

中国数智融合发展洞察

©2022.7 iResearch Inc.



VUCA时代，市场变化加速。企业需要更加敏捷而准确的数智化决策，这些决策应当是分钟级的而非天级的，应当是基于全量数据的而非局部数据的，应当是基于准确数据的而非基于“脏数据”的，应当是业务人员和数据分析人员任意发起的而非是通过复杂流程和多部门配合才能实现的。



传统的数仓或者湖仓分离架构让数智融合和企业敏捷决策变得困难：数据孤岛存在，决策无法基于全量数据；数据来回流转，成本高、周期长、时效差。基于存储-缓存-计算分离，湖-仓-AI数据统一元数据管理的Serverless，可在数据量、成本、效率、敏捷方面取得最优解。



开源为数智生态贡献重要力量，但这不预示所有企业需通过开源产品自建数智平台。实际上，大多企业聚焦自己核心业务，选择性能稳定、无须运维、数智融合、端到端自动化与智能化的商业化数智平台，ROI会更高。当然，平台应与主流开源产品具有良好继承性，如此，更加灵活开放，企业的IT人才补给成本也更低。

中国数智融合发展背景

1

企业数智融合的痛点及应对

2

数智融合典型实践

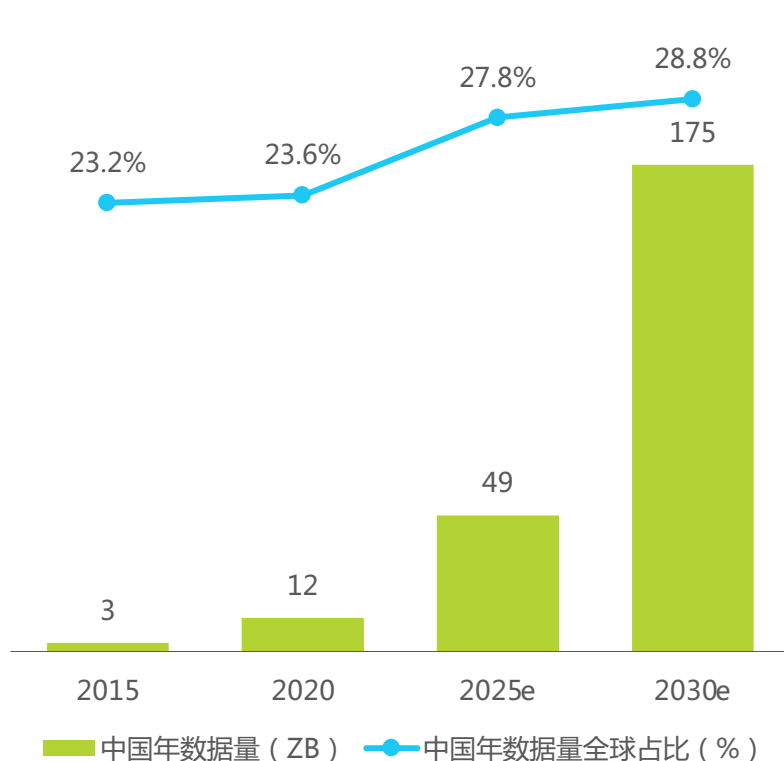
3

数据量和非结构化数据占比上升

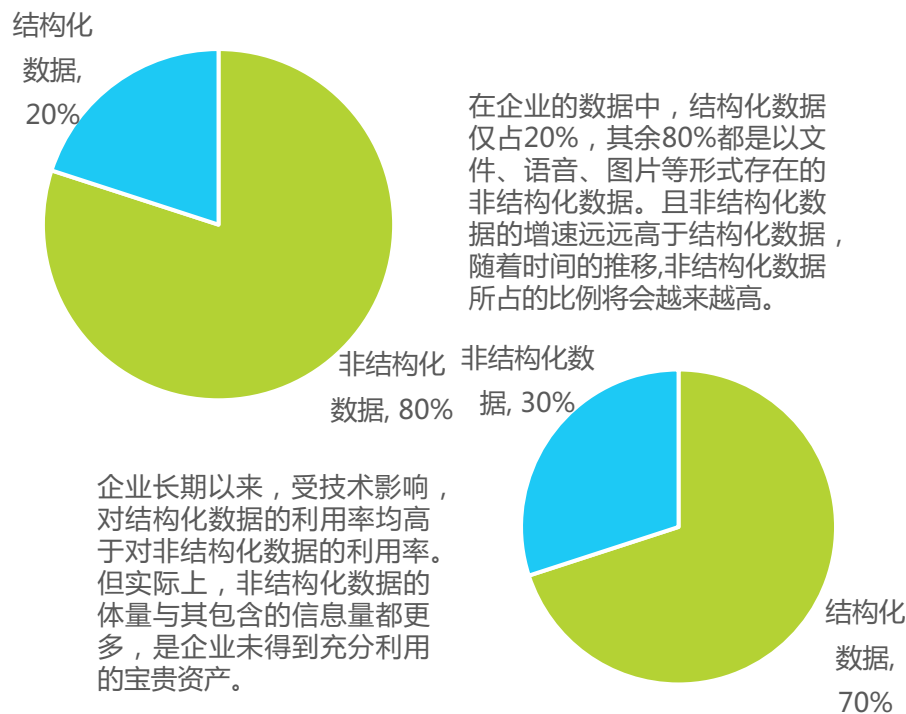
统一管理，统一查询使用，成为新的挑战

全球数据量以59%以上的年增长率快速增长，其中80%是非结构化和半结构化数据，中国数据量的上升较全球更为迅速。数据量和非结构化数据的上升，使得基于对象存储的数据湖越来越为普及。此时，如何使用统一管理，统一查询使用，成为新的挑战。

2015-2030年中国数据量规模及全球占比



企业内结构化数据与非结构化数据占比及使用情况



来源：中国电信招股说明书，艾瑞咨询研究院整理及绘制。

来源：艾瑞咨询研究院自主研究及绘制。

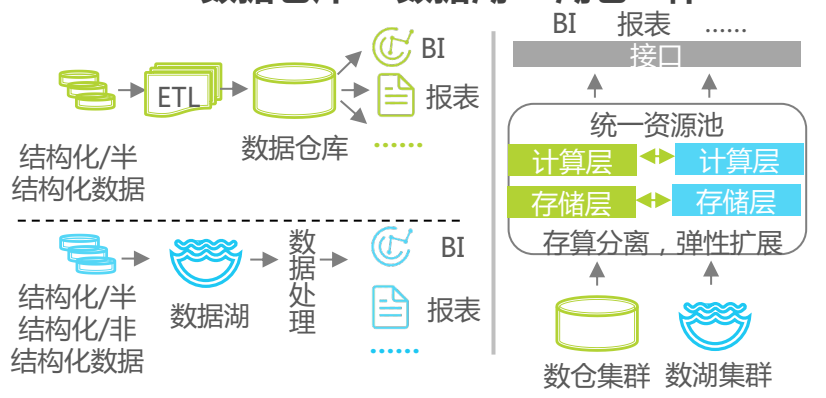
数据多源异构成为常态

数据从“汇聚才可被用”到“链接即可被用”

在传统数仓中，多源数据经ETL过程并集中入仓，方可被使用。该方式有许多不足：第一，因有复杂的ETL过程及大量数据的传输，数据实时性难以保障，因此分析常必须T+1才可完成；第二，数据的全量存储和存储成本之间难以取舍，因此必须提前抉择保留哪些数据，随着数据种类的逐渐增多，这很难做到；第三，对于异常值的下钻、回溯等，无法回溯到最为原始的数据。随着应用场景的增多，数据库的种类也逐渐丰富，如更适应物联网场景的时序数据库、更适应知识谱图应用的图数据库，等等。

