

2025

中国Data&AI数据 基础设施白皮书

AI创生时代，数智融合新底座

出品机构：甲子光年智库

发布时间：2025.10

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

在AI创生时代，Data&AI数据基础设施是新世界的底座

- 当前，全球正经历一场由地缘政治重塑和人工智能技术革命双重驱动的深刻变革。这两大趋势交织并进，正以前所未有的速度重构世界经济与社会发展格局。面对全球化向区域化、供应链向本土化转型的外部环境，以及人工智能从一项前沿技术向核心生产力跃迁的内部需求，中国正处于一个关键的历史转折点。麦肯锡研究显示，生成式AI有望为全球经济贡献约7万亿美元的价值，而中国有望贡献其中约2万亿美元，接近全球总量的三分之一。IDC的预测也印证了这一趋势，预计到2028年，全球AI领域IT总投资将增至8159亿美元，而中国的AI总投资规模将突破1000亿美元，五年复合增长率高达35.2%。这些数据表明，AI技术革命不仅是一场技术革新，更是一场将催生全新生产模式和经济增长点的大变局。
- 在这场变革中，各类组织机构，无论是企业还是政府，都面临着巨大的挑战和机遇。埃森哲的研究表明，中国企业正经历一场“挤压式转型”，转型窗口期更短，挑战更多。传统的、孤立的数据系统（如数据仓库或割裂的数据平台）已难以支撑AI应用对实时性、多模态数据处理和高弹性算力的需求。企业迫切需要打破数据孤岛，解决高质量数据供给不足、模型与业务场景割裂以及数据安全合规等痛点。这些挑战正倒逼数据基础设施实现一次范式跃迁，从单一的数据存储与分析工具，向支撑全业务流程的智能化底座演进。
- 数据厂商要想抓住机遇并在新一轮竞争中脱颖而出，关键在于构建一种全新的数据基础设施。这种基础设施应具备以下核心能力：
 - 必须将组织内部的数据生产、加工治理与业务终端AI模型的应用视为一个动态持续的生产过程，以满足业务的动态连续性需求。
 - 需要摒弃当前许多企业采用的单场景Agent应用建设模式。正如传统数据库时代的“数据烟囱”，这种模式将导致数据和能力再次割裂，无法形成全局协同效应。Gartner预测，到2028年，至少15%的日常工作决策将由Agentic AI自主做出，这要求基础设施必须能够支持大规模、跨场景的智能体部署。
 - 一体化的数据基础设施需要具备集中式建设和集中式管理的能力基盘，从而确保数据资产的统一治理、安全合规与高效利用。同时，它也必须面向业务侧即价值端，形成分散式的赋能能力，让各业务单元能够灵活调用和创新。
- 在技术层面，新型数据基础设施的核心路径是“湖仓一体”与“AI原生”的深度融合。它将数据湖的开放性和灵活性与数据仓库的结构化管理、ACID事务特性相结合，形成一个统一的数据底座。同时，它将AI模型的训练、部署与数据治理紧密耦合，实现从数据到智能的无缝流转。这种架构不仅是技术上的整合，更是理念上的革命，它为Data&AI深度融合体提供了坚实基石。支撑数据价值逐级跃升，从“点”（功能提效）到“线”（组织创新），到“面”（行业协同），再到“体”（产业带动），最终形成支撑未来智能社会的国家战略，成为智能革命的核心引擎。
- 本白皮书旨在系统梳理中国Data&AI数据基础设施的发展现状、技术路径与实践经验，为大型组织提供前瞻性参考与可行性指南。人类开启信息化革命后的几十年里，全球竞争格局一直都由美国引领。然而，在进入AI新时代后，中国持续追赶的成果逐渐显露，尤其在通信、互联网、人工智能等领域出现了引领全球的创新技术产品和服务，而Data&AI数据基础设施正处在中国这些优势领域的交叉地带，必将成为中国在AI竞争中持续保持优势、赢下智能革命的必争之地。我们相信，当数据与AI在新底座的支撑下真正融合一体，新世界的无限可能便将由此开启。

数据应用与人工智能从相互独立到逐渐融合

- 在数据作为生产要素被广泛认知后，数据应用和人工智能加速融合。高质量数据集的积累加快了大模型的成熟，大模型又反过来让各组织机构更好地发挥数据要素的生产力，开始从深度融合走向一体化。

