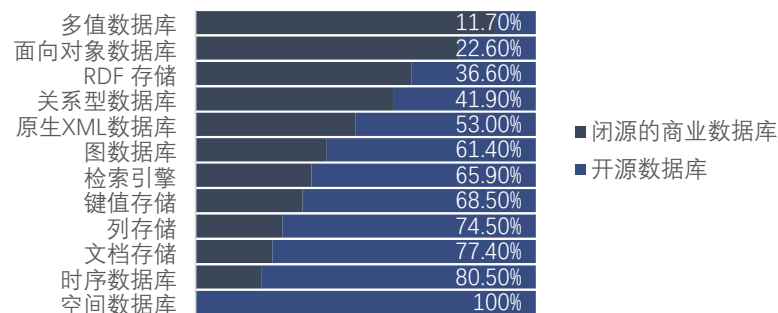
因素六

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。

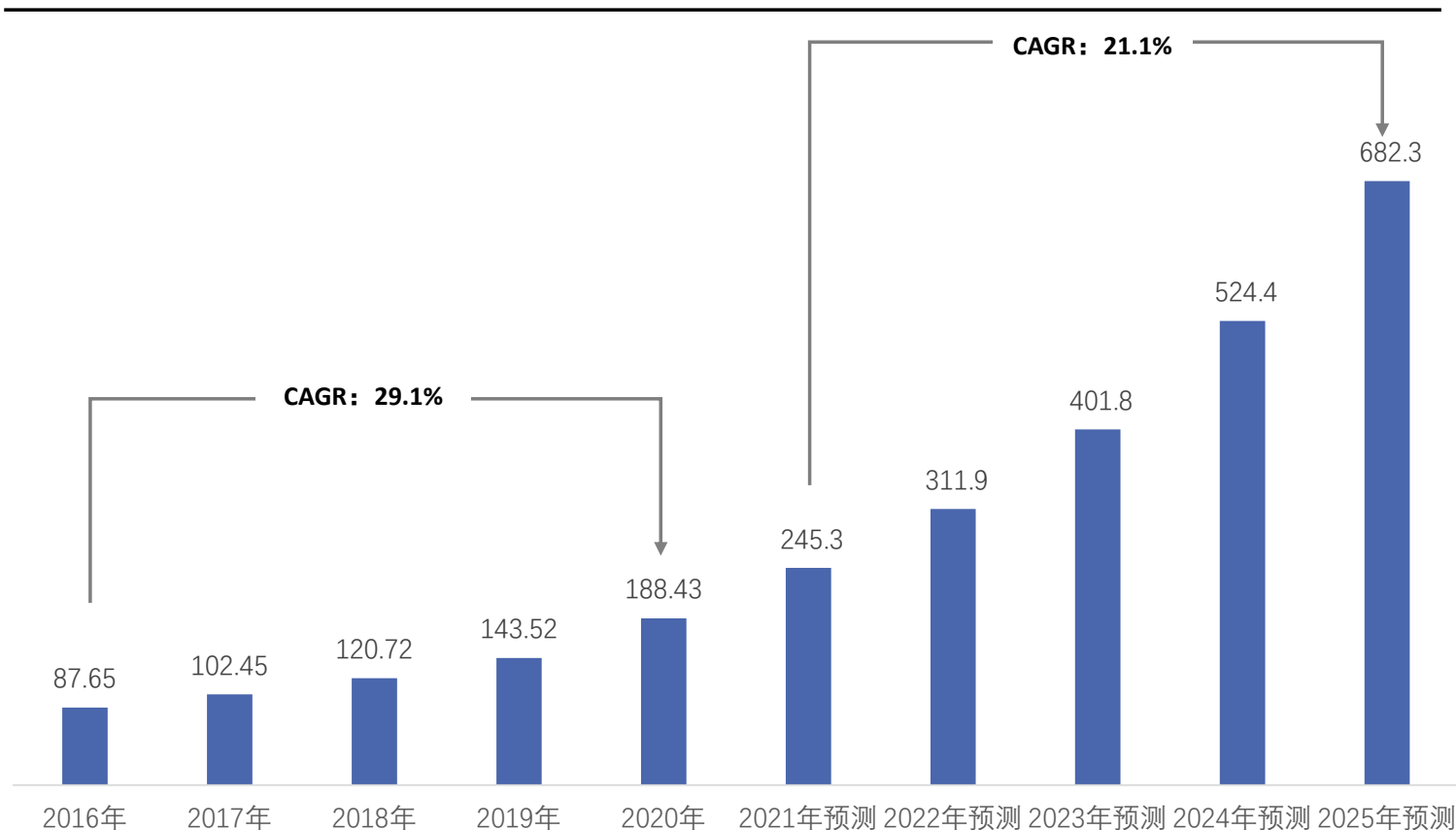
不同数据库类型的开源数据库流行度占比



来源：DB-Engines, 头豹研究院

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

2016-2025年中国数据库市场规模及预测 【亿元】



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。

来源：中国信通院、国家统计局、Scale Grid，头豹研究院

中国数据库行业整体风险适中，其中技术风险适中，政策风险较小，市场风险适中

数据库行业风险

- **政策风险较小：**出于国家发展战略考虑，中国政府将计算机软件行业作为重点发展领域，其中就包括数据库行业

政策风险



宏观调控的不确定性：

- 不排除政府由于经济因素、政治因素、宏观调控等因素，出台针对行业的限制政策，若数据库企业不能及时应对，将造成政策风险

技术风险



技术更迭：

- 由于技术更迭速度提高，数据库企业在短时期内不能满足客户和市场的更高要求

研发周期长：

- 云数据库对传统部署数据库的完善及替换成本较高，数据的迭代升级的完善仍需要时间沉淀

- **技术风险较大：**短期来看5G、大数据和云计算的进一步发展和普及将加速数据库产品的升级和迭代，这将刺激用户对数据资源和数据管理更高的需求，若不能及时满足，市场将面临价格战的恶性竞争

- **市场风险较适中：**数据库产品对于技术和资金双密集型产业，因此数据库厂商会通过规模经济、技术壁垒以及稳定的质量来维持市场地位，格局变动周期较长

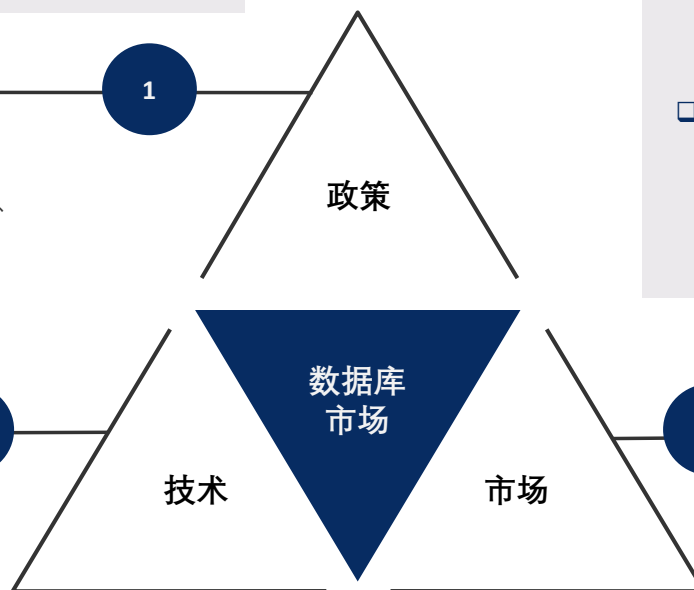
市场风险

垄断风险：

- 数据库行业技术壁垒明显又是资金密集型产业，市场参与者易通过规模经济形成低成本壁垒，进而垄断市场

上游数据库基础设施供应商前向一体化风险：

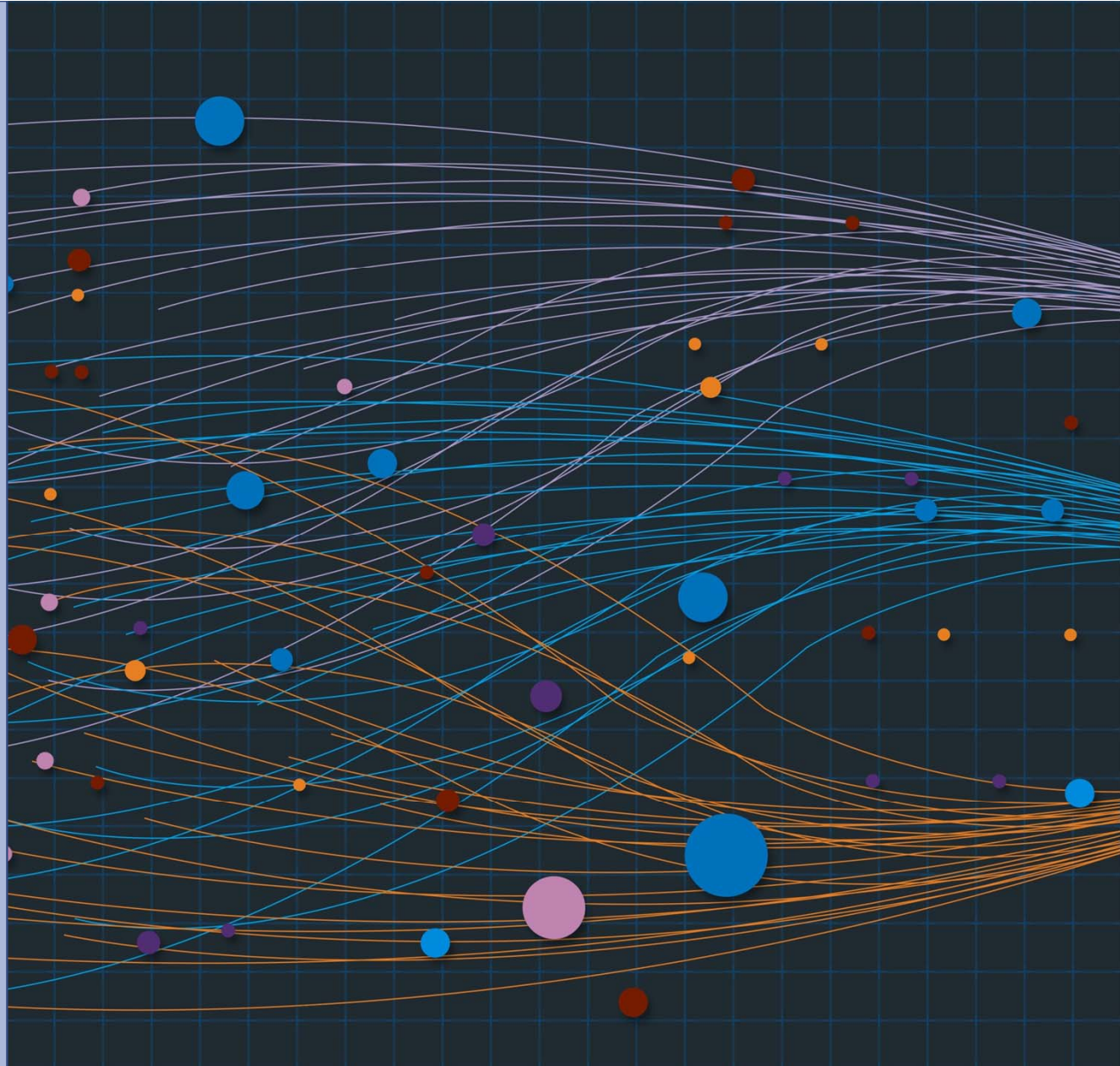
- 数据计算与存储解除耦合，为保住数据存储物理设备及服务器利润，上游厂商或将加深布局数据库软件业务