综上，多源异构、分布存储、现用现传、统一查询与应用的架构，逐渐被敏捷型企业认可。

数据仓库vs数据湖vs湖仓一体



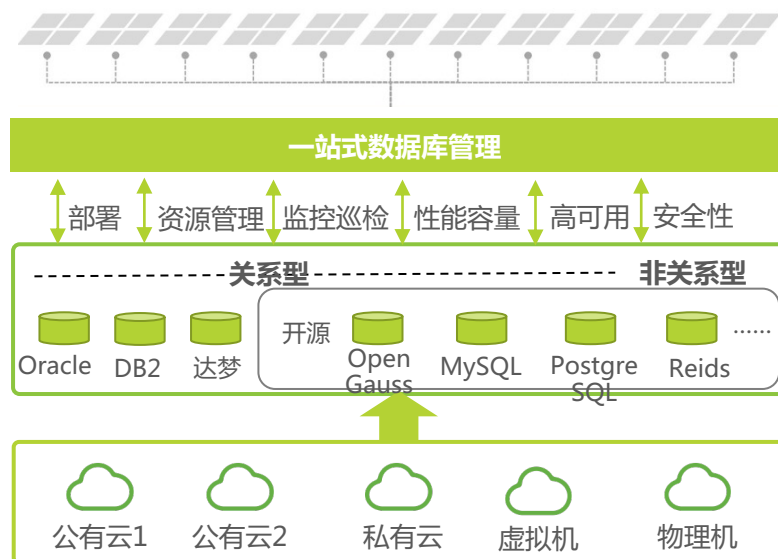
数据仓库vs湖仓一体

“湖仓一体”作为数据处理统一底座，提供实时处理多引擎、多数据类型能力，避免数据移动建模，降低数据处理的成本。

数据湖vs湖仓一体

“湖仓一体”弥补Hadoop下数据湖实时数据处理的缺失，降低事后数据治理难度，提升了大数据应用性能。

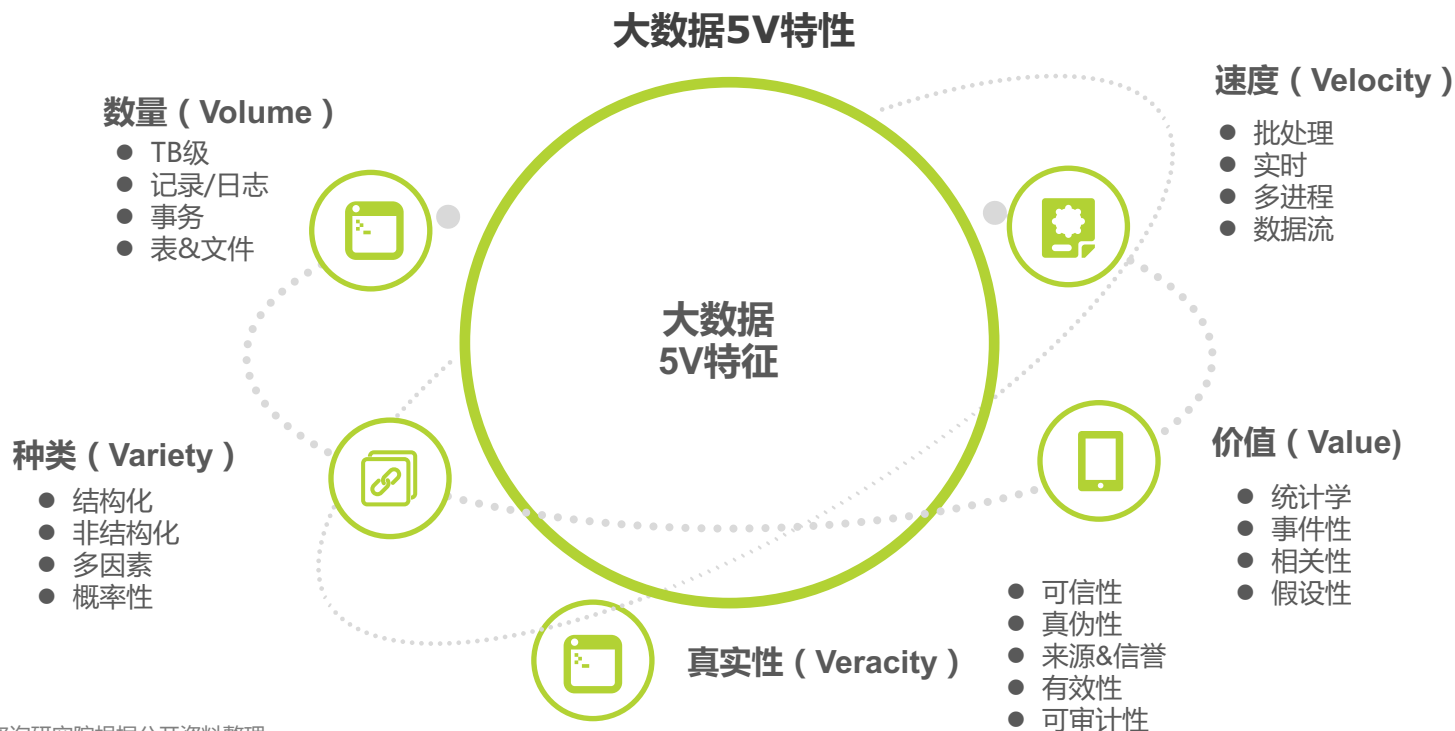
数据库的多源性



大数据的5V价值有待进一步释放

可从平台性工具入手，进而解决思维和技能的问题

大数据产业作为以数据生成、采集、存储、加工、分析、服务为主的战略性新兴产业，提供全链条技术、工具和平台，孕育数据要素市场主体，深度参与数据要素全生命周期活动，是激活数据要素潜能的关键支撑，是数据要素市场培育的重要内容。目前，大数据产业仍存在数据壁垒突出、碎片化问题严重等瓶颈约束，大数据容量大、类型多、速度快、精度高、价值高的5V特性未能得到充分释放。这其中既有思维、技能的要素，又有工具的要素，三者也并非割裂存在，一般来说，性能稳定、简单易用的全链条平台工具有助于消除思维的“不敢”和技能的“不会”，化解掉5V特性释放的原始阻力，使得大数据更加普适化。



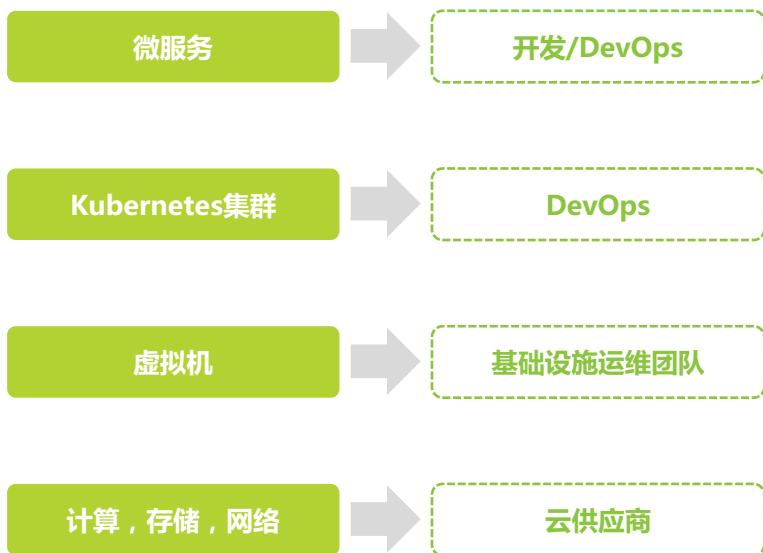
来源：艾瑞咨询研究院根据公开资料整理。

云原生：从微服务走向Serverless

从PaaS到FaaS，基础设施被更深层次地托管和“屏蔽”

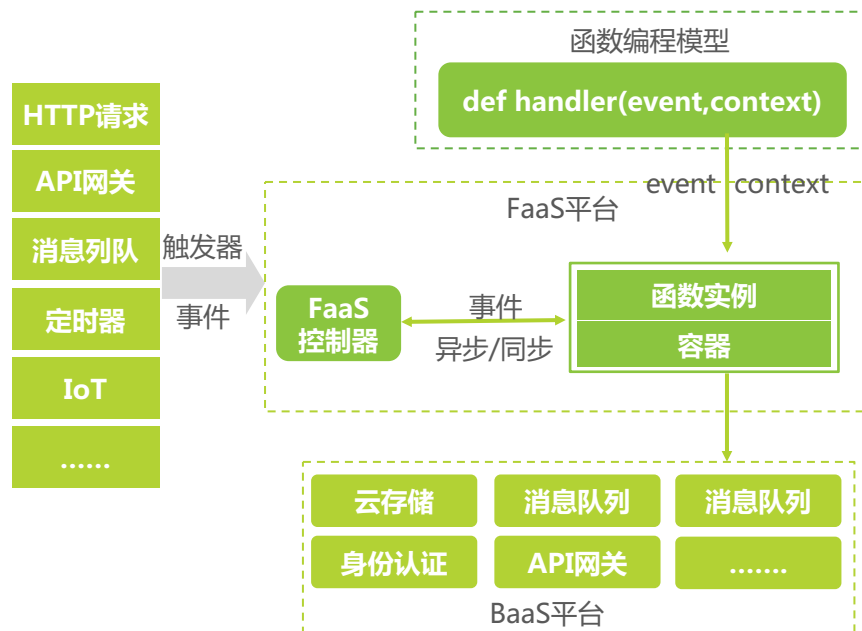
当前，微服务的生态和实践已经比较成熟，其设计方法、开发框架、CI/CD工具、基础设施管理工具等，都可以帮助企业顺利实施，然而其仍有许多不足：（1）粒度仍然比较大。（2）开发仍有较高门槛。（3）微服务基础设施管理、高可用和弹性仍然很难保证。（4）基础设施的成本依然较高。而Serverless中，开发者不再需要将时间和资源花费在服务器调配、维护、更新、扩展和容量规划上，这些任务都由平台处理，开发者只需要专注于编写应用程序的业务逻辑。如果再结合低零代码，则“编写应用程序”的难度也大为降低，企业内的技术人员更加贴近业务。