数据应用与人工智能的融合，推动架构与能力实现跨越式发展

- 第一阶段的感知智能主要依赖分类的非结构化数据，第二阶段生成式AI推动多类混合数据的治理体系发展完善，到了第三阶段的Agentic AI则对多模态数据融合与动态调度提出更高要求。数据应用与人工智能技术正从“AI+数据”的松散组合迈向“数据即AI、AI即平台”的深度融合阶段。
- 新阶段需要一个AI原生的、一体化的技术底座平台，为Agentic AI的场景化应用提供持续的工程化能力和稳定的技术底座支撑，类似建筑的地基工程，为组织长期提供多模态数据融合处理、动态异构资源调度，构建“数据-模型-Agent-业务”闭环系统。这个平台我们称为Data&AI一体化基础设施，它正飞快成为新一代智能系统的核心，围绕“数据流+Agent流”帮助组织构建闭环、自主的智能生态。

数据基础设施成为AI时代的“核心生产工具”

- 数据基础设施逐步从“支撑决策”升级为“协同驱动智能”，成为组织机构实现数据价值流转与智能闭环的“核心生产工具”。
- 其中，Data&AI一体化数据基础设施是最终能够长期持续支撑人工智能规模化落地的一体化基础软件平台，不仅能提供基础底座支撑，更重要的是通过平台持续、实时、动态工程化能力实现打通数据存储、治理、计算与AI模型开发的全链路，构建出了“Data for AI”和“AI for Data”的双向赋能体系。

Data&AI数据基础设施的价值链条呈现“点-线-面-体”递进

- Data&AI数据基础设施通过湖仓一体架构统一纳管全域数据资产，并以AI原生设计支撑模型的高效训练与实时推理，实现数据到智能的闭环转化，驱动业务持续创新，进而支撑数据价值逐级跃升，从“点”（功能提效）到“线”（组织创新），到“面”（行业协同），再到“体”（产业带动），为数字经济高质量发展提供软件技术基础设施，是智能革命的核心引擎。

Data&AI数据基础设施的能力体现在开发、架构、调度、AI原生和安全运营

- Data&AI数据基础设施让数据平台和AI智能工具融合成为统一系统性底座，其关键能力主要体现于融合开发、架构、资源调度、AI原生和安全运营五个方面。五个能力方向侧重点不同，各厂商在其专业方向上各有侧重。