Chapter 4

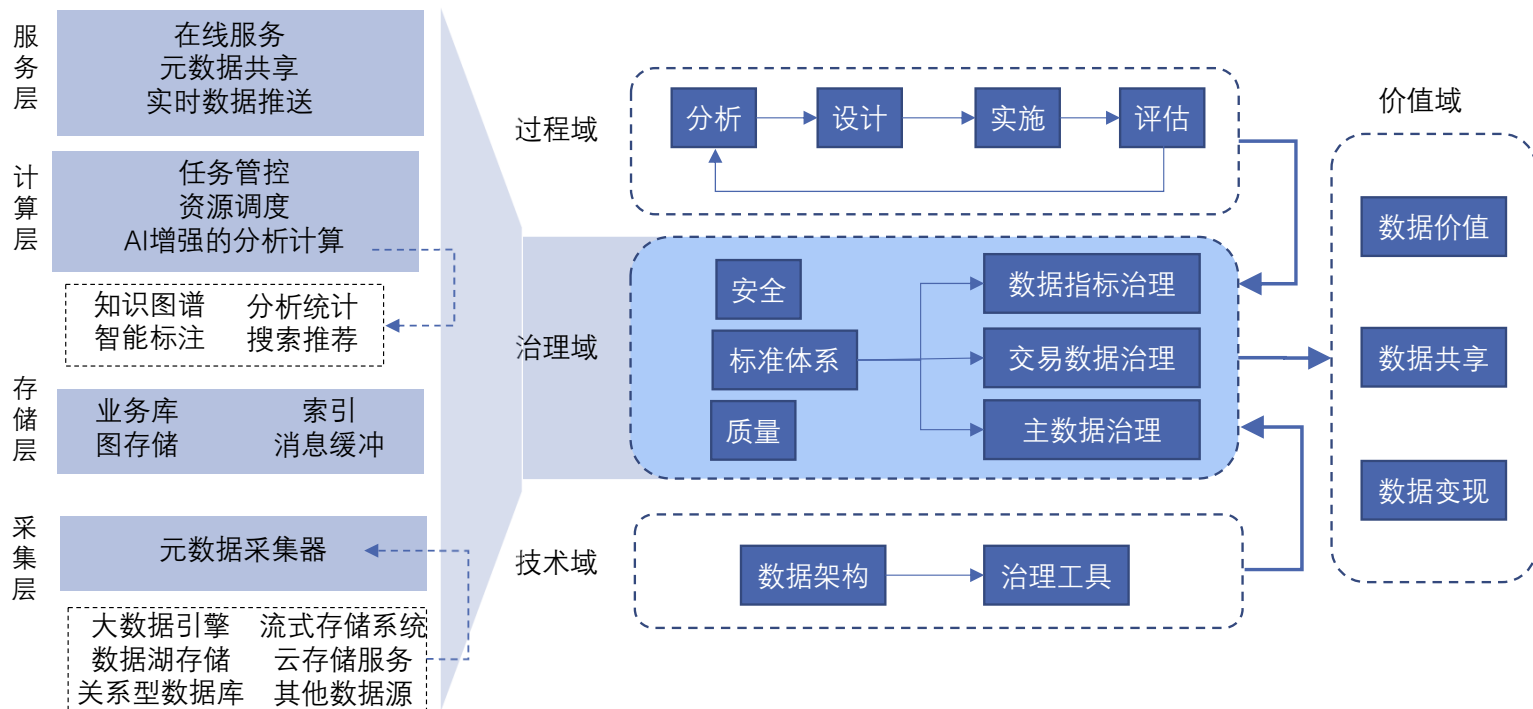
数据治理

- 定义
- 发展路径
- 数据资产管理措施
- 应用现况
- 市场规模
- 企业对比分析



为了适应新时代基于数据的生产、分析、决策的需求，包括数据管理和数据价值变现的数据治理成为了关键

数据治理宏观架构



■ 数据治理 Data Governance

随着企业信息系统的不断发展，数据的管理和管治变得更为复杂，不仅包含了企业内部网，还包含了企业外部环境数据的外部网；数据不仅要支撑企业的运行管理，更要开始为企业对环境的认识和决策提供帮助。数据治理的概念应运而生，广义的数据治理包括数据管理和数据价值“变现”，具体包含数据架构、主数据、数据指标、时序数据、数据质量、数据安全等一系列数据管理活动的集合。

头豹研究院将数据治理定义为：在数据驱动的业务与生产环境中，围绕具备更高可用性的数据标准与数据模型，规划构建“采集、存储、管理、服务”一整套制度体系的系统工程。

来源：DAMA, 阿里云, 航天云网, 中国石油, 头豹研究院

数据治理是数据管理的延伸和进阶，集成不同数据治理工具，构建数据中台，建设全域数据治理是趋势中的发展路径

数据治理演变历程

编码时代

主要采用发布主数据代码为主，采用接口开发形式分发数据。

1994年-2004年

数据处理 Data Processing

在早期为了更高效的检索和管理数据，通过计算机开始对数据进行采集、加工、转换和传输，将原始的、无序的和非结构化的数据转换为结构化的数据。

主数据时代

主要采用国外主流产品，采用SOA，B/S架构，为ERP等经营管理系统服务。

2005年-2011年

数据管理 Data Management

随着数据量的不断增加，开始对计算机数据进行更专业化的存储、检索、控制，包括文件和数据库的介入、数据处理系统的管理等。

信息标准化时代

主要采用自主ESB数据交换平台，采用微服务架构，实现动态模拟、移动应用等。

2011年-2017年

数据管治 Data Administration

随着数据之于效率提升的重要性不断凸显，开始关注一个组织结构的计算机系统中所存储的数据、信息作为组织机构的资源的管理，包括数据的分析、分类、维护、流动、应用等。

数据治理时代

主要采用数据中台和微服务架构初步替代传统架构，对内和对外系统提供云服务。

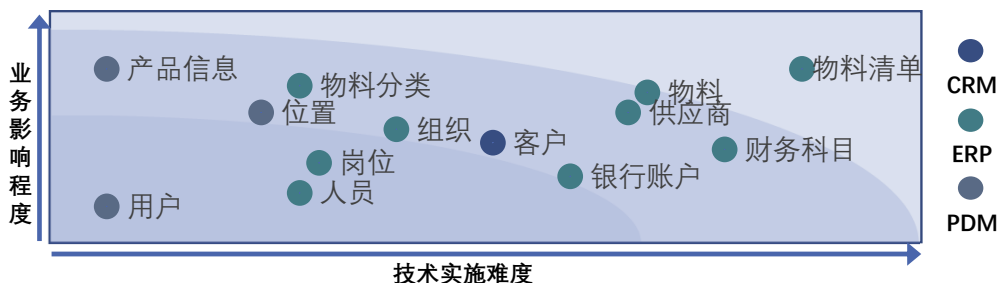
2018年-

数据治理 Data Governance

在数据驱动的业务与生产环境中，围绕具备更高可用性的数据标准与数据模型，规划构建“采集、存储、管理、服务”一整套制度体系的系统工程。

来源：工业大数据应用技术国家工程实验室、腾讯分析，头豹研究院

数据管理节点的业务影响与实施难度



来源：中国移动，头豹研究院

- 21世纪以来，基于互联网的企业信息系统（企业内部网和外部网）的发展，企业数据的管理和管治更为复杂，不仅包含了企业内部的各种产品设计、生产、管理数据，还包含了与企业外部运行环境和竞争环境相关的一切数据；数据不仅要支撑企业的运行和管理，更要为企业对环境的把握和决策服务。
- 在不同环节单独实施不同格式的主数据管理系统给企业埋下了从研发到生产再到销售的整个流程信息脱节的隐患，限制了生产效率和发展边界。CRM、ERP、PDM等企业管理系统亟需贯通信息流，打破数据孤岛，数据治理厂商要培养企业用户数据管治的思维，兼容服务企业信息流上下游。
- 相似的，单一提供比如报表、BI的数据管理厂商也需要兼容从下而上的数据结构才能发挥作用，提前将统一且规范的数据作业布局在数据过程的边界，保证数据传递的一致性与有效性，为创造数据价值打造稳定可靠的数据平台，是数据治理厂商趋势中的发展路径。