微服务中，大量运维仍未被托管



来源：《华为serverless核心技术与实践》，艾瑞咨询研究院整理及绘制。

典型的serverless架构



来源：《华为serverless核心技术与实践》，艾瑞咨询研究院整理及绘制。

人工智能：需要大规模准确数据哺育

人工智能应用引发数据治理需求

企业在部署AI应用时，数据资源的优劣极大程度决定了AI应用的落地效果。因此，为推进AI应用的高质量落地，开展针对性的数据治理工作为首要且必要的环节。而对于企业本身已搭建的传统数据治理体系，目前多停留在对于结构性数据的治理优化，在数据质量、数据字段丰富度、数据分布和数据实时性等维度尚难满足AI应用对数据的高质量要求。为保证AI应用的高质效落地，企业仍需进行面向人工智能应用的二次数据治理工作。

AI应用对数据治理需求

AI应用的数据要求

数据规模

传统数据治理多以人为面向对象，基于有限数据容量进行聚合类信息展示，AI可接纳数据量远远大于人所接纳的数据量和信息量，且**可用高质量数据越多，模型质量和准确性越好**。

数据类型

AI应用，尤其是知识图谱搭建，需要大量半结构化和非结构化数据支持来开展工作。因此AI应用在**结构化数据基础上，将半结构化或非结构化数据纳入数据源并支持上层分析应用**。

数据质量

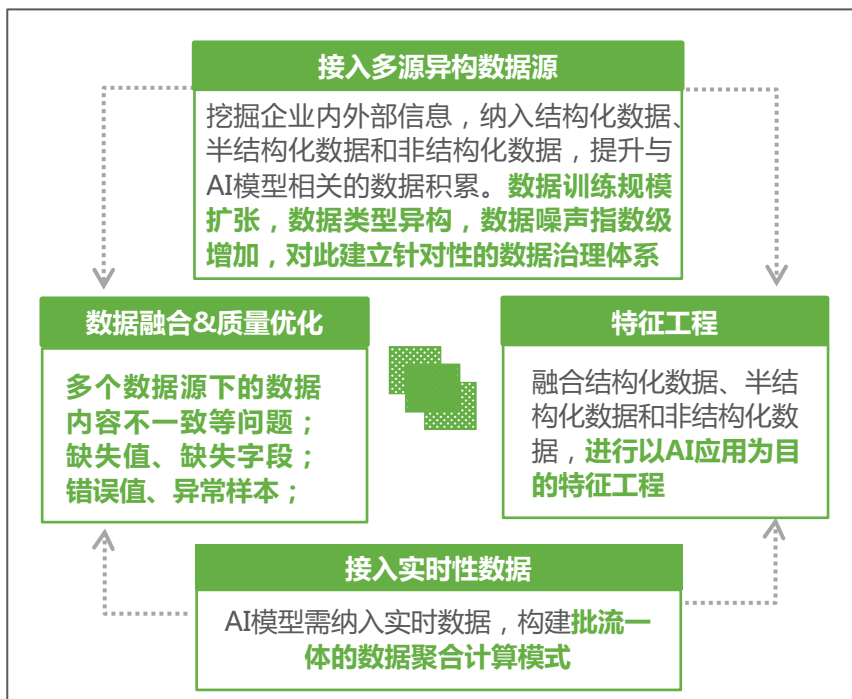
AI模型对数据高度敏感，其质量优劣极大程度影响AI模型的应用效果，因此AI数据源需极力规避“garbage in, garbage out”的问题发生，**多维度的质量检查成为必修课**。

数据实时性

AI模型对实时性要求高，大部分应用需基于实时数据实现分析、推荐和预警等目的，**支持AI应用的数据源更强调具备实时性接入能力**。

基于AI应用的数据治理需求

数据治理的需求传导



业务敏捷需要IT架构“去过程化”

通过抽象解耦、水平扩展、自动化与智能化实现去过程化

VUCA时代，市场变化加速，通过数据来分析和决策的需求，也有了更高的不确定性。当这些需求提出，通过一套复杂的IT流程和漫长的等待，变得不再现实，IT架构的去过程化变得极为重要。去过程化是指减少或完全去掉原始数据/原子能力与业务需求之间的中间数据/步骤，或使中间数据/步骤无须人为干预，自动化、智能化完成。其可实现架构的简单化、扁平化，同时可对业务需求实时响应，以进一步实现敏捷和创新。架构一开始就放弃“精细梳理方可使用”以及“梳理完成千万别动”思想，用全量原始数据保障读时模式，有助于打破“僵”与“乱”的悖论，使得企业用更少的“能量”便可以维持数字化系统的持续运行。

去过程化四大支撑

抽象与解耦

将IT架构抽象成存储、处理、应用三层，处理层又拆分成原子能力和数据模型，当不确定的需求来临，现将数据、能力、展现与应用形式进行组合进行处理。

稳定且可水平扩展的基础设施

应用的敏来自于基础的稳，上层的简来自于底层的繁，基础设施稳定性更加重要。

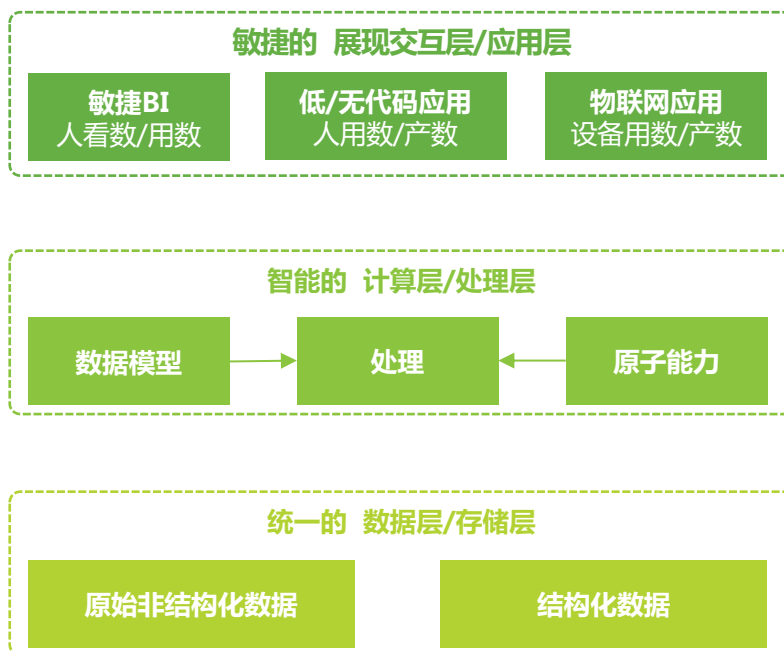
高性能计算与网络

存算分离和读时模式往往存在更多的重复数据传输与计算，对网络和计算要求更高。

自动化与智能化

通过智能化完成基础性能优化，降低硬件压力或硬件成本；通过智能化完成部分过程的自动化，从而屏蔽“人”视角下的该过程。

去过程化分层示例



相关标签

低代码/无代码
敏捷BI
GraphQL/JsonAPI

统一查询语言
统一接口
统一角色与权限

微服务
泛化模型
智能决策

冷热温数据分层
RDMA
存算分离

对象存储/数据湖
数据仓库
湖仓一体

中国数智融合发展背景

1

企业数智融合的痛点及应对

2

数智融合典型实践

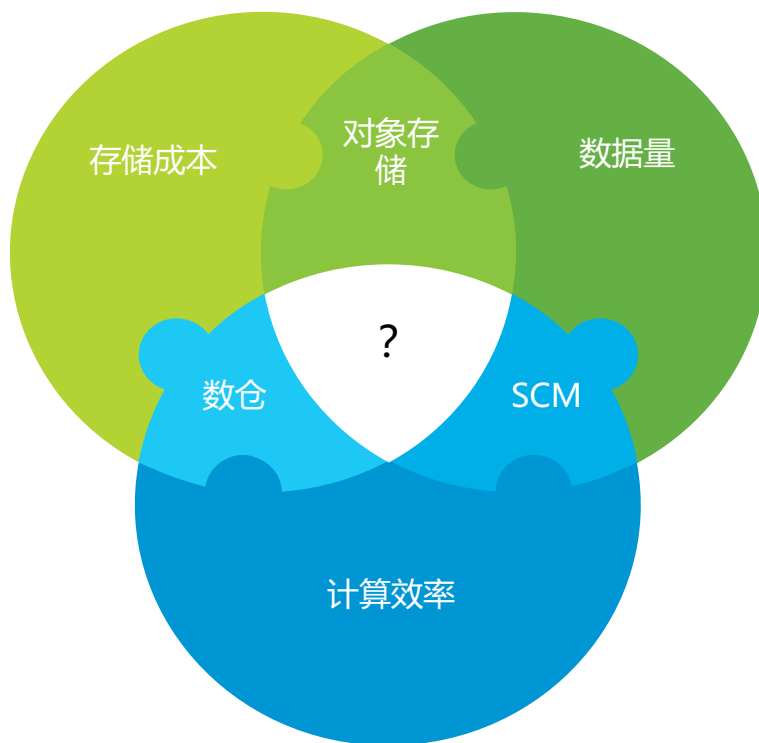
3

痛点一：数据量-成本-效率难以兼得

不可能三角需要更高维的技术去打破

在传统架构中，数据量、存储成本和计算效率是一组不可能三角。如果不考虑数据量和数据类型，那么一个传统的数仓或者单体的DBMS即可满足；不考虑计算效率，那么基于HDFS或者公有云对象存储即可满足，当下价格仅约0.1元/G/月，并持续下降，归档存储等价格更低；不考虑存储成本，可使用非易失性存储，其拥有一般硬盘的无限容量和断电保护特性，却有接近于内存的性能。