目录 >>>>

Part 01

新趋势：
AI 时代，数据范式迈入数据基础设施阶段

Part 02

新定义：
Data&AI 数据基础设施，支撑 AI 规模化落地

Part 03

新范式：
Data&AI 数据基础设施的核心能力模型

Part 04

新应用：
Data&AI 数据基础设施典型应用

目录 >>>>

Part 01

新趋势

AI 时代，数据范式迈入数据基础设施阶段

1.1 数智融合成为时代特征

1.2 双向互动成就范式跃迁

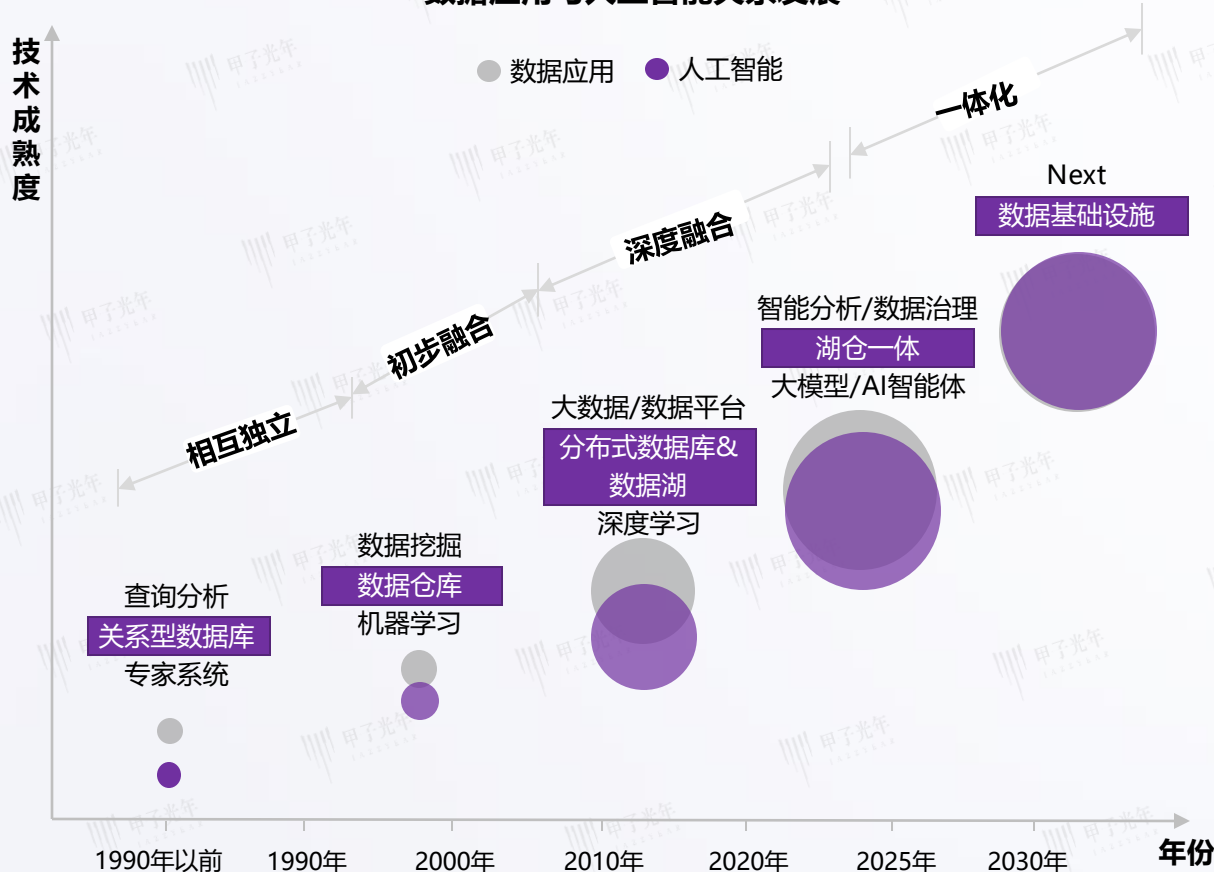
1.3 应用深化推高融合要求

1.4 数智一体化成行业共识

数据应用与人工智能从相互独立到逐渐融合

- 数据应用和人工智能技术出现已经很久，但最初二者并没有太多交集。数据应用和早期人工智能技术处于“各自为政”状态。
- 随着社会信息化程度提高，数据资源不断增多，依赖数据的人工智能算法，如深度学习开始崭露头角；同时数据分析也越来越多吸纳人工智能算法模型，发展出数据挖掘技术。数据应用与人工智能开始了融合之路。

数据应用与人工智能关系发展



资料来源：甲子光年智库整理

- 随着互联网、移动互联网出现，全社会数据出现爆炸式增长，大数据和数据平台概念开始出现，同时人工智能进入依赖海量数据的深度学习阶段，数据应用和人工智能技术迈入深度融合阶段。
- 在数据作为生产要素被广泛认知后，数据应用和人工智能加速融合，高质量数据集的积累加快了大模型的成熟，大模型又反过来让各组织机构更好的发挥数据要素的生产力，数据应用和人工智能开始从深度融合走向一体化。
- 支撑数据应用和人工智能的关键基础设施也经历了关系型数据库、数据仓库、分布式数据库及数据湖、湖仓一体的多阶段发展。

数据应用的发展是人工智能技术发展的重要基础条件

人工智能技术发展在本世纪经历三次范式跃迁：

- 1.感知智能：**此阶段主要解决的是让机器能够感知和理解外部世界的信息，如图像、声音、文本等。该阶段AI依赖相互独立的图像、音频、文本等非结构化数据。
- 2.生成式AI：**此阶段的重心是让机器能够进行逻辑推理、生成新的内容，并处理更加复杂和抽象的任务。该阶段AI依赖结构化、单/多模态数据集。
- 3.Agentic AI：**这是目前AI发展的最新趋势，其核心是让AI具备规划、决策和执行并完成复杂任务的能力。该阶段AI依赖多模态数据为主。



资料来源：甲子光年智库整理

数据应用的发展也经历了三次范式跃迁：

- 1.数据库阶段：**主要以传统关系型数据库、数据仓库为主，辅以分布式文件系统、NoSQL数据库等。数据形态以结构化数据为主，与感知AI数据形态融合度不高。
- 2.数据平台阶段：**出现云数据仓库、数据湖等新载体，其核心思想是打破数据孤岛，通过统一的数据治理和建模，将数据能力以服务形式提供给业务部门。
- 3.数据基础设施阶段：**是“湖仓一体”与AI原生融合的新型架构，旨在将数据的全生命周期管理与AI模型的全生命周期管理融合到同一个平台上，支撑企业业务在AI应用端侧场景实现持续、稳定的落地服务。