中国数据资产管理和应用尚处于摸索阶段，企业及政府部门尚未建立及完善有效的管理和应用数据模式，影响大数据在各领域的共享和应用

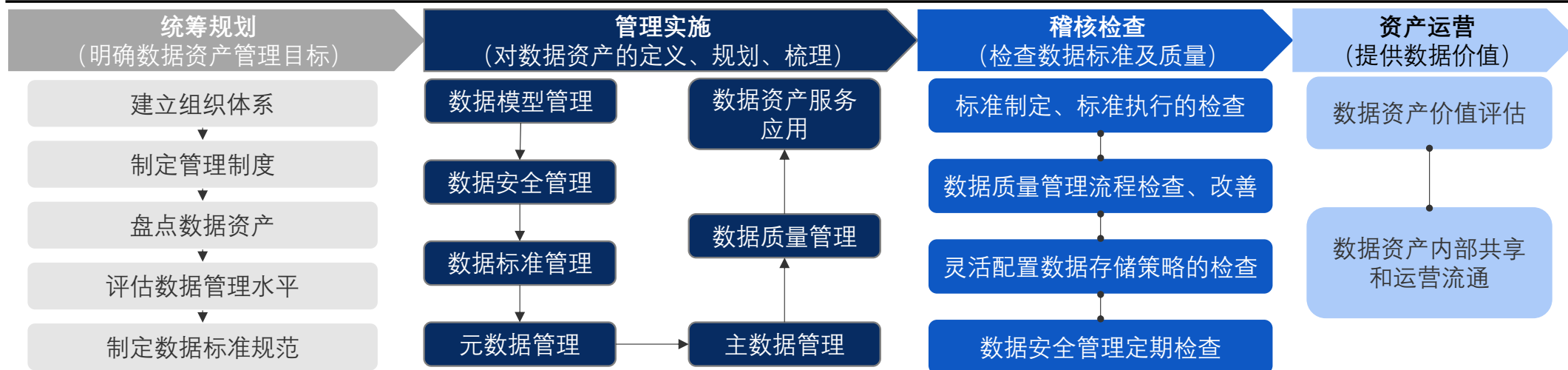
➤ 数据资产管理发展滞后，影响大数据共享及应用

中国大数据资产管理和应用尚处于摸索阶段，一方面，大部分企业和政府部门尚未建立有效的管理和应用数据模式，如数据价值评估、数据成本管理等，大部分企业及政府部门数据基础较薄弱，数据标注尚未统一，数据质量参差不齐，数据孤岛化亟待解决，影响大数据在各领域的共享和应用。

➤ 利好政策助力大数据管理能力评估体系逐步建立

大数据是保障生产实施及组织运营的重要生产资料，伴随管理对象逐渐复杂，数据处理技术日渐成熟，数据应用愈发广泛，政府在数据资产管理方面的力度持续加大。2019年9月工信部公开征求《工业大数据发展指导意见》，意见指出政府将结合工业领域大数据管理的特点和需求，提升工业大数据资源管理能力，系统构建工业大数据管理能力评估体系，同时鼓励并引导企业全面且系统的提升工业大数据管理能力。

数据资产管理实施步骤

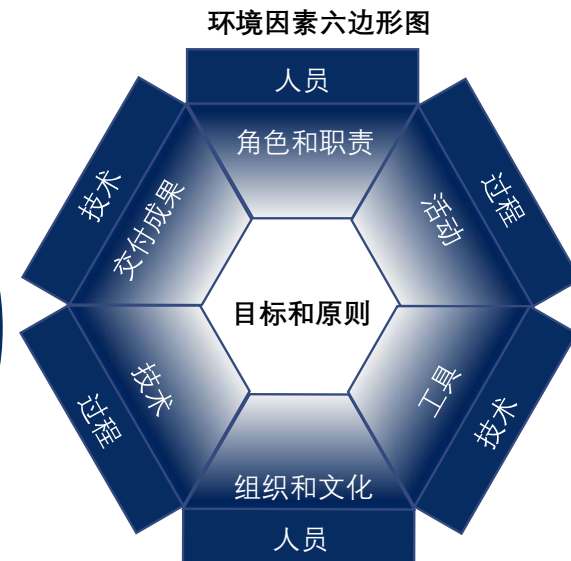
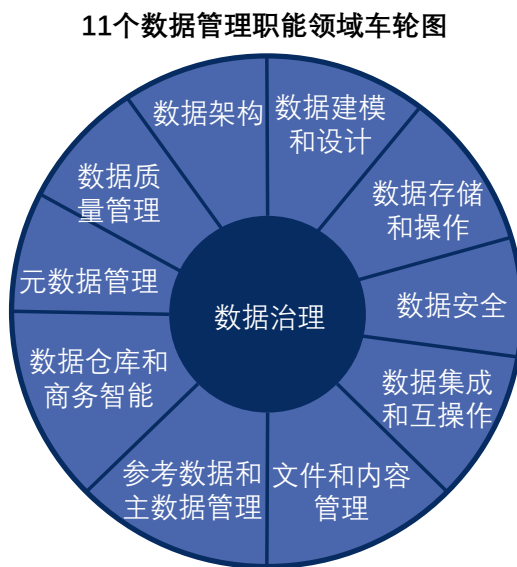
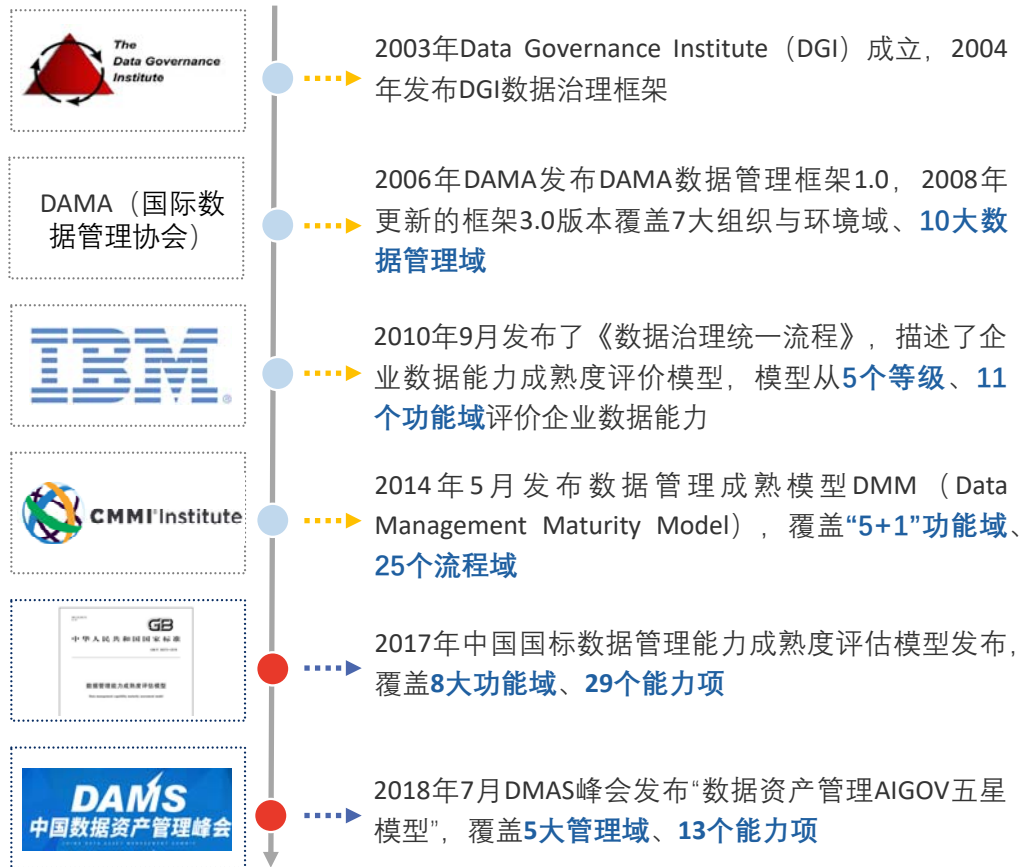