传统架构下数据量、存储成本和计算效率的不可能三角



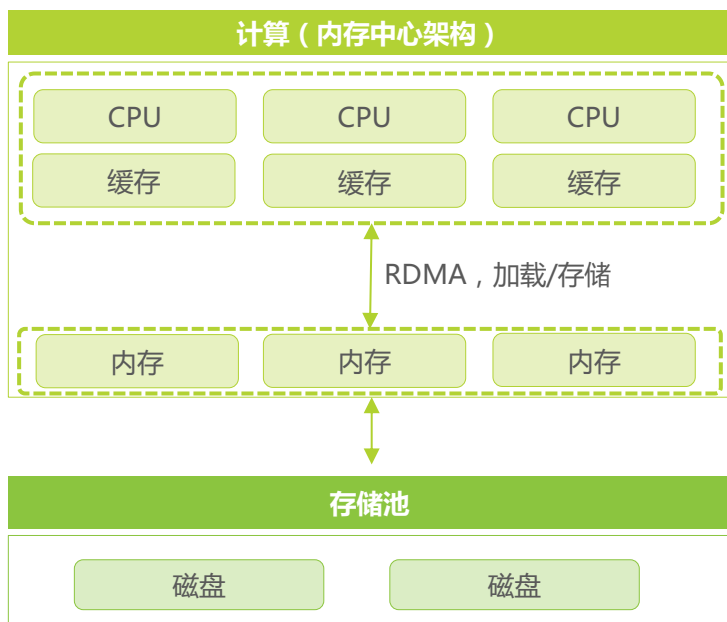
应对一：存储-缓存-计算三层分离

以内存为中心的架构，在大数据量下降低成本、保持性能

为了使数据充分共享，降低均摊成本且打破数据孤岛，存算分离架构产生，存储和计算各自弹性伸缩，按需使用。但此时，因存储拉远，IO成为瓶颈，性能有所下降，因此需要缓存层来存储高IO的热数据，并最终形成以内存为中心的架构。

从必要性看，以计算为中心架构已经无法适应当前数据生态发展：数据方面，大数据、人工智能等以数据为中心的工作负载快速发展；云方面，数据湖存算分离架构存储访问性能低，不支持实时分析。从可行性看，介质、网络、协议的高速发展驱动架构转型：SCM填补了内存纵向扩展的介质空白；缓存一致性标准的争夺进入白热化；高速内存直连协议及技术（如华为1520，InfiniBand，Converged Ethernet）使得内存的远程直接访问不再是障碍。

以内存为中心的系统架构示意图



来源：艾瑞咨询研究院自主研究及绘制。

以内存为中心的架构技术优势

核心价值与技术优势

性能：

内存密集型和分布式应用减少数据搬运/拷贝、提升性能

- HPE以内存为中心系统提升Spark性能15倍
- HPE分布式图计算场景性能提升128倍
- MemVerge分级大内存提升深度学习性能20倍

成本：

内存池化和升级

- 内存池化提升内存资源利用率
- 通过更低成本的SCM获得大容量内存扩展能力

故障解耦合：

CPU和内存的故障不相互影响

弹性增强：

CPU和内存可独立扩缩容

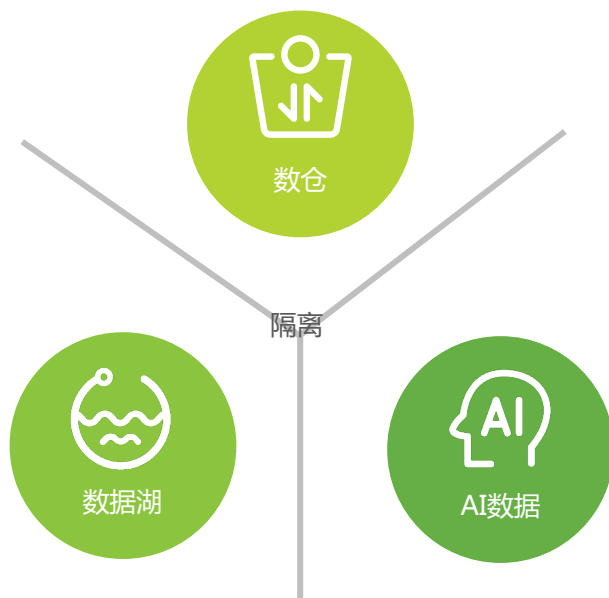
来源：艾瑞咨询研究院自主研究及绘制。

痛点二：仓-湖-AI数据形成新孤岛

要么隔离，要么迁移，均无法适应全量、敏捷、低成本需求

数据分析和AI分析经过多年的发展，出现了很多面向不同任务的专用数据系统：数仓系统处理结构化数据，规模不够大；基于对象存储的大数据系统处理海量数据和非结构化数据；AI系统一般是数据存储在本地。这些专用系统要么无法打通，形成新的数据孤岛，要么不同业务的开发要迁移数据，耗费存储和网络资源，数据准备慢、等待周期长，且面临后期数据不一致的风险，发现异常时数据的下钻、溯源等也相对困难，无法适应市场环境快速变化下敏捷数据分析的需求。

数仓、数据湖和AI数据形成新的数据孤岛



数据迁移导致副本增多、资源浪费、管理困难



应对二：统一元数据到中心节点

Master-Slave架构，以集中管理代替集中存储

把数仓、数据湖、AI数据的目录、数据权限、事务一致性、多版本管理等能力都统一到一个中心点，依赖于这个中心点来访问数据，这样数据的利用就不会被孤立的系统束缚。这种分布式存储，统一管理的Master-Slave架构，类似于计算领域的Mapreduce。这种方式：首先，可以打破数据孤岛、让一份数据在多个引擎间自由共享，例如同一个表格可以被不同的分析工具做分析，既可以跑数仓任务，也可以做大数据和机器学习任务，不同的用户角色不管用什么工具访问数据，都有一致的权限，一致的事务控制；其次，可以避免数据来回迁移而造成资源的浪费；再次，任何环节都可以看到自己权限下的全量数据，例如ML工程师可以利用整个数据湖的数据做特征训练；最后，所有模型均基于唯一事实来源（原始数据），避免不同团队基于不同数据分析造成结果不一致，且一旦发现异常可以便捷地下钻、回溯。

统一元数据示意图



痛点三：开源产品丰富，但开发运维难

开发成本高，运维成本高，技术与与时俱进难，风险大

尽管在云、数、智体系下，开源产品极为丰富，但企业安全、稳定地驾驭，TCO并不低。**在开发上**，企业一般需花费20-1000人力年的时间，不能满足业务敏捷性；**在运维上**，人工运维，事后补救，宕机频繁，耗时耗力；**在技术更新上**，开发人员难以与时俱进，资源浪费严重；**在IT风险上**，企业将面对IT团队自身的风险（复杂架构下，团队离职无人接手）以及开源产品的漏洞风险（如log4j4漏洞事件），还可能面对因经验不足选型错误的风险；**在体验上**，因产品自产自销，复用率低，技术团队一般只保障基础需求，对于降低业务人员使用难度、提升使用体验的附加性需求响应度低。并且，这些基础的开发、运维等，与企业核心业务常无必然联系，并不会带来企业核心竞争力的提升，导致企业数智化的ROI较低。

企业利用开源产品自行搭建数智平台面临的困难



应对三：DataOps和MLOps融合

享受成熟产品的红利，兼顾与开源产品的继承和包容性

企业在数智化选型中，应首先明确自身的核心竞争力和能力边界，摒弃“重即好”思想，以更加轻盈的Serverless、Lowcode/Nocode、SaaS等方式享受社会分工和先进技术的红利。以数智融合为例，抛开IaaS层，企业自研还需掌握Kubernetes+Docker生态、Java+Hadoop生态、Python+Pytorch/Tensorflow生态、SQL生态.....即便成功对接，往往也离好用、敏捷相去甚远，最终往往只形成指标长期不变的静态报表。而与此同时，业界已存在较为领先的一站式数智平台，让数据工程师甚至业务人员以简单、熟悉的工具/语言，甚至拖拉拽即可在全域数据内使用预置AI算法，打通大数据和人工智能，使得DataOps和MLOps融合，使数据和模型的开发成本大为降低，周期大为缩短。