- 数据应用与人工智能技术的发展呈现出协同演进的趋势，每一次AI范式的跃迁本质上都建立在数据范式的进化基础上。

数据应用与人工智能的融合，推动架构与能力实现第三次范式跃迁

数据应用和人工智能范式跃迁



资料来源：甲子光年智库整理

- 从感知智能依赖分类的非结构化数据，到生成式AI推动多类混合数据的治理体系，再到Agentic AI对多模态数据融合与动态调度提出更高要求，数据应用与人工智能技术正从“AI+数据”的松散组合迈向“数据即AI、AI即平台”的深度融合。
- Agentic AI需要与真实世界进行实时、动态的交互，这对其数据处理能力提出了更高要求。它不仅需要处理多模态数据（文本、图像、音频、视频），还需要处理实时流数据，并能够根据环境变化进行动态调度。这是“数据基础设施”所应具备的核心能力。
- 数据基础设施是一个AI原生的、一体化的平台，旨在为Agentic AI等新兴应用提供底层的技术支持。它与人工智能技术基于多模态数据融合、动态调度，构建“数据-模型-Agent-业务”闭环系统。
- 第三阶段的到来标志着Data&AI一体化基础设施成为新一代智能系统的核心，未来平台将围绕“数据流+Agent流”构建闭环、自主的智能生态。

数据基础设施成为数据要素化阶段支撑智能化变革的“核心生产工具”

- 随着数字化迈向智能化阶段，人工智能的开发和应用对数据能力提出更高要求，尤其在多模态、高质量、实时反馈等方面。

中国数字化水平演进历程与数据应用需求情况



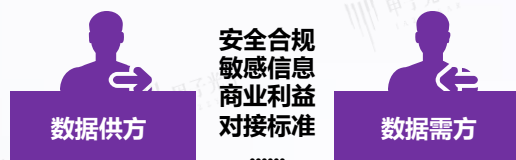
资料来源：甲子光年智库整理

- 产业数字化阶段的数据平台以存储和分析为核心，难以支撑模型训练、智能体部署等新型智能化场景，这促使组织机构需要进一步提升数据能力，将数据平台从支撑性、辅助性工具向AI原生融合的智能底座演进。
- 数据基础设施因此从“支撑决策”升级为了“协同驱动智能”，成为组织机构实现数据价值流转与智能闭环的“核心生产工具”。
- 正如生产工具是衡量一个社会或组织生产力水平的核心标志，数据生产工具的先进程度也是衡量社会数字化、智能化发展水平的核心标志。

人工智能应用深化到城市、产业层级，依赖数智一体化基础设施

产业痛点

- 在国家加快培育数据要素市场、推动人工智能工程化应用的背景下，从制度机制到平台支撑仍存在关键瓶颈，严重制约了数据资源价值在产业级层面的释放与AI能力的高效转化，面临的典型数据问题可归结为以下三类：
- 需求拉动型障碍：数据“拿不出、调不动”；
- 供给推动型失焦：数据“起什么作用、怎么用”；
- 供需定向型低效：数据“范围窄、复用差”。



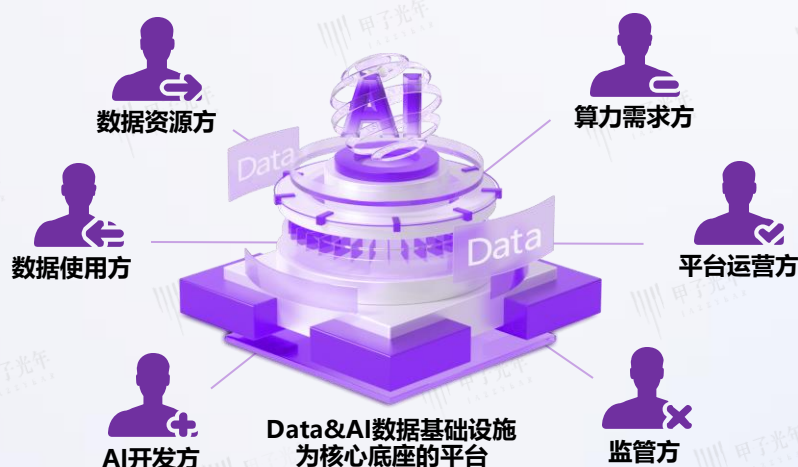
产业方案

- 构建一个能够统筹制度保障与工程落地的核心平台——监管沙盒；
- 城市级人工智能数据应用开发平台正是在该背景下应运而生的基础性工程。平台通过打通“产品交易—模型调优—合规评测”全流程闭环路径，来集成制度资源、平台能力、数据场景，破解数据价值释放与AI能力落地间的断链问题，推动构建全国数据要素与大模型融合发展工程。