来源：工信部、信通院云计算与大数据研究所、大数据技术标准推进委员会、头豹研究院

虽然在数据资产管理领域起步较晚，但相较海外数据资产管理模型，数据资产管理“AIGOV五星模型”在关注数据资产管理的全局性和应用性等方面具有较大优势

中国及海外国内外数据资产管理标准/模型及框架

国际数据管理协会DAMA-DMBOK2理论框架

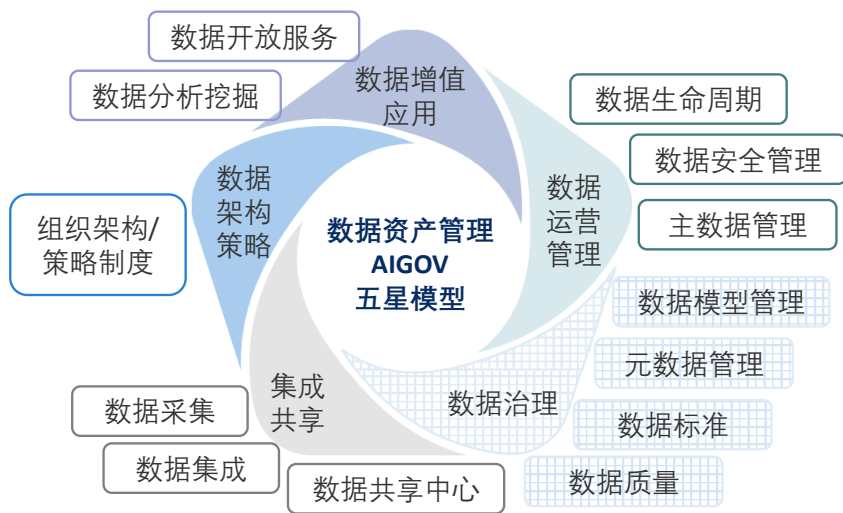


来源：国际数据管理协会DAMA，头豹研究院

来源：数据观，中国电子技术标准化研究院，信通院，工业互联网产业联盟，头豹研究院

虽然在数据资产管理领域起步较晚，但相较海外数据资产管理模型，数据资产管理“AIGOV五星模型”在关注数据资产管理的全局性和应用性等方面具有较大优势

数据资产管理AIGOV五星模型内容及价值



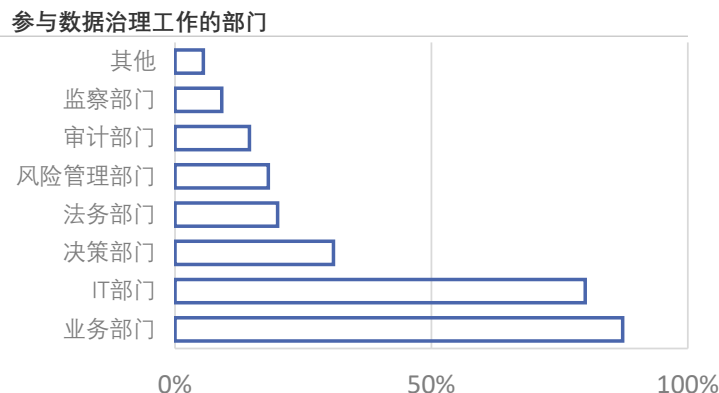
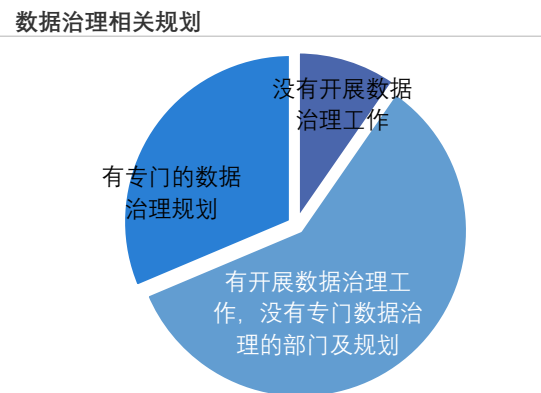
数据资产管理AIGOV五星模型价值：实现数据资产保值与增值

- 助力企业科学制定数据资产管理规划
- 规避常见实施过程中的数据资产管理风险
- 沉淀企业数据资产知识图谱
- 推动企业内外数据共享与应用

来源：工信部，信通院云计算与大数据研究所，大数据技术标准推进委员会，头豹研究院

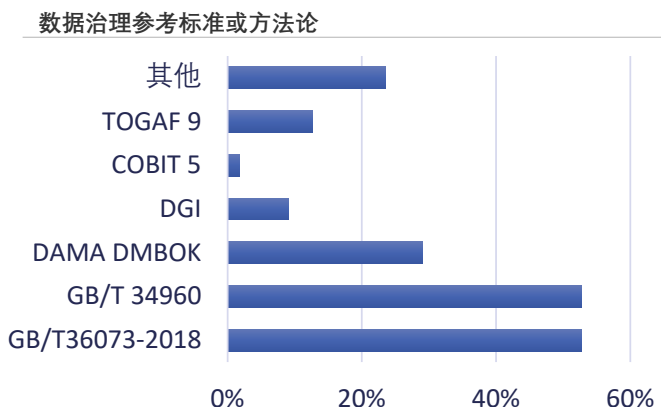
在利好政策的推动下，部分企业自创数据资产管理模型，针对海外数据资产管理模型的弊端及弱势，自主研发适合中国数据资产管理现状的模型。例如新炬网络自主设计的“AIGOV五星模型”现已覆盖五大管理域、十三个能力项，相较于海外数据资产管理模型，新炬网络研发的数据资产管理五星模型在全局性和应用性等方面具有较大优势。数据资产“AIGOV五星模型”的五大管理域分别从数据架构、数据集成共享、数据治理、数据运营管理及数据增值应用五方面进行重构和治理，旨在实现数据资产保值和增值。截至2020年3月，新炬网络已与广东、浙江、安徽、湖南等省份的省级运营商建立合作关系，新炬在各省的营运中心派驻专业数据资产管理团队，协助当地运营商建设自身的数据资产管理体系，现已协助多省用户建立数据资产管理制度、流程以及规范。

当前企业在数据治理的应用中，存在数据治理组织机制不健全、数据治理体系不完善、缺乏数据战略规划、缺乏标准指导的问题



企业数据治理存在的困难

缺乏统一数据模型管理规范, 数据模型与实际不符	缺乏统一的数据服务	对企业的整体数据脉络和分布情况无法掌握	源端系统数据质量问题严重
各业务系统的数据质量不可控	缺乏统一数据标准, 无法实现数据融合	缺乏专业的数据管理机制	缺乏保障安全的相关策略



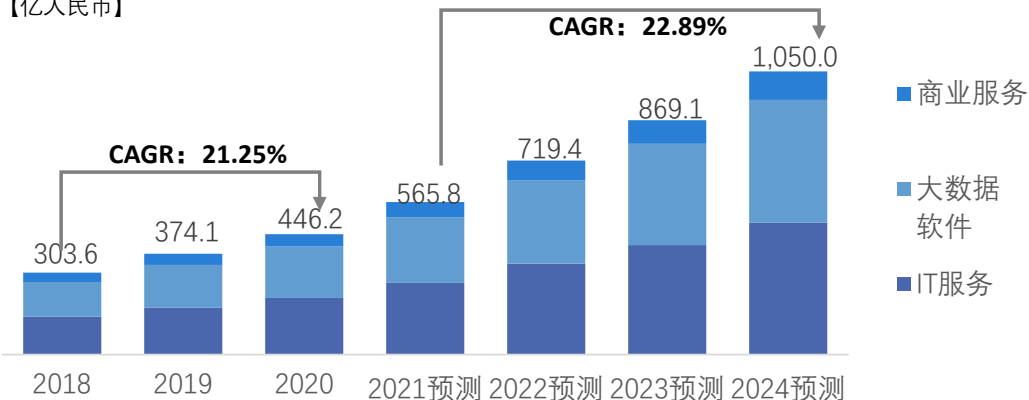
- ❑ 数据治理企业应用现存问题
1. 大部分企业缺乏专门的数据治理规划来知道数据治理的具体实施。企业内部设置专职数据治理部门和数据治理岗位仍在少数，大多数据治理规划都还停留在兼职部门和岗位层面。
 2. 数据治理组织是数据治理工作的主体。目前大多数企业的数据治理工作主要由业务部门和IT部门配合完成。而包括总裁办、董事长秘书处等决策部门对数据治理的参与度不够，使得数据治理缺乏自上而下的组织。
 3. “缺乏统一”是企业数据治理工作遇到最多的问题，制定统一标准是企业数据治理工作的重中之重。而还没有一项标准或理论方法能够被绝大多数企业参考。
- ❖ 随着数字化转型进程中，企业对数据治理的重视度不断进阶，数据治理技术和工具的不断智能化、高效化，标准体系的研制、宣贯不断深化，数据治理的市场将持续扩张，数据治理在企业数字化建设的投入占比将增加。

来源：信标委大数据标准工作组、御数坑、头豹研究院

随着市场需求扩大，数据支持方案趋于成熟，国家数据安全政策的立法完善，数据治理市场规模在预测期间以33.83%的年复合增长率于2024年达到325.5亿人民币的价值

大数据服务市场规模，2018-2024年预测

【亿人民币】



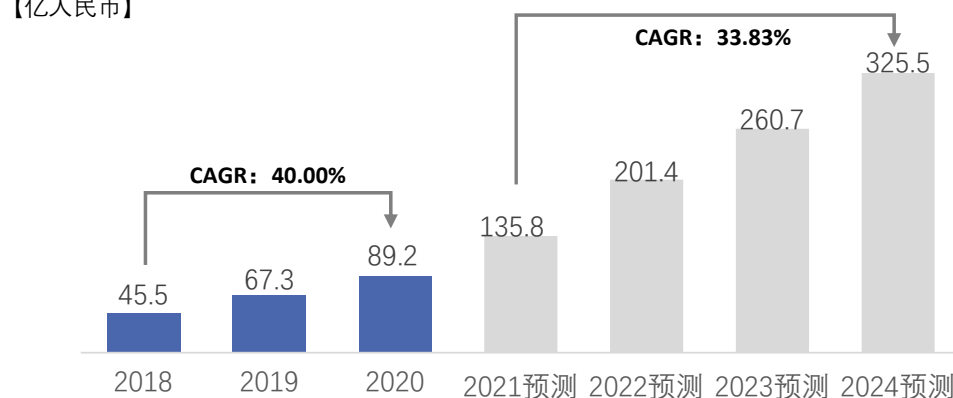
来源：沙利文

2021年9月将施行《中华人民共和国数据安全法》，不断增长的监管和合规要求，对数据治理市场需求的刺激最为突出。更多的企业将不得不重视并建立数据治理组织机制，加强标准规范体系的应用，着重数据治理专业人才的培养。同时，数据治理在标准化体系化的进程中，保证了数据质量，更有效地释放数据价值，支撑企业的管理与决策，促使企业加大对数据治理的投入。