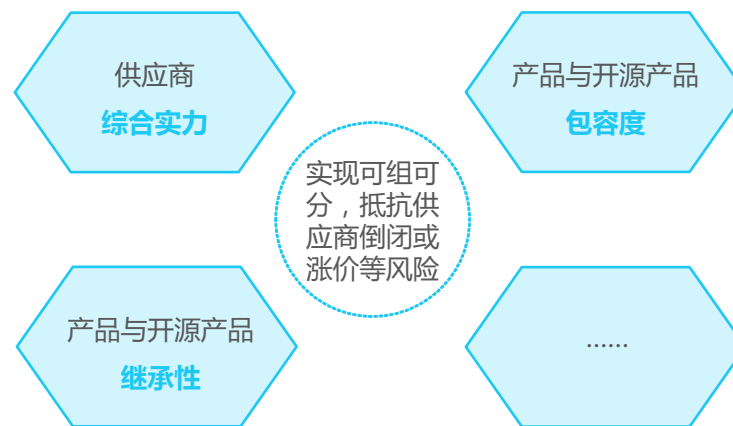
企业选择基于开源产品自研，不少时候是出于一种怕被“绑架”的防御心态，以化解供应商倒闭或涨价等风险。为此，企业可从供应商综合实力，与开源产品的包容度和继承性等方面综合考虑，做到可组可分，灵活装配。

将大数据和人工智能打通的DataOps和MLOps

简单、易用、全局能力调用



灵活装配

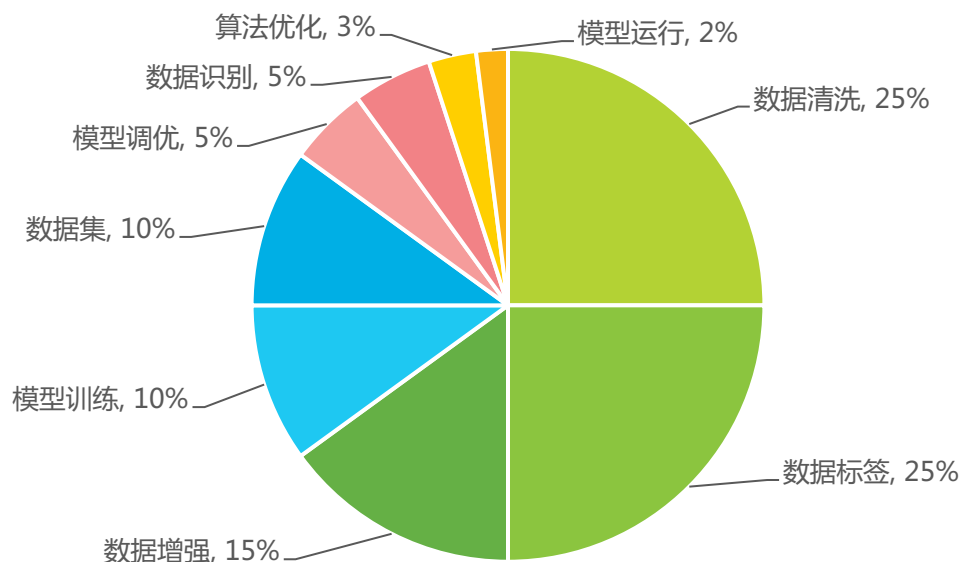


痛点四：数据准备工作复杂低效

数据长生命周期决定了其复杂性

数据质量至关重要，错误数据致使企业做出错误决策。数据的生成、采集、存储、加工、分析、服务、安全、应用长生命周期，使得数据的准备工作复杂、低效。例如，数据工程师开发大量的ETL任务，依赖大量算力资源，运行成本高，作业管理复杂，时间周期长，而此时数据分析师和AI开发者都需要等待ETL任务执行完，才能做相应的分析工作和建模工作。Cognilytica调研显示，机器学习中，超过80%的时间都用于数据的准备，预示着大量的数据工作其实与企业的经营目标并不相关，只是不得以而为之。并且，当任务不能便捷地执行和即时的反馈，偏业务侧的数据分析师常主动放弃“不太重要的”需求和对数据的深入探索。显然，这些都不符合企业数智化转型的真正目标。

机器学习中数据准备工作的时间分配



应对四：端到端的自动化与智能化

低零代码实现自动化，AI反哺数据实现智能化

实现数据应用的敏捷化和去过程化，需要在整个数据链条的端到端实现自动化和智能化。自动化一般用低/零代码实现：一方面可以屏蔽软硬件差异和复杂的底层技术，以便于理解的拖拉拽和少量代码，来降低使用门槛；另一方面，可以基于规则，配置自动化的工作流，以ifttt的方式减少重复工作量。智能化是指基于半监督或无监督的学习，自动发现数据管理中的规则，在Data4AI的同时，实现AI4Data，目前人工智能已经用于数据集成、数据质量、数据建模、数据安全与访问控制、数据关联、数据洞察等多个场景中。另外，低/零代码常和人工智能结合使用：将人工智能的统计意义上的规则，融入到低/零代码的逻辑化的流程中。

人工智能在数据管理中的应用（AI4Data）

 数据质量检查	 数据模型管理	 主数据管理	 元数据管理	 数据安全
<ul style="list-style-type: none">• 基于机器学习，确定数据阈值。• 对完整性、规范性、一致性、准确性、唯一性、时效性进行检查。• 脏数据自动识别订正。	<ul style="list-style-type: none">• 利用聚类 and 知识图谱确定实体间关系。• 利用知识图谱等进行数据血缘分析。	<ul style="list-style-type: none">• 人工智能帮助企业识别主数据。• 人工智能帮助定义和维护数据匹配规则。	<ul style="list-style-type: none">• 人工智能实现对非结构化数据的采集和关键信息的提取。• 人工智能帮助维护元数据。• 人工智能帮助实现元数据的整合。	<ul style="list-style-type: none">• 隐私级自动标注。• 数据传输监控。

中国数智融合发展背景

1

企业数智融合的痛点及应对

2

数智融合典型实践

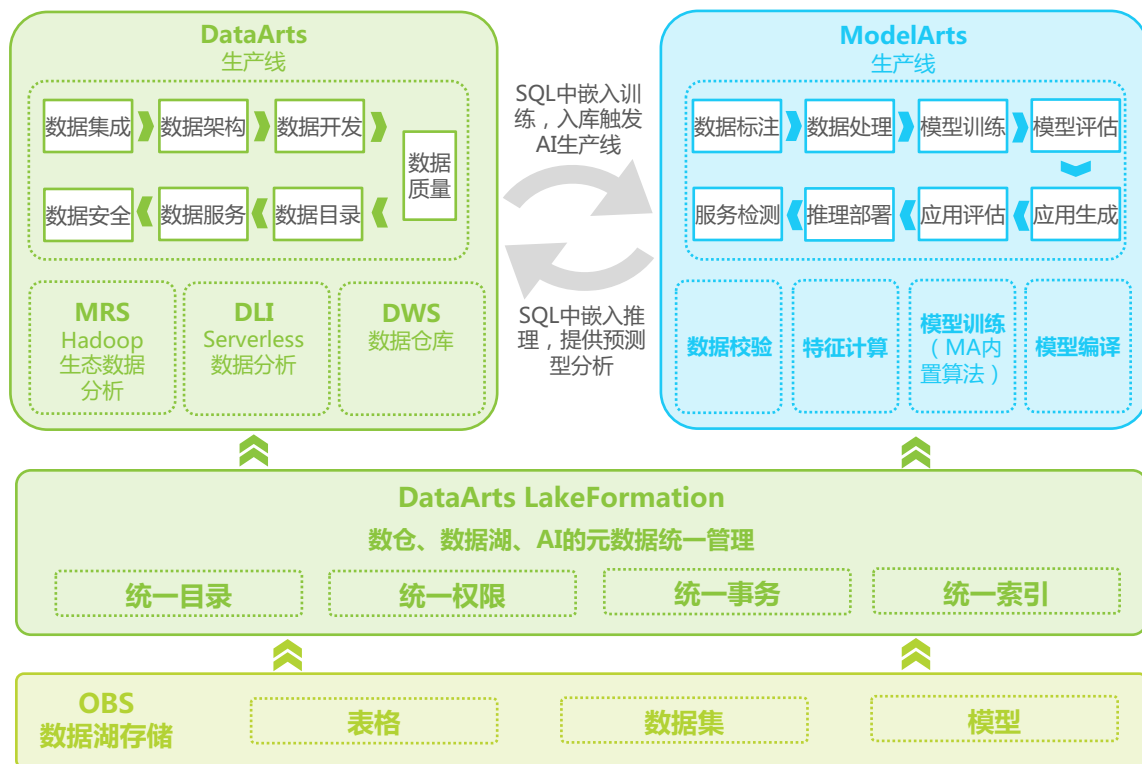
3

华为云DataArts+ModelArts

独家创新架构：兼顾成本与性能，实现统一管理、一数多用，数智融合实现敏捷用数，全流程实现自动化与智能化

华为云通过DataArts和ModelArts融合架构，打通了大数据和人工智能。统一了元数据，使得一数多用，打破数据孤岛，同时避免数据来回迁移。存储-缓存-内存三层分离，兼顾存储成本和计算性能。DataOps和MLOps结合，让企业不同部门、不同角色可以以擅长的方式敏捷用数。低零代码和人工智能反哺数据（AI4Data），使得全流程实现自动化与智能化。