技术难题

- 行业数据规模化聚集、治理、资产化及参与模型训练等落地难题；
- 卫星遥感、具身智能、自动驾驶、医疗健康、金融保险、工业制造等各行业异构数据资源统一接入与集中管理；
- 多类型GPU/CPU异构算力；
- 及时识别潜在风险并，确保数据合规性与算法可信性。

城市级人工智能数据应用开发平台示意



数据平台厂商持续补强AI能力，Data&AI一体化成为行业共识

- 市场需求推动数据平台的定位转变，从“被动的容器”升级为“主动的生产力载体”。数据平台的核心作用不再局限于存储数据，更要能生产、处理面向AI原生的数据，成为数据的“产地”与“工艺”。
- 行业响应用户需求逐步形成共识，全球多家领先厂商不约而同都通过一系列战略性收购，持续增强自身AI能力，以形成从数据接入、处理到数据、AI融合开发和应用部署的一体化能力，清晰地展现了Data&AI深度融合已成为数据平台行业的共识。
- 国际代表厂商Databricks公司从非结构化数据出发向结构化数据和AI能力扩展；Snowflake公司则从结构化数据出发向非结构化数据和AI能力扩展。
- 国内代表厂商阿里云通过自研+收购（如Data Artisans）构建起了Data&AI一体化端到端全栈能力；科杰科技主要通过自研完善了Data&AI一体化端到端基础软件能力。

代表性公司Data&AI一体化布局情况

Databricks	 大规模、复杂数据处理计算能力，深耕 非结构化数据 与实时分析	• Tabular 提高数据湖中的数据互操作性，提升 结构化数据处理能力	• Neon 为开发AI智能体应用提供高性能数据支撑	• MosaicML 为湖仓一体平台注入 生成式AI模型定制能力	• Tecton 为客户提供端到端的特征管理能力
	 处理海量结构化数据的能力，在 结构化数据处理和BI领域 占据统治地位	• Applica 增强了处理 非结构化数据 的能力	• Streamlit 填补了平台在数据可视化和应用开发方面的能力缺失	• TruEra 提供模型治理、性能监控和可信AI能力 • Neeva 注入了先进的AI搜索技术	• Myst.AI 为平台注入时间序列预测和AI建模能力 • LeapYear 强化了平台在数据安全、隐私保护和合规性方面的能力
•上一阶段核心能力		•融合开发	•平台架构	•智能原生	•安全运营
科杰科技	 以“湖仓一体”为架构核心打造的支持多模态计算和混合部署的Data&AI一体化平台	自研多模态计算引擎核心，提升文本/图像/音视频等多模态数据融合开发能力 • Data studio& AI studio	自研AI-in-Lakehouse技术架构，将湖仓一体引擎、OLAP数据分析和 AI 技术统一 • AI-in-Lakehouse	自研构建AI-Native为核心，涵盖MaaS自推理、Agent自迭代、模型训推一体的LLOPS，MaaS-Agent全链路智能能力 • Keen AI/Keen DSP	自研实现Data&AI双向就绪，集中管控与分散赋能并行能力 • Keen Asset/ Governance
	 超大规模、高并发、企业级数据处理与云原生调度能力	获得来自Apache Flink社区的流处理技术 • Data Artisans	自研EB级云原生数据仓库、实时湖仓一体分析引擎和海量对象存储 • Hologres 3.0	自研千亿参数模型；整合自研Easy系列低代码平台和DeepRec高性能推荐引擎 • 通义大模型/PAI平台	自研隐私计算平台，自研数据运维与数据安全解决方案 • Data Trust

资料来源：甲子光年智库整理

目录 >>>>

Part 02

新定义

Data&AI 数据基础设施，支撑AI规模化落地

2.1 赛道定义：Data&AI数据基础设施，AI时代的“核心数字底座”

2.2 核心特征：Data&AI数据基础设施，AI原生的一体化架构

2.3 边界厘清：Data&AI数据基础设施与传统大数据平台的差异

2.4 价值体现：Data&AI数据基础设施的五层价值体系

Data&AI数据基础设施，AI时代的“核心数字底座”