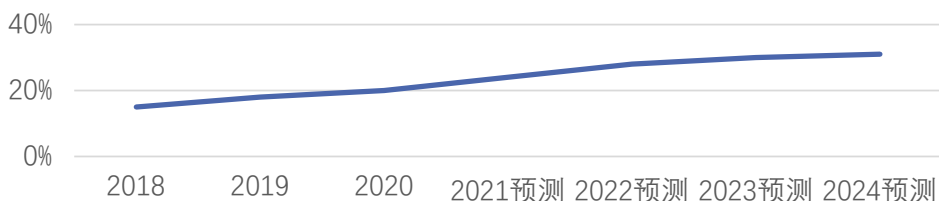
数据治理市场规模，2018-2024年预测

【亿人民币】



来源：头豹研究院

数据治理在大数据平台建设中的平均预算占比



来源：爱分析，头豹研究院

美林数据、神策数据、亿信华辰是三家有着截然不同的发展历程的数据公司，在数据治理领域上将是殊归同途



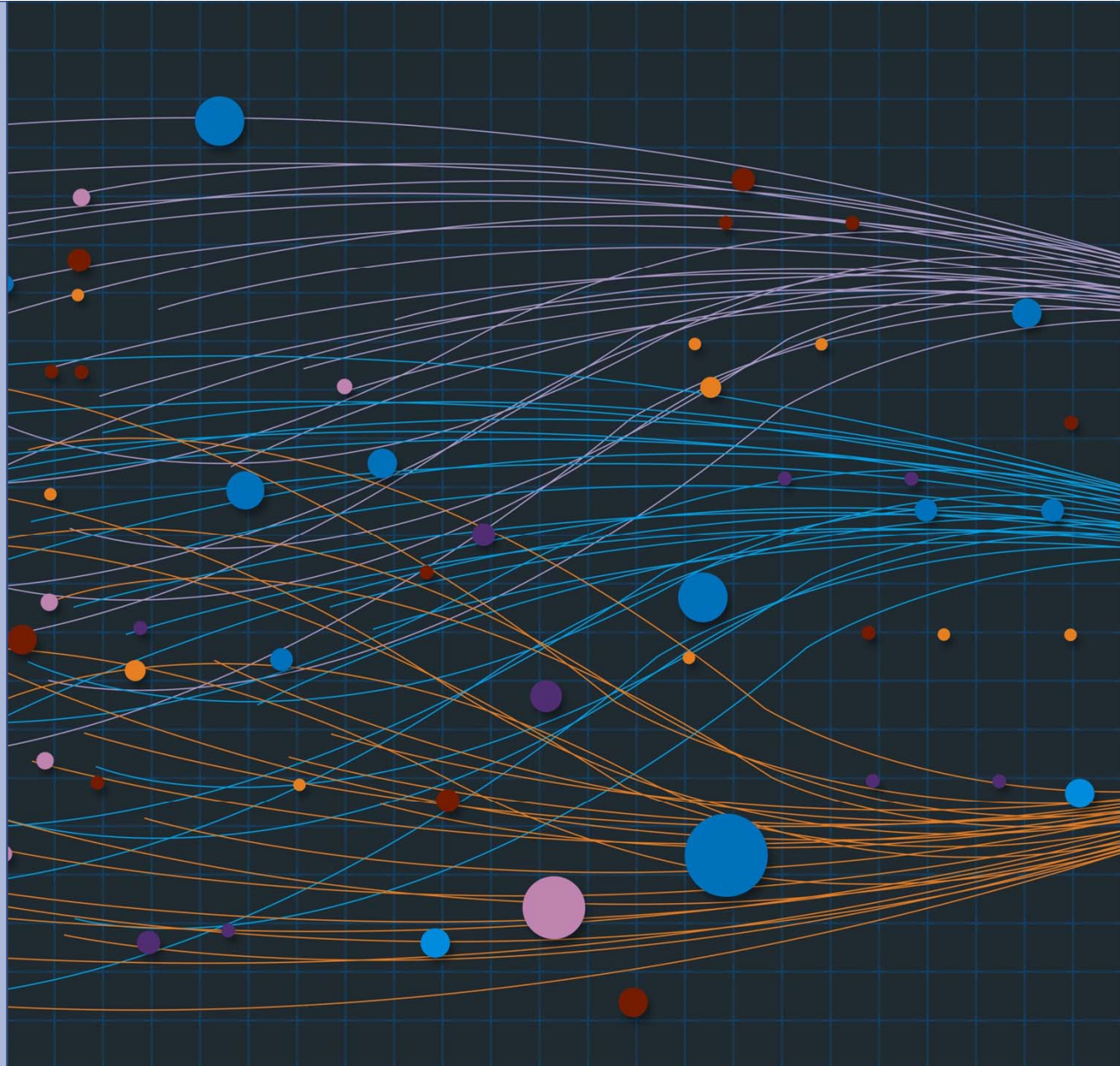
数据治理领域企业对比

	发展逻辑	业务布局	业务布局	发展逻辑
美林数据	<p>美林在国内工业大数据、工业互联网领域有着长期的数据分析与数据治理技术积累，尤其是制造、能源领域。未来，新基建加速落地，美林能够在智能制造、智慧能源领域进一步提供下沉专业化服务，将抓住数据化、网络化、智能化融合发展的契机，助推传统工业企业数字化转型。</p>	<p>以数据分析和数据治理为核心： 美林数据主攻数据分析和数据治理，向企业客户提供数据资产管理、数据分析与挖掘、数据开发应用。</p> <ul style="list-style-type: none"> □ 数据分析 □ 数据治理 □ 运维服务 □ 业务管理 □ 培训服务 		<p>美林是典型的技术驱动型企业，目前新三板上有望达到创业板和科创板的条件。但由其工业领域的定位所限，美林的数据治理和数据分析产品都以定制化开发为主，导致成本较高，从2020年公司年报可见，营业成本高于营收，首度亏损。同时，电力及军工制造工业级项目的回款周期较长，应收账款占总资产比重的45.41%，现金流量流动性较差，虽美林已在年报中强调国企客户的偿付违约风险较低，但仍要关注坏账损失的可能性。</p>
神策数据	<p>神策围绕用户级大数据分析和需求，从互联网运营切入，以私有化部署和全量采集的定位，瞄准了对数据安全和全生命周期数据分析有更高需求的市场。最终从数据分析定义了公认的行业最高标准：30分钟完成私有化部署、单日入库千亿条数据、亿级日活实时在线分析等。夯实了神策在市场的立足点，逐步形成了咨询层、服务层、产品层的未来业务发展架构。</p>	<p>以数据治理为补充： 神策数据主攻数据分析和营销科技服务，为了构建统一的数据标准体系，发展数据治理业务。</p> <ul style="list-style-type: none"> □ 数据分析 □ 数据治理 □ 业务管理 		<p>神策数据未公开过财务数据，但其独到的市场定位与高速技术迭代使得其短短五年就获得了飞速成长的市场表现，是当之无愧的数据领域独角兽公司。A轮400万美元，B轮1,100万美元，C轮4,400万美元，C+轮3,000万美元的融资路径也在证明资本市场看好神策的估值与未来。</p>
亿信华辰	<p>亿信华辰从商业智能BI软件出身，15年的时间，以高比例的研发投入，涉及数据存储、数据处理、数据治理、数据分析，提供数据产业全链服务，覆盖100+细分领域。亿信华辰通过OEM合作为客户提供技术方案的支持；产品/服务优势互补降低项目开发成本；提供全方位服务支持，与客户共同进行项目交付缩短项目期这三种方式提供服务。使得其在维持高额研发投入的同时，快速在各行业铺开，以技术研发满足服务需求，以服务经验反哺技术能力的提升。</p>	<p>数据产业链业务布局： 亿信华辰的业务板块涉及数据存储、数据处理、数据治理、数据分析，提供全链服务。</p> <ul style="list-style-type: none"> □ 数据分析 □ 数据治理 □ 数据处理 □ 数据存储 		<p>亿信华辰由华宇软件控股80%，其余20%留作员工激励。2020年华宇软件的年报中显示，亿信华辰总资产2.55亿元，净资产1.10亿元，报告期内净利润206.37万元，应收款项存在9,800万元。与美林数据类似，亿信华辰也有大量客户来自电力、制造、通信，回款周期长的通病亿信华辰没能幸免。从上述发展逻辑中展示，亿信华辰已然找到方法改善结算模式，创新推动项目进度，创造回款条件。</p>

Chapter 5

工业领域的管理

- 数据资源
- 产业链
- 中游业务模式
- 需求分析
- 应用价值



数据是工业数据管理行业的核心竞争资源，伴随工业大数据与新技术领域联系更加紧密，工业大数据的商业价值将得以进一步释放

工业大数据的来源渠道共分为三类：（1）企业运营管理相关的业务数据（**企业运营管理数据**）；（2）工业制造过程中产生的数据（**生产数据**）；（3）企业外部数据。伴随工业大数据与云计算、物联网、边缘计算、人工智能等技术融合与发展，工业大数据的数据获取量将逐渐提高，工业大数据将实现商业价值最大化。

企业运营管理数据：该类数据来自企业信息化范畴，包括企业资源计划(ERP)、产品生命周期管理(PLM)、供应链管理(SCM)、客户关系管理(CRM)和能耗管理系统（EMS)等，此类数据是工业企业传统意义上的数据资产。

生产数据：该类数据是在工业生产过程中，装备、物料及产品加工过程的工况状态参数、环境参数等生产情况数据，通过MES系统实时传递，目前在智能装备大量应用的情况下，此类数据量增长最快。

企业外部数据：该类数据包括工业企业产品售出之后的使用、运营情况的数据，及大量客户名单、供应商名单、外部的互联网数据等。

中国工业大数据源特征

数据类别	收集渠道	采集难度	数据利用率	数据服务合作方	数据应用场景
企业运营管理数据	企业内部各运营管理部门	★★★☆☆	★★★★★	<ul style="list-style-type: none"> 工业软件服务商 大数据技术服务方 数据挖掘算法服务提供商 其他 	<ul style="list-style-type: none"> 智能化设计 智能化生产
生产数据	企业内部生产线	★★★★☆	★★★★☆☆		<ul style="list-style-type: none"> 协同制造 智能化服务
企业外部数据	企业内部及企业外部	★★★★☆	★★★☆☆		<ul style="list-style-type: none"> 智能化服务 其他

来源：头豹研究院

中国工业数据管理行业上游资源提供商为中游参与者提供多样化软硬件资源，中游工业大数据服务商为下游工业领域各细分行业用户提供工业大数据产品及服务

中国工业数据行业产业链

上游：软硬件服务商

- ✓ 工业大数据软硬件服务行业利润率近**10%**
- ✓ 美国、日本、德国合计占据全球传感器市场份额的**60%**
- ✓ 中国中高端传感器进口率由2007年的**99%**降至2018年的**85%**