华为云数智融合平台创新架构



来源：华为云，艾瑞咨询研究院整理及绘制

华为云数智融合平台核心价值



来源：华为云，艾瑞咨询研究院整理及绘制

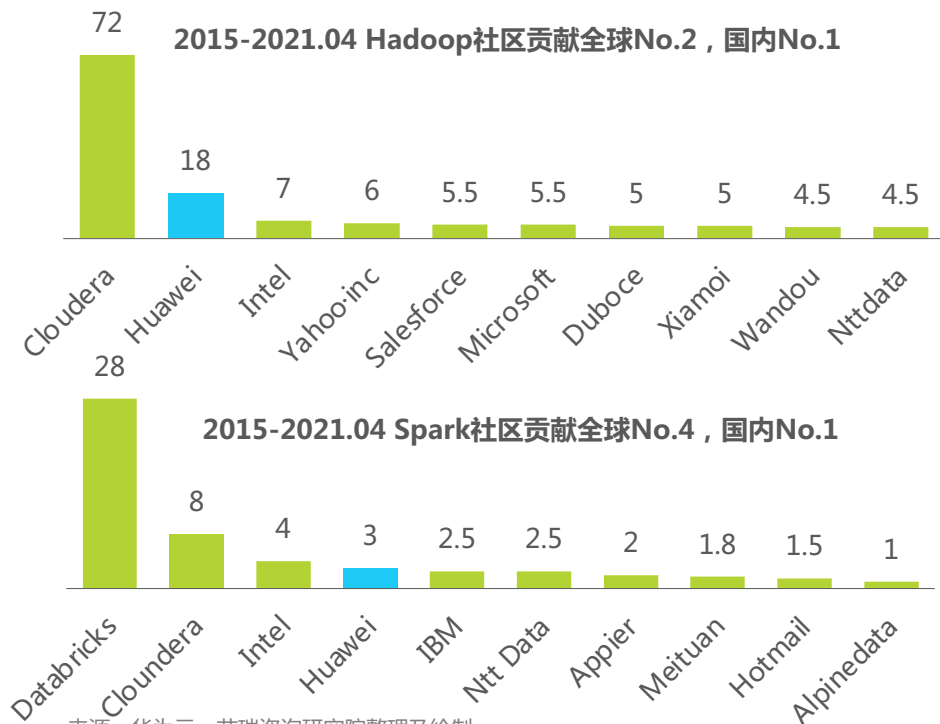
华为云DataArts+ModelArts

软硬融合，行业实践经验和开源生态结合

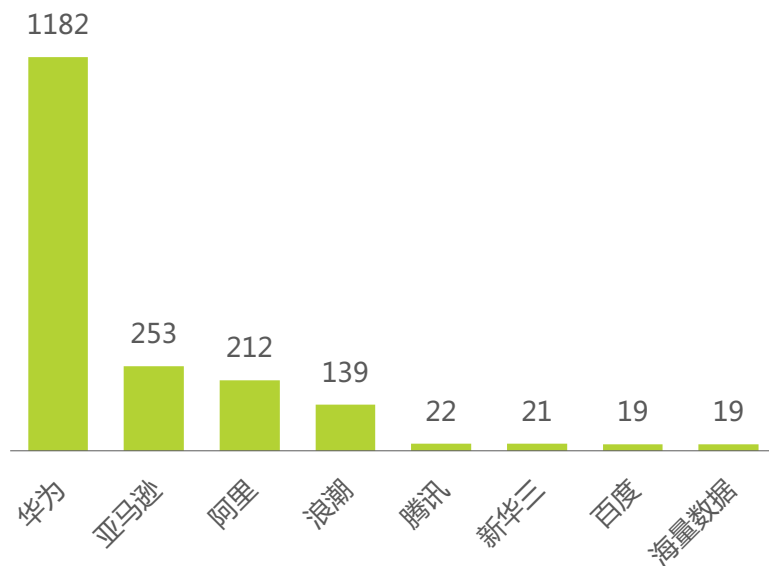
另外，华为云在软硬融合、行业实践经验和开源生态方面，也具有一定优势：在软硬融合方面，华为云对计算、存储和网络的底层技术更为擅长，例如，在内存池化的关键技术之一RDMA的专利申请数量上，华为具有明显优势。在行业实践经验上，华为云一直秉承“一切皆服务”的原则，深扎行业，在互联网和传统政企方面均积攒了大量实践经验，并把这些经验返回、沉淀到产品中。在开源生态方面，华为云在Hadoop和Spark社区中贡献度均较高，这使得华为云对这些开源产品在安全、稳定性等方面有着更深的技术理解，同时使得DataArts对这些开源产品的主流版本的兼容性更好。

华为在开源社区的贡献度

高性能RDMA网卡技术专利数量



来源：华为云，艾瑞咨询研究院整理及绘制。



来源：智慧芽专利数据库，艾瑞咨询研究院整理及绘制。

IT服务业 (1/2)

数智融合助力IT服务商降本增效提质

IT服务行业范围较广，包括前期的IT咨询与培训，中期的定制开发、系统集成、部署实施，后期的IT运维升级、IT运营管理，以及贯穿全程的IT安全保障等。对于IT服务企业而言，随着业务量增长，普遍面临“数据基数庞大，搬迁上云难”“数据持有成本高”“数据治理不佳，形成数据孤岛”“产品开发技术门槛高，运维成本高，存在安全隐患”“数据准备工作复杂，难以聚焦业务本身”等问题。数智融合下的存算分离、元数据统一、DataOps 和 MLOps融合、端到端的自动化与智能化等创新技术，帮助开发人员实现所想即所得，助力IT服务企业实现降本增效、业务创新、提升客户满意度，从而在产业发展转折点抓住市场机会，实现企业品牌质的飞跃。

IT服务业环节



IT服务业发展痛点vs数智融合

1. 数据体量大，存储成本高，算力要求高，运维开销大

“存储+缓存+计算”三层分离架构提供存算性能的弹性伸缩和按需使用，通过serverless模式实现秒级弹性扩容和管理运维全托管，满足IT服务业存算性能的同时降低运维开销。

2. 数据治理不佳，数据在系统间流转不畅，形成数据孤岛

人工智能算法模型应用到数据全生命周期治理，通过对元数据的统一管理，解决传统数据分析和AI模型之间“数据搬家”的问题，实现数据在不同引擎间的自由流动，消除数据孤岛。

3. 产品开发门槛高，数据准备耗时长，难以聚焦业务本身

低/无代码的集成开发平台，为使用者屏蔽底层技术，同时将AI能力植入到数据开发过程，将能力服务化、技术组件化，通过分层解耦和复用，即插即用，敏捷交付，降低开发与运维成本。

来源：艾瑞咨询研究院自主研究及绘制。

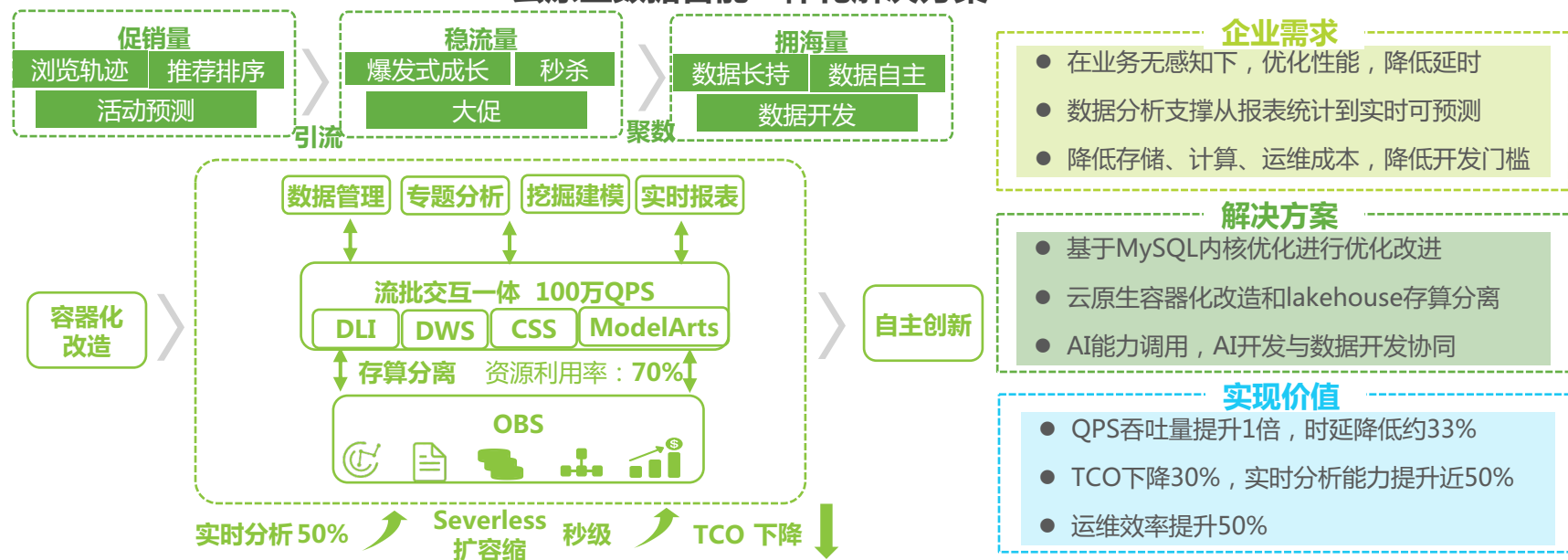
来源：艾瑞咨询研究院自主研究及绘制。

IT服务业 (2/2)