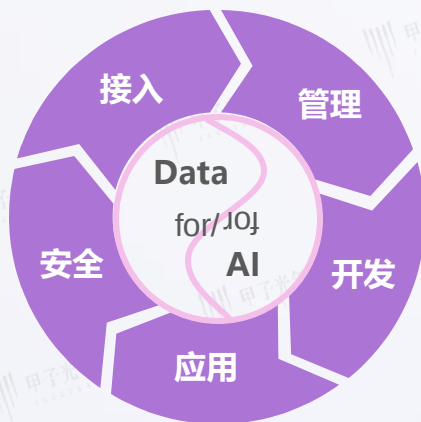
- 国家数据基础设施是从数据要素价值释放的角度出发，面向社会提供数据采集、汇聚、传输、加工、流通、利用、运营、安全服务的一类新型基础设施，是集成硬件、软件、模型算法、标准规范、机制设计等在内的有机整体。
- 国家数据基础设施在国家统筹下，由区域、行业、企业等各类数据基础设施共同构成。网络设施、算力设施与国家数据基础设施紧密相关，并通过迭代升级，不断支撑数据的流通和利用。

国家数据基础设施概念——软硬件+服务+管理的生态



资料来源：《国家数据基础设施建设指引》，甲子光年智库整理

Data&AI数据基础设施概念——基础软件平台



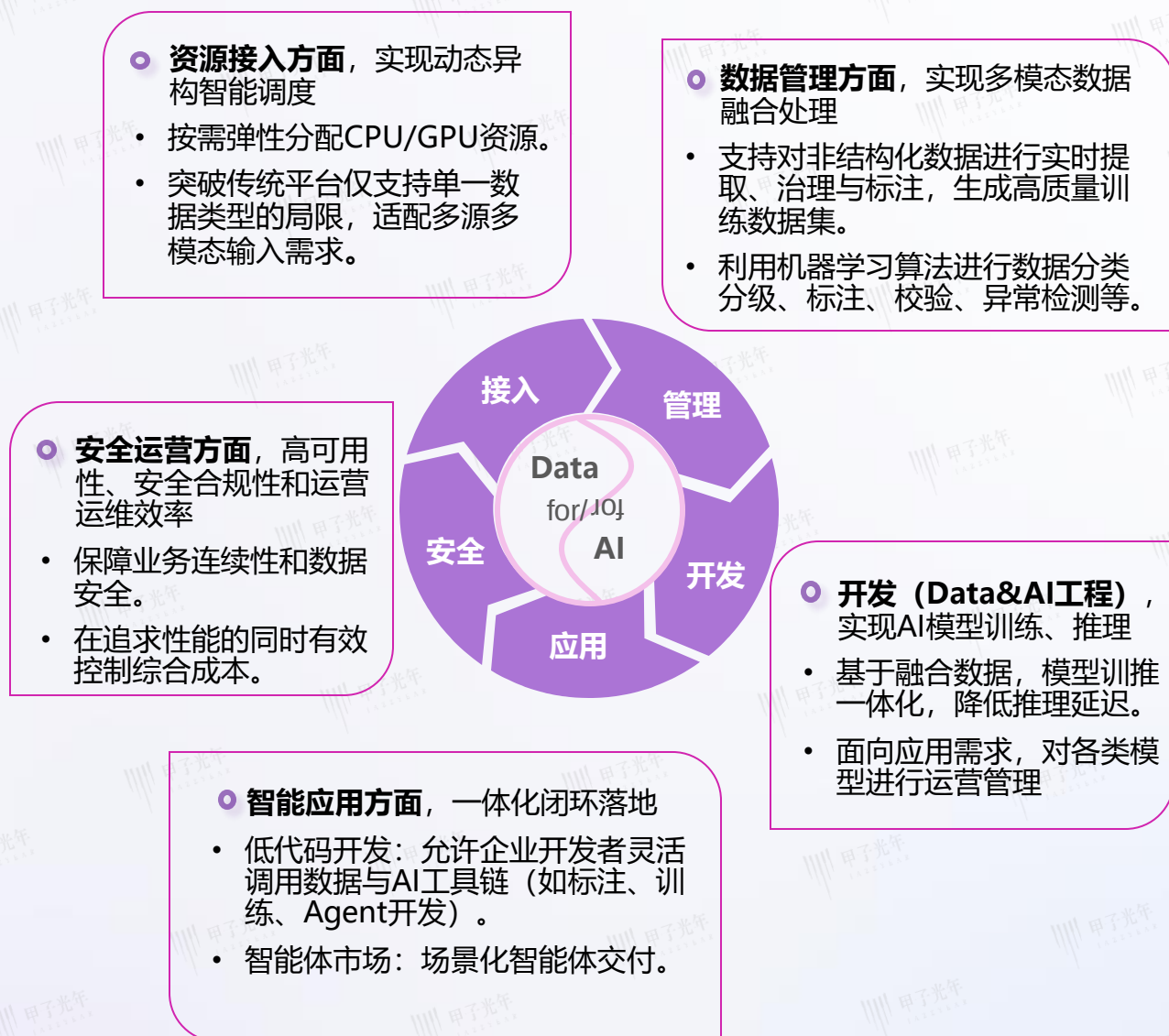
资料来源：甲子光年智库整理

- Data&AI数据基础设施是为支撑人工智能规模化落地而构建的一体化基础软件平台，核心目标是打通数据存储、治理、计算与AI模型开发的全链路，实现“Data for AI”和“AI for Data”的双向赋能，融合开发Data工程和AI工程。其本质是传统大数据平台的升级形态，通过AI原生架构重构数据处理范式，满足企业智能化转型的底层需求。

Data&AI数据基础设施的核心特征是AI原生的一体化架构驱动

- 构建“Data for AI”和“AI for Data”的动态持续双向就绪平台体系，以实现持续双向赋能是Data&AI一体化基础设施的核心目标。实现Data与AI的全链路无缝衔接。
- 云原生的存算分离是Data&AI一体化基础设施架构上的关键技术特征。多模态计算引擎和异构资源调度引擎是平台核心技术实现。
- 低代码/自然语言交互是Data&AI一体化基础设施交互体验上的突出特点。业务人员可通过自然语言生成SQL或代码，大幅降低数据使用和AI应用开发门槛。

Data&AI数据基础设施概念详解



Data&AI数据基础设施与大数据平台的核心差异在于数智一体和AI原生

- Data&AI数据基础设施处在数据应用的第三代，三代产品的差别主要体现在：