传感器提供商



云服务提供商



中游：工业大数据服务商

- ✓ 中国工业大数据行业中游行业利润率为**12.7%**
- ✓ 大数据服务商综合实力较强，工业大数据服务商专业性较强

专业工业大数据服务商



大数据服务商



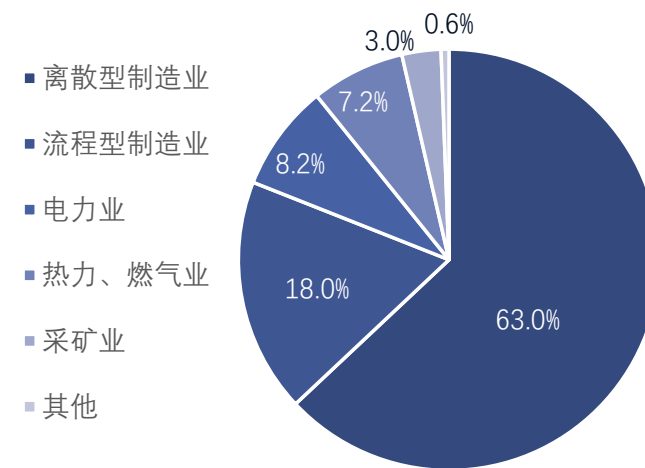
工业互联网企业



下游：行业用户

- ✓ 离散型制造企业用户收益 (**71亿元**) 居行业首位，其次为流程型 (**20.5亿元**) 等行业

中国工业大数据用户行业结构占比统计

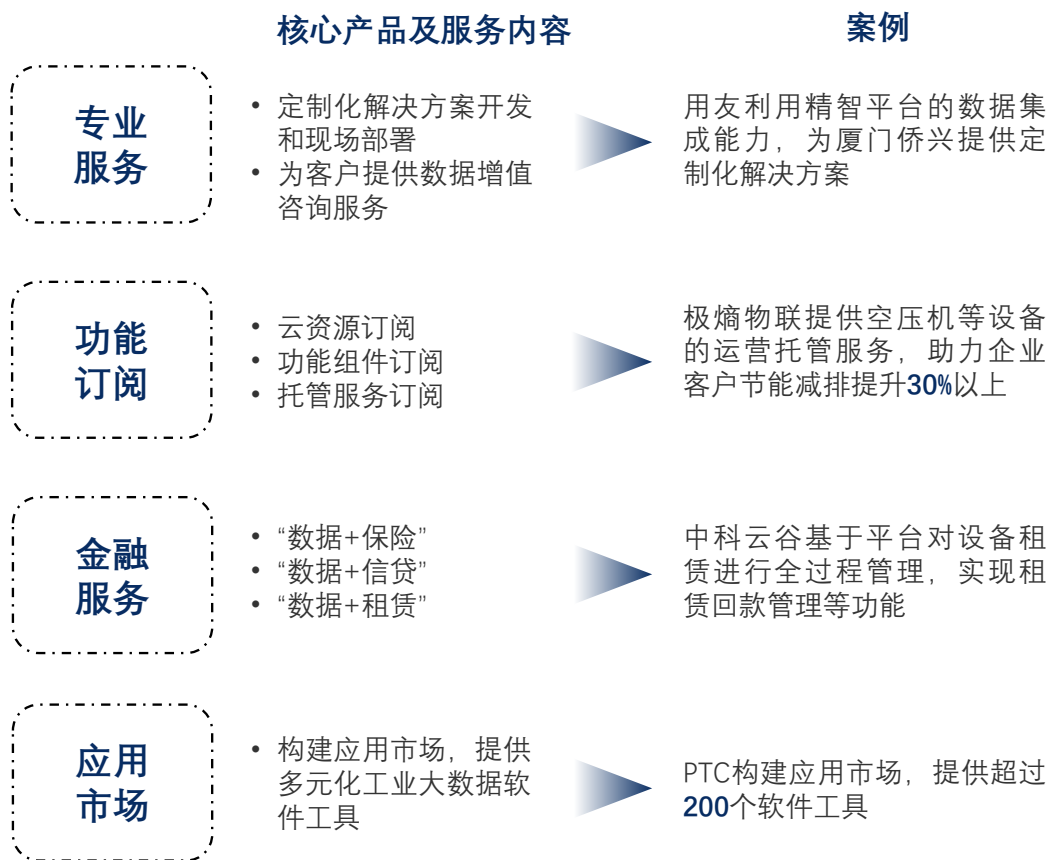


来源：头豹研究院

中国工业大数据行业上游资源供应商为工业大数据软硬件服务商，为中游服务商提供保障工业大数据产品和服务正常运转所需的多样化软硬件资源，以及大数据存储、网络 and 计算相关的软硬件产品和服务，行业利润率近**10%**。中国工业大数据行业中游参与者按照企业类型可分为专业工业大数据服务商、大数据服务商等，行业利润率为**12.7%**，中游工业大数据服务商为下游各领域企业用户及政府用户提供工业大数据产品及服务。

专业服务为工业大数据服务商的核心业务模式；部分工业大数据服务商通过将核心产品向云端迁移，以低成本、灵活交付的优势吸引企业客户采购

工业大数据服务商的业务模式



➤ 专业服务为工业大数据服务商的核心业务模式

工业大数据服务商的业务模式包括专业服务、功能订阅、金融服务、应用市场四种。其中**专业服务**是现阶段工业大数据服务商的核心业务模式，原因为工业体系具有专业性与复杂性，通用化产品及解决方案难以满足工业企业客户对数据增值的需求。

➤ 部分工业大数据服务商升级商业模式，提高服务水平

传统工业软件企业将软件能力转化为**平台PaaS及SaaS服务**，一方面以**低成本、灵活交付**优势吸引更多用户采购，另一方面借助PaaS、SaaS核心技术提升数据采集及分析能力，更深层次挖掘工业数据价值。

用友、金蝶、Ansys、Infor等工业软件企业已将核心软件产品向云端迁移；华龙讯达与腾讯合作，借助腾讯云计算、大数据技术实现业务数据的云端存储管理和快速处理。

部分工业大数据服务商构建平台**简化**数据连接，借助边缘计算、AI分析等技术深化数据分析能力、提升工业大数据服务解决方案技术水平和能力。例如日立推出Lumada平台，利用大数据处理和AI分析等技术助力客户形成资产运维、能耗优化、产线监测、工艺优化。

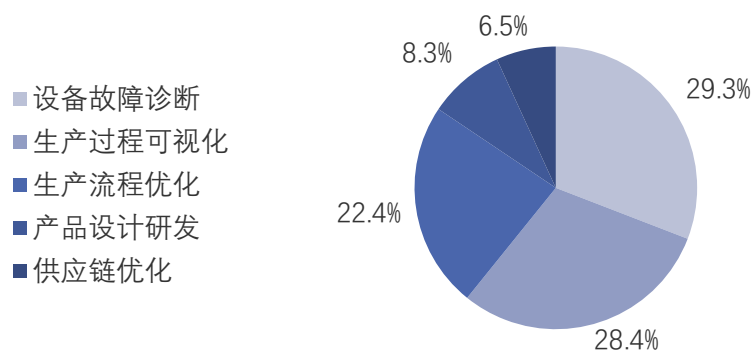
来源：工业互联网产业联盟、头豹研究院

工业大数据技术及产品在工业生产各环节可实现多方要素覆盖，加速优化工业生产品命周期各环节，提升工业企业智能化程度

工业智能化转型需求



中国工业大数据市场产品结构（按销售额计），2020年



来源：信通院云计算与大数据研究所，大数据技术标准推进委员会，头豹研究院

中国工业企业的智能化、数字化转型需求持续上涨

工业作为国民经济的重要物质生产部门，是推动国民经济发展的基础动力，伴随新一代信息技术与工业的深度融合，中国工业迎来产业变革，新的生产方式、产业形态、商业模式及经济增长点推动中国工业企业向智能化、数字化转型。

同时工业4.0概念的出现和普及，以及《中国制造2025》等利好政策的颁布，对工业发展提出新的要求，例如

- （1）保障工业制造本身价值化的同时减少生产过程的浪费；
- （2）在原有自动化的基础上实现“自省”；
- （3）在工业制造过程中达到零故障、零忧虑、零意外、零污染等。

工业4.0的实现以数字化整合为主要驱动力，工业智能化转型需求的增长加速工业大数据商业化产品的落地。

工业大数据商业化产品契合工业智能化转型需求

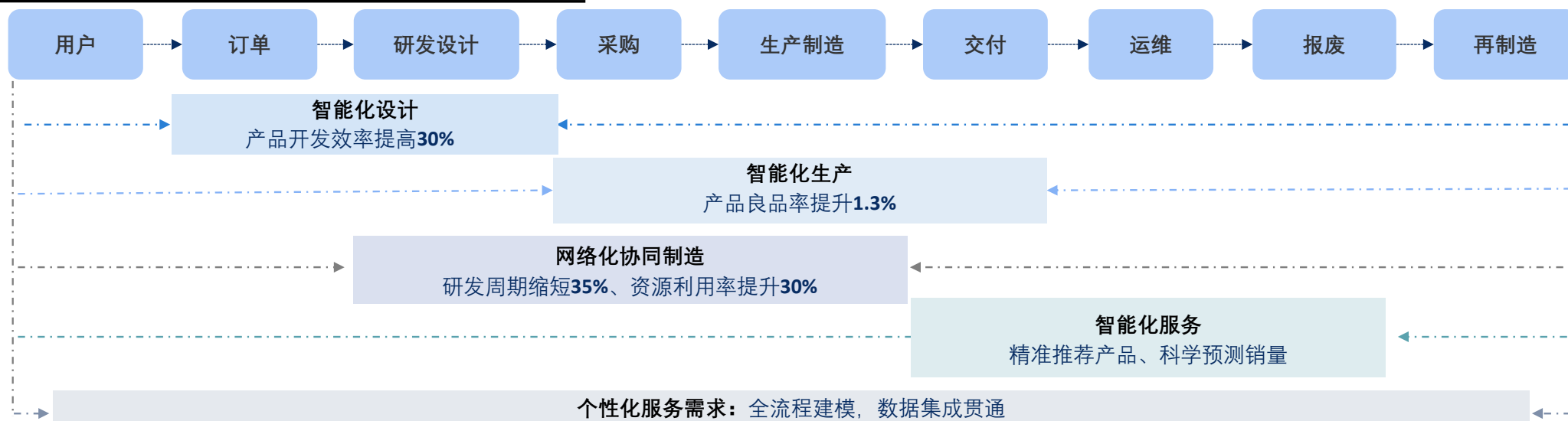
伴随信息化与工业化的深度融合，中国工业企业在生产各环节产生的数据量日益丰富，数据对工业生产的驱动作用日渐突出。