数智融合在IT服务企业的落地

梦饴集团定位于新电商基础设施提供商，提供基于SaaS店铺工具外加一体化商品与服务结合的供应链解决方案，旗下饴店以去中心化的品牌特卖平台，拥有超百万店主，月交易总额达10亿人民币；爱库存提供国际化库存交易供应链平台。随着业务增长，公司遇到“云迁移难，数据持有成本高、数据治理差、产品开发门槛高，开发周期长且后期运维成本高”等难题。依托华为云在行业市场的云迁移经验和丰富的服务和工具，实现高效、稳定、可靠的云迁移；通过存算分离技术，实现秒级弹性扩缩容，降低用数成本；通过端到端的自动化和智能化，将AI算法模型应用到数据全生命周期治理，为AI开发提供高质量数据；通过DataOps 和 MLOps融合，使大数据开发和AI开发协同，降低开发门槛，聚焦业务本身。华为云一体化解决方案为梦饴集团带来“资源高效”、“应用敏捷”、“业务智能”、“安全可信”等多方面的智能升级。

云原生数据智能一体化解决方案



来源：梦饴集团，艾瑞咨询研究院整理及绘制。

网约车行业（1/2）

业务模式灵活创新，数智融合赋能网约车行业个性化开发

网约车经营服务依托互联网技术构建服务平台，整合供需信息，匹配最优的车辆和驾驶员，提供预约出租汽车服务。网约车产业链包括上游车辆及技术服务供应商，中游网约车服务商以及下游应用场景。受国家政策鼓励，中国获许的网约车平台数量从2020年底的214家增长至2022年3月底的267家，未来市场空间广阔。作为互联网时代产物，网约车行业与车联网、大数据等密不可分，面临数据安全、客户需求、数据打通、应用创新等层面的众多挑战。存算分离技术提供性能的弹性伸缩，满足用车高峰的算力高并发，统一元数据打破大数据、数仓、AI的“数据孤岛”，增强网约车企业数据洞察，DataOps 与 MLOps的融合降低开发技术门槛，方便企业开发人员聚焦业务创新，敏捷交付。

网约车行业产业链



网约车行业发展痛点vs数智融合

1. 传统数仓存算弹性不足，难以应对高峰期高并发需求

网约车服务场景先天具备波峰波谷特性，除资源利用率要求外，还具有实时性和长时执行的要求。数智融合下的“存储+缓存+计算”三层分离具备弹性伸缩、持续服务和节省成本等优势，与网约车行业场景需求天然契合。

2. 多平台数据难以打通复用，“数据孤岛”现象严重

网约车行业是车联网、大数据、人工智能等新技术的融合产物。具备强互联网和大数据属性，网约车企业运营需对接多个数据平台，数据难以连接互动造成“数据孤岛”。数智融合统一元数据，让数据在多个引擎间流动共享，充分发挥数据资产价值。

3. 业务模式创新多样，多源数据、技术等导致系统“熵增”

为提升市场竞争力和市占率，网约车行业下游应用场景不断推陈出新，业务模式的创新多样必然导致数据体系的“熵增”。数智融合下的DataOps 与 MLOps深度融合，数据工程师可使用熟悉的工具调用AI能力，协同数据开发与AI开发，支撑行业业务模式的不断创新。

来源：艾瑞咨询研究院自主研究及绘制。

来源：艾瑞咨询研究院自主研究及绘制。

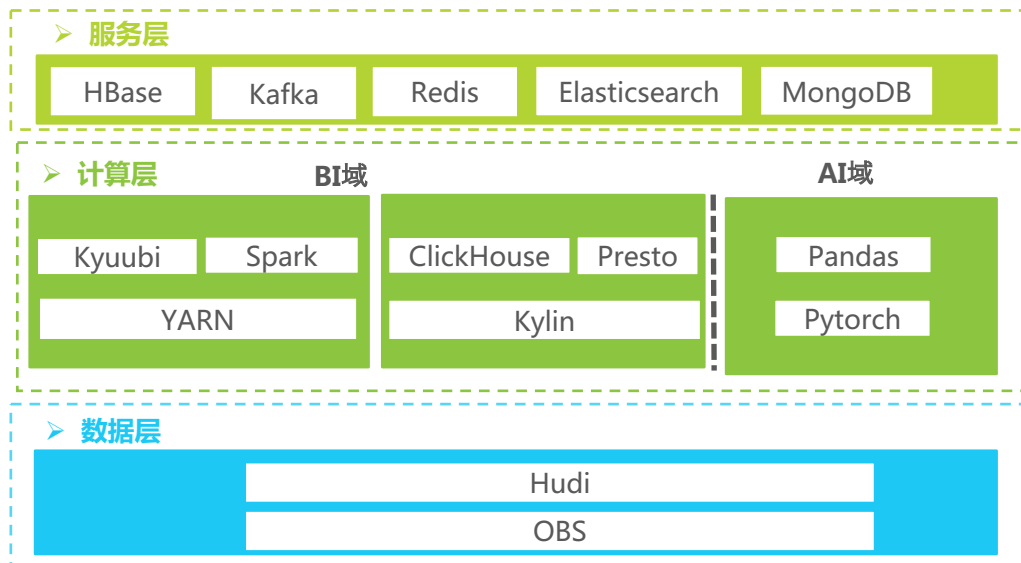
网约车行业（2/2）

数智融合在网约车企业的落地

T3出行是南京领行科技股份有限公司打造的智慧出行生态平台，公司以“科技引领 愉悦出行”为使命，致力于成为能够为用户提供“安全、便捷、品质”出行服务的科技创新型企业。截止2022年7月，T3出行登陆全国91个城市，累计注册用户超1亿，单日订单峰值破300万单。随着业务扩张和用户数量增加，T3出行原系统在支撑海量数据和提供出行服务方面出现“传统数仓难以解决“长尾支付””、“AI架构无法为行车安全提供支撑”、“存算一体架构无法应付订单峰值”等问题。为满足业务发展，提供安全高效的出行服务，T3出行不断对出行产品迭代更新，而华为云数智融合产品在这段升级之旅中发挥了巨大作用。

T3出行+华为云智能数据湖MRS

湖仓一体（对象储存+数据湖格式+存算分离）



来源：T3出行，艾瑞咨询研究院整理及绘制。

T3出行+华为云DRS数据迁移解决方案

痛点一：数据库种类多，迁移难度大

- 华为云DRS支持其他云、本地IDC、ECS自建数据库等不同平台之间的迁移，支持60+类不同的数据库链路，充分满足T3出行的不同数据库迁移需求。

痛点二：时间紧任务重，要在30h内完成

- 华为云DRS针对不同实例，通过优化参数、同步方式和架构，并结合其特有的限流能力，确保迁移期间源端业务正常平稳运行，成功在指定时间内完成全量数据同步。

痛点三：业务不停机，系统持续平稳运行

- 华为云通过RDS适配canal能力，结合DRS高可靠和自动化运维能力，大大提高数据订阅稳定性。

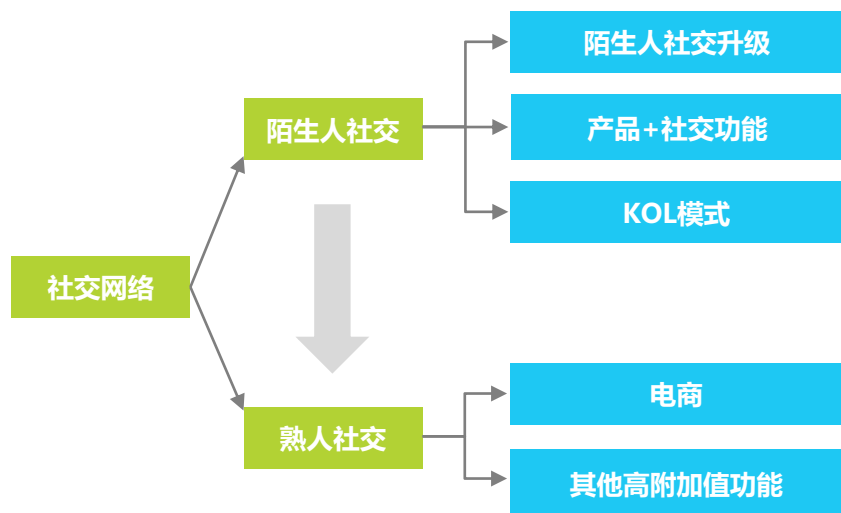
来源：T3出行，艾瑞咨询研究院整理及绘制。

社交网络行业（1/2）

数智融合赋力社交网络行业向兴趣化、垂直化创新发展

社交网络源于互联网服务，为拥有相同兴趣与活动的用户提供各种联系、交流的交互通路。目前中国社交网络行业基于5G通信、人工智能、虚拟现实等先进技术，为用户提供沉浸式的社区服务。随着行业的内容生态向兴趣化和垂直化方向演进，用户群体的进一步泛化以及用户数量的爆发式增长，社交网络行业面临用户数据安全、后台数据治理效果不佳、存储资源紧张、算力不足以应对业务敏捷相应等诸多挑战。此时，数智融合提供的存算架构分离、基于人工智能算法模型的数据全生命周期治理、产品的低门槛敏捷开发等先进技术，满足了社交网络行业对于如VR虚拟社交等可预见的新型社交场景的技术支撑，以及对于Z世代移动网民全覆盖的需求。