 - 第一代，以关系型数据库为代表，国际厂商如Oracle公司；数据形态以结构化为主，数据模型固定（ER图预定义表结构）；应用处理以批处理为主，GB级数据，无实时性要求；技术架构以烟囱式架构为主（单机数据库+应用服务器）。
 - 第二代，以混合存储服务为代表，国际厂商如Snowflake公司；数据形态半结构化/非结构化数据占比提升至50%+，采用混合存储(SQL+NoSQL，如HBase+Phoenix)；应用处理以分布式批处理为主，PB级数据，实时计算开始萌芽；技术架构采用分布式架构(Hadoop生态+微服务)，云原生架构(容器化+K8s)。
 - 第三代，以湖仓一体为代表，代表厂商如Databricks公司、科杰科技；数据形态非结构化数据占比超80%（音视频等多模态数据），采用多模态数据统一存储与处理，高维矩阵与张量存储，多模态复杂数据预处理，统一Unity Catalog；应用处理以批流一体为主，万亿参数模型，实时推理计算；技术架构采用数据湖仓一体，分布式弹性扩展架构。
- Data&AI数据基础设施在各类企业级数据处理应用场景下，可以原生贴合Agentic AI时代企业基于数据的智能化业务需求。

Data&AI数据基础设施与大数据平台差异

		传统大数据平台	Data&AI数据基础设施
资源接入	资源利用	存算一体	存算分离
	计算引擎	静态调度	动态调度
数据管理	数据类型	以结构化为主	原生支持多模态
	实时性	批量处理为主，延迟高	并行处理+实时动态处理
Data&AI工程化	AI集成	需外接ML平台，模型开发独立	Data&AI一体化
	工具链	缺乏AutoML/大模型运营支持	支持AutoML/训推一体/大模型运营/智能体
	应用开发	依赖代码开发，周期长	低代码生成场景化智能体
	治理模式	规则引擎+人工标注，成本高	AI驱动自动化，智能标注、异常检测
智能应用	数据血缘	局部追溯，跨系统难打通	端到端全局可追溯
	典型场景	BI报表、历史数据分析	实时决策、生成式AI、智能体应用
	响应速度	需求交付周期长，依赖人工开发	需求交付周期短，自动化、智能化程度高
	运维成本	多系统独立运维，复杂度高	全链路监控+自动弹性伸缩
平台架构	架构	中心化数据仓库，数据与AI割裂	分布式湖仓一体+向量数据库
	数据定位	静态资产，被动调用	全域动态生产资料，主动驱动模型优化