工业大数据技术及产品覆盖从订单到研发设计、采购、生产、交付、运维、报废及再制造的生命周期的全过程，现已落地的工业大数据产品类型包括设备故障诊断、生产过程可视化、生产流程优化、供应链优化等。

2020年销售额占比居首位的工业大数据产品为设备故障诊断产品，占据工业大数据产品销售总额的29.3%，其次为生产可视化产品，占据工业大数据产品销售总额的28.4%。

工业大数据技术及产品在工业生产各环节可实现多方要素覆盖，加速优化工业生产品生命周期各环节，提升工业企业智能化程度

工业数据管理应用场景及应用价值



来源：信通院云计算与大数据研究所，大数据技术标准推进委员会，头豹研究院

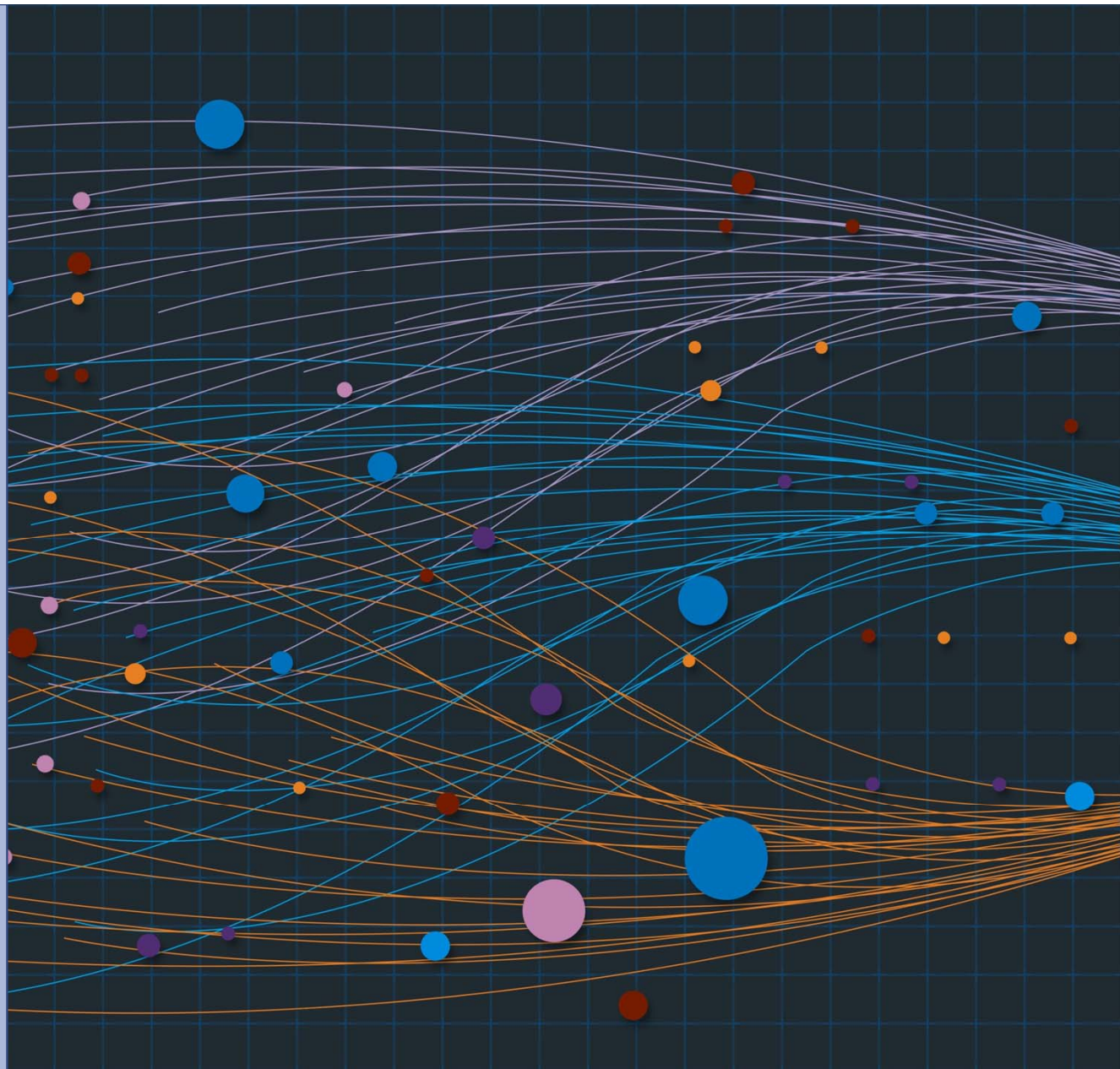
工业大数据产品的商业化落地及广泛应用助力企业实现产品的智能化设计、智能化生产以及智能化服务等目标，加速优化产业链各环节，例如

- (1) **在设计环节**，工业大数据的应用可有效提高研发人员创新能力、研发效率可提升**30.0%**；
- (2) **在生产环节**，企业可通过采集和汇聚设备运行工业数据、工艺参数、质量检测数据、物料配送数据和进度管理数据等生产现场数据，利用大数据技术分析和反馈并在制造工艺、生产流程、质量管理、设备维护、能耗管理等具体场景应用，实现**生产过程的优化、产品质量的监测和异常追溯**，产品良品率可提升**1.3%**，研发周期缩短**35.0%**、资源利用率提升**30.0%**；
- (3) **在市场营销环节**，企业可利用大数据挖掘用户需求和市场趋势，建立用户对商品需求的分析体系，挖掘用户深层次的需求，寻找机会产品，进行生产指导和后期市场营销分析；
- (4) **在售后服务环节**，数据驱动企业创新服务模式，从被动服务、定期服务发展成为主动服务、实时服务。中国工业大数据产品的优势加速其在工业领域不同场景的应用和普及，推动工业大数据行业发展。

Chapter 6

附录

- 数据管理厂商图谱



数据管理厂商图谱



图谱企业位置不涉及排名与先后顺序
来源：数据猿、各公司官网，头豹研究院

名词解释

- ◆ 大数据：Big Data，无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。
- ◆ 元数据：Metadata，又称中介数据、中继数据，为描述数据的数据（data about data），主要是描述数据属性（property）的信息，用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。
- ◆ 主数据：MD Master Data，指系统间共享数据（例如，客户、供应商、账户和组织部门相关数据）。
- ◆ 结构化数据：又称行数据，结构化数据由明确定义的数据类型组成，由二维表结构实现逻辑表达。结构化数据严格遵循数据格式与长度规范，主要通过关系型数据库进行存储和管理。
- ◆ 非结构化数据：数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据，非结构化数据包括所有格式的办公文档、文本、图片、图像和音频/视频信息等等。
- ◆ 数据仓库：Data Warehouse，是企业所有级别的决策制定过程，提供所有类型数据支持的战略集合。
- ◆ 数据湖：Data Lake，是一个集中式存储库，允许您以任意规模存储所有结构化和非结构化数据。
- ◆ 高级分析：主要通过深度学习、数据准备自动化、自然语言处理技术将数据转化为可实现商业价值。
- ◆ Hadoop：是一种分布式系统架构，充分利用集群的威力进行高速运算和储存。核心设计框架是为海量的数据提供存储的HDFS和为海量的数据提供计算的MapReduce。
- ◆ 联机事务处理：OLTP，是一个在线交易处理系统。OLTP系统的主要是记录事务时的当前更新，插入和删除。
- ◆ 联机分析处理：OLAP是一个在线分析处理系统。存储输入的历史数据。它允许用户查看多维数据的不同摘要。
- ◆ 数据挖掘：Data Mining指从大量的数据中通过算法搜索隐藏于其中信息的过程。
- ◆ 决策支持系统：DSS，是解决非结构化问题，服务于高层决策的管理信息系统。
- ◆ 主管资讯系统：EIS，是服务于组织的高层经理的一类特殊的信息系统。
- ◆ 商业智能：Business Intelligence指用现代数据仓库技术、线上分析处理技术、数据挖掘和数据展现技术进行数据分析以实现商业价值。
- ◆ 数据重力：强调数据在哪里收集就应该在哪里处理，这样操作的效率和成本都是最优的。换句话说，就是不必把数据运送到处理所在的位置，而是把处理推送到数据所在的位置。
- ◆ 数据智能：Data Intelligence，是以数据作为生产资料，通过结合大规模数据处理、数据挖掘、机器学习、人机交互、可视化等多种技术，从大量的数据中提炼、发掘、获取知识，为人们在制定决策时提供有效的数据智能支持，减少或者消除不确定性。
- ◆ 数据资产：Data Asset 指由企业拥有或控制的，能为企业带来未来经济利益的，以物理或电子方式记录的数据资源。
- ◆ 数据治理：Data Governance，作为数据管理的一个核心职能，对数据资产管理行使权力和控制的活动集合。