社交网络行业产业链



社交网络行业发展痛点vs数智融合

1. 用户体量大，后台数据量剧增，存储资源紧张，算力不足

“存储+缓存+计算”三层分离架构提供存算性能的弹性伸缩和按需使用，通过serverless模式实现秒级弹性扩缩容和管理运维全托管，支撑社交网络行业存算性能的同时降低运维成本。

2. 数据多源异构显著，数据治理难度大，信息孤岛现象严重

社交网络行业数据类型丰富，形式多样，多源异构现象普遍，数智融合通过统一元数据到中心节点，打破仓-湖-AI数据间的数据孤岛，实现一份数据在多个系统间自由流动。

3. 产品开发门槛高，开发人员疲于数据准备，难以聚焦业务本身

数智融合提供DataOps和MLOps融合，不同角色各用所长，各取所需，将AI能力植入到数据开发过程，将能力服务化、技术组件化，通过分层解耦和复用，让开发人员聚焦核心业务本身。

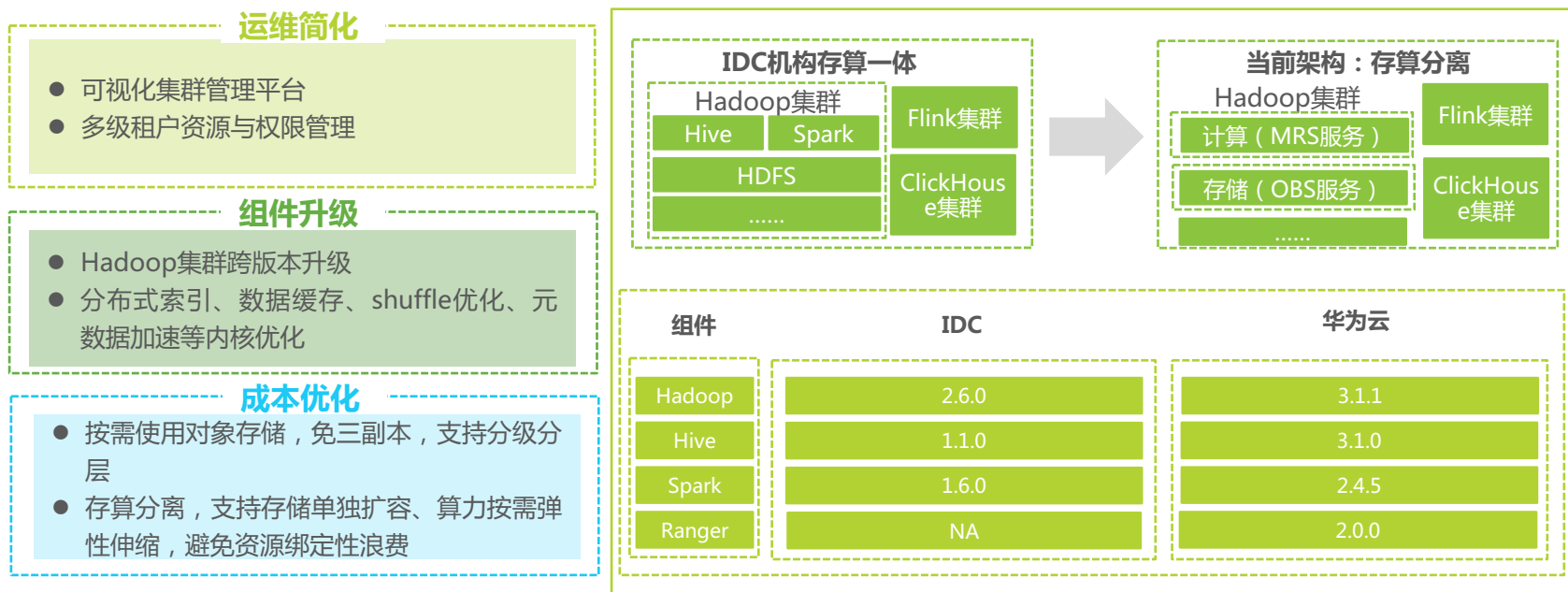
社交网络行业（2/2）

数智融合在社交网络行业的落地

脉脉是1.2亿职场人都在用的职场社区和求职平台，基于“实名/职业认证”和“人脉网络引擎”帮助职场人拓展人脉、交流合作、求职招聘，收获机遇。通过职场社交和招聘两大核心业务，发挥自身产品优势，帮助职场人士和企业打造职业形象、树立品牌，提前进行高效连接，收获职场机遇。2020年注册用户数量破亿，成为“内容社区+招聘服务”职场全价值链平台，实现了自身的高速发展。随着业务场景和用户数量增加，脉脉原有的系统架构在支撑巨量用户数据和提供社交平台服务方面面临“存储资源紧张，算力弹性不足，运维负担大”、“数据安全与数据治理要求升级”等挑战。华为云数智融合实现了脉脉的智能用数和业务的敏捷响应，在竞争激烈的社交网络行业勇争上游。

脉脉+华为云数智融合核心价值

脉脉+华为云智能数据湖MRS



来源：脉脉，艾瑞咨询研究院整理及绘制。

来源：脉脉，艾瑞咨询研究院整理及绘制。

艾瑞新经济产业研究解决方案



行业咨询

- 市 场 进 入 为企业提供市场进入机会扫描，可行性分析及路径规划
- 竞 争 策 略 为企业提供竞争策略制定，帮助企业构建长期竞争壁垒



投资研究

- IPO行业顾问 为企业提供上市招股书编撰及相关工作流程中的行业顾问服务
- 募 投 为企业提供融资、上市中的募投报告撰写及咨询服务
- 商业尽职调查 为投资机构提供拟投标的所在行业的基本面研究、标的项目的机会收益风险等方面的深度调查
- 投后战略咨询 为投资机构提供投后项目的跟踪评估，包括盈利能力、风险情况、行业竞对表现、未来战略等方向。协助投资机构为投后项目公司的长期经营增长提供咨询服务

关于艾瑞




艾瑞咨询是中国新经济与产业数字化洞察研究咨询服务领域的领导品牌，为客户提供专业的行业分析、数据洞察、市场研究、战略咨询及数字化解决方案，助力客户提升认知水平、盈利能力和综合竞争力。

自2002年成立至今，累计发布超过3000份行业研究报告，在互联网、新经济领域的研究覆盖能力处于行业领先水平。

如今，艾瑞咨询一直致力于通过科技与数据手段，并结合外部数据、客户反馈数据、内部运营数据等全域数据的收集与分析，提升客户的商业决策效率。并通过系统的数字产业、产业数据化研究及全面的供应商选择，帮助客户制定数字化战略以及落地数字化解决方案，提升客户运营效率。

未来，艾瑞咨询将持续深耕商业决策服务领域，致力于成为解决商业决策问题的顶级服务机构。

联系我们 Contact Us

 400 - 026 - 2099

 ask@iresearch.com.cn



企 业 微 信



微 信 公 众 号

法律声明

版权声明

本报告为艾瑞咨询制作，其版权归属艾瑞咨询，没有经过艾瑞咨询的书面许可，任何组织和个人不得以任何形式复制、传播或输出中华人民共和国境外。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

免责条款

本报告中行业数据及相关市场预测主要为公司研究员采用桌面研究、行业访谈、市场调查及其他研究方法，部分文字和数据采集于公开信息，并且结合艾瑞监测产品数据，通过艾瑞统计预测模型估算获得；企业数据主要为访谈获得，艾瑞咨询对该等信息的准确性、完整性或可靠性作尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或所表述的观点均不构成任何建议。

本报告中发布的调研数据采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，该数据仅代表调研时间和人群的基本状况，仅服务于当前的调研目的，为市场和客户提供基本参考。受研究方法和数据获取资源的限制，本报告只提供给用户作为市场参考资料，本公司对该报告的数据和观点不承担法律责任。

为商业决策赋能

EMPOWER BUSINESS DECISIONS