名词解释

- ◆ 公有云：Public Clouds 指第三方提供商用户能够使用的云，通过互联网使用，可能是免费的或成本低廉的。
- ◆ 私有云：Private Clouds 指为一个客户单独使用而搭建的，公司拥有基础设施，或部署在一个安全的主机托管场所。
- ◆ 混合云：Hybrid Cloud 是公有云和私有云两种服务方式的结合，提供弹性需求，同时保障安全和控制。
- ◆ 数据沙箱：Data Sandbox, 通过实现数据所有权和数据使用权分离，以确保数据流通过程安全可控的产品。
- ◆ 流数据（或数据流）：指在时间分布和数量上趋于无限的一系列动态数据集合体，数据的价值随着时间的流逝而降低，因此必须实时计算给出秒级响应。
- ◆ 流计算：对流数据执行实时加载实时计算，对包含单条记录或几条记录的微批量数据进行相对简单的模型计算。
- ◆ 并行计算：一种一次可执行多个指令的算法，目的是提高计算速度，及通过扩大问题求解规模，解决大型而复杂的计算问题。
- ◆ 分布式计算：Distributed computing，是一种把需要进行大量计算的工程数据分割成小块，由多台计算机分别计算，在上传运算结果后，将结果统一合并得出数据结论的科学。
- ◆ 内存内计算：把计算单元嵌入到内存里面，这样的话内存既是一个存储器，也是一个计算机，它并不需要从内存中读取数据，数据是直接进出CPU的。不但不受内存的性能限制，而且还提高了效能比。
- ◆ 小型机：一种介于PC服务器和大型机之间的高性能计算机，一般认为，传统小型机是指采用RISC、MIPS等专用处理器，主要支持UNIX操作系统的封闭、专用的计算机系统。
- ◆ 大型机：又名大型主机，使用专用的处理器指令集、操作系统和应用软件。故此，大型机不仅仅是一个硬件上的概念，更是一个硬件和专属软件的有机整体。
- ◆ ACID理论：是指数据库管理系统（DBMS）在写入或更新资料的过程中，为保证事务（transaction）是正确可靠的，所必须具备的四个特性：原子性（atomicity，或称不可分割性）、一致性（consistency）、隔离性（isolation，又称独立性）、持久性（durability）
- ◆ BASE理论：通过牺牲强一致性来获得可用性，并允许数据段时间内的一致，但是最终达到一致状态。BASE是三个短语的缩写：基本可用（Basically Available），软状态（Soft State），最终一致性（Eventually Consistent）。
- ◆ CAP理论：指的是在一个分布式系统中，一致性（Consistency）、可用性（Availability）、分区容错性（Partition tolerance）。CAP 原则指的是，这三个要素最多只能同时实现两点，不可能三者兼顾。
- ◆ DBMS：数据库管理系统(Database Management System)是一种操纵和管理数据库的大型软件，用于建立、使用和维护数据库，简称DBMS。
- ◆ 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。

名词解释

- ◆ 大规模多级并行处理能力MPP：massively parallel processing，是采用大量处理单元对问题进行求解的一种并行处理技术。
- ◆ MVC架构：M是指业务模型，V是指用户界面，C则是控制器，使用MVC的目的是将M和V的实现代码分离，从而使同一个程序可以使用不同的表现形式。
- ◆ SOA架构：面向服务的架构（SOA）是一个组件模型，它将应用程序的不同功能单元（称为服务）进行拆分，并通过这些服务之间定义良好的接口和协议联系起来。
- ◆ 微服务架构：微服务架构 = 80% 的 SOA 服务架构思想 + 100% 的组件化架构思想 + 80% 的领域建模思想，系统中的各个微服务可被独立部署，各个微服务之间是松耦合的。每个微服务仅关注于完成一件任务并很好地完成该任务。
- ◆ DevOps：Development和Operations的组合词，是一组过程、方法与系统的统称，用于促进开发（应用程序/软件工程）、技术运营和质量保障（QA）部门之间的沟通、协作与整合。
- ◆ AI Ops：是应用人工智能（AI）包括大数据、分析和机器学习的功能来改进 IT 运营的方法。
- ◆ NoOps：使开发、部署和维护，以便可以将人从循环中移除，一切都得到了处理。
- ◆ LXC轻量级容器：Linux Container容器是一种内核虚拟化技术，可以提供轻量级的虚拟化，以便隔离进程和资源。
- ◆ Zabbix：一个基于WEB界面的提供分布式系统监视以及网络监视功能的企业级的开源解决方案。能监视各种网络参数，保证服务器系统的安全运营；并提供灵活的通知机制以让系统管理员快速定位/解决存在的各种问题。
- ◆ 数字孪生：数字孪生是充分利用物理模型、传感器更新、运行历史等数据，集成多学科、多物理量、多尺度、多概率的仿真过程，在虚拟空间中完成映射，从而反映相对应的实体装备的全生命周期过程。
- ◆ APM应用性能管理：Application Performance Management，是一个比较新的网络管理方向，主要指对企业的业务应用进行监测、优化，提高企业应用的可靠性和质量，保证用户得到良好的服务，降低IT总拥有成本。
- ◆ CMDB配置管理数据库：Configuration Management Database，配置管理数据库包含一个组织的IT服务使用的信息系统的组件的所有相关信息以及这些组件之间的关系。

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，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。

“不懂就不折腾”

既然不懂投资，我就选择信任。
一觉醒来看到买的基金又赚钱了，
好的，那我继续睡会儿。

—— 中正达广基金

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 价值 | 平衡 | 快乐 | 爱
ZHONGZHENG DAGUANG FUND

证监会核准的独立基金销售机构（沪证监许可[2015]85号）

过往业绩不预示未来表现，市场有风险，投资需谨慎

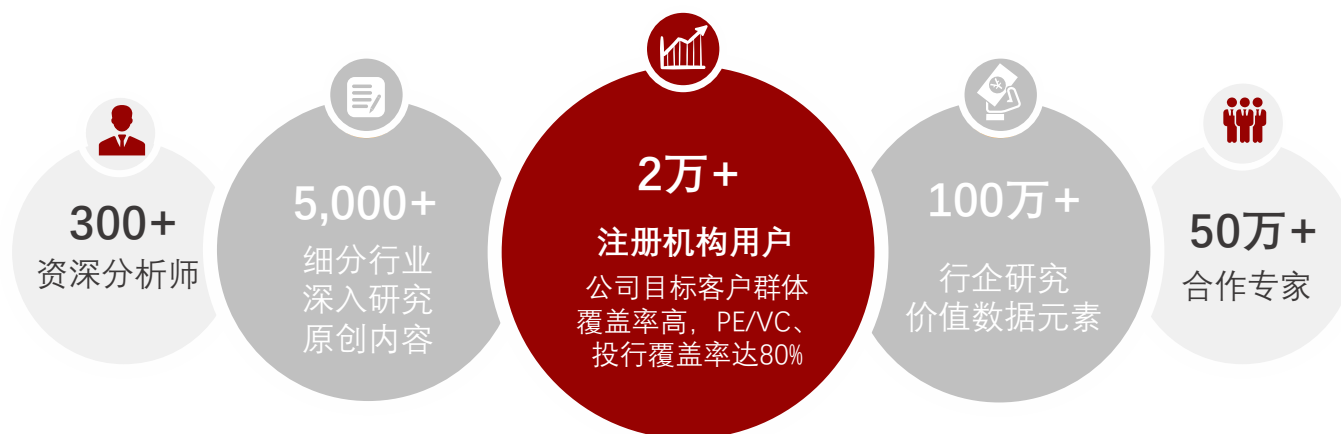
中正达广基金
充实美好生活

中正达广基金 价值 | 平衡 | 快乐 | 爱
ZHONGZHENG DAGUANG FUND

证监会核准的独立基金销售机构（沪证监许可[2015]85号）

头豹研究院简介

- ◆ 头豹是中国领先的原创行企研究内容平台和新型企业服务提供商。围绕“协助企业加速资本价值的挖掘、提升、传播”这一核心目标，头豹打造了一系列产品及解决方案，包括：**报告/数据库服务、行企研报服务、微估值及微尽调自动化产品、财务顾问服务、PR及IR服务**，以及其他企业为基础，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，通过开放合作的增长咨询服务等
- ◆ 头豹致力于以优质商业资源共享研究平台，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

研究咨询服务

为企业提供定制化报告服务、管理咨询、战略调整等服务

企业价值增长服务

为处于不同发展阶段的企业，提供与之推广需求相对应的“内容+渠道投放”一站式服务

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

研报阅读渠道

◆ 头豹官网：登录 www.leadleo.com 阅读更多研报

◆ 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

◆ 行业精英交流分享群：邀请制，请添加右下侧头豹研究院分析师微信



扫一扫
进入头豹微信小程序阅读报告



扫一扫
实名认证行业专家身份

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

头豹领航者计划介绍

头豹共建报告

2021年度
特别策划

Project
Navigator
领航者计划



每个季度，头豹将于网站、公众号、各自媒体公开发布**季度招募令**，每季公开**125个**招募名额



头豹诚邀各行业**创造者、颠覆者、领航者**，知识共享、内容共建



头豹诚邀**政府及园区、金融及投资机构、顶流财经媒体及大V**推荐共建企业

沙利文担任计划首席增长咨询官、江苏中科院智能院担任计划首席科创辅导官、财联社担任计划首席媒体助力官、无锋科技担任计划首席新媒体造势官、iDeals担任计划首席VDR技术支持官、友品荟担任计划首席生态合作官……



1

企业申请共建

2

头豹审核资质

3

确定合作细项

4

信息共享、内容共建

5

报告发布投放

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。



www.leadleo.com 400-072-5588

©2021 LeadLeo

85

头豹领航者计划与商业服务

研报服务

共建深度研报
撬动精准流量



传播服务

塑造行业标杆
传递品牌价值



FA服务

提升企业估值
协助企业融资



头豹以**研报服务**为切入点，根据企业不同发展阶段的资本价值需求，依托**传播服务**、**FA服务**、**资源对接**、**IPO服务**、**市值管理**等，提供精准的商业管家服务解决方案

资源对接

助力业务发展
加速企业成长



IPO服务

建立融资平台
登陆资本市场



市值管理

提升市场关注
管理企业市值



扫描二维码
联系客服报名加入



www.leadleo.com 400-072-5588

©2021 LeadLeo

读完报告有问题？ 快，问头豹！你的智能随身专家



扫码二维码即刻联系你的
智能随身专家