Data&AI数据基础设施的价值链条呈现“点-线-面-体”递进

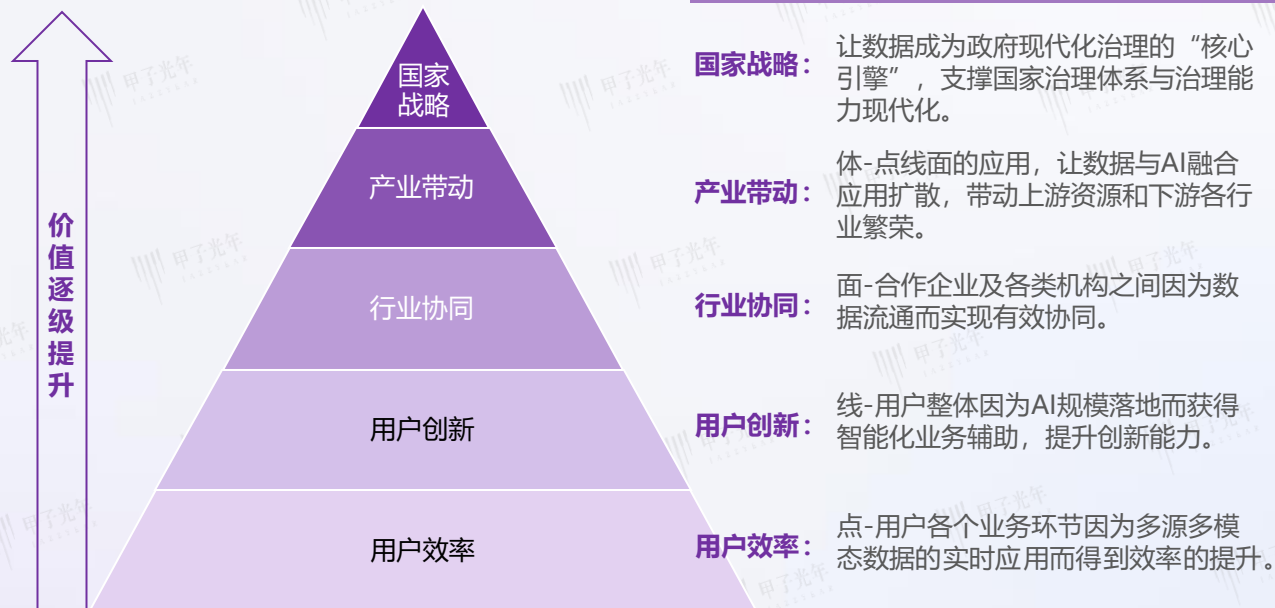
- Data&AI数据基础设施的核心价值就在于通过湖仓一体架构统一纳管全域数据资产，并以AI原生设计支撑模型的高效训练与实时推理，实现数据到智能的闭环转化，驱动业务持续创新，最终形成支撑未来智能社会的国家战略，成为智能革命的核心引擎。

甲子光年五级价值模型分解评估：

- 1.用户效率价值：**Data&AI数据基础设施以AI原生降低智能应用门槛，解决“用数难”，满足用户数字化升级，提升运行效率的需求。
- 2.用户创新价值：**AI规模化应用推动用户构建多模态实时处理能力，数据基础设施成为支撑模型训练、Agent开发的核心底座，满足用户AI落地需求，实现业务的智能化创新。
- 3.行业协同价值：**推动行业整体技术能力升级，提高AI数据协同能力。Data&AI数据基础设施作为底层支撑产品，覆盖各类用户和场景，实时响应业务场景变化；满足跨组织的数据流通需求和安全管控。
- 4.产业带动价值：**Data&AI数据基础设施促进数据与AI规模化应用，带动算力服务、算法开发、数据治理；激发新兴业态，如数据资产运营商、数据交易平台等新业态不断涌现。
- 5.国家战略价值：**以可信流通与基础设施赋能，加速数据要素市场化进程。融合AI、大数据、云计算、边缘计算等技术，形成新型基础设施和新型生产工具。构建自主可控的数据基础设施体系，保障关键领域安全。

甲子光年智库——五级价值评估模型

五级价值评估模型



资料来源：甲子光年智库整理

目录 >>>>

Part 03

新范式

Data&AI 数据基础设施的核心能力模型

3.1 能力模型：五大关键能力要素

3.2 架构范式：产品服务4+1架构体系

3.3 厂商类型：覆盖不同服务层次的四类厂商

3.4 类型能力：典型厂商类型与能力侧重各不同

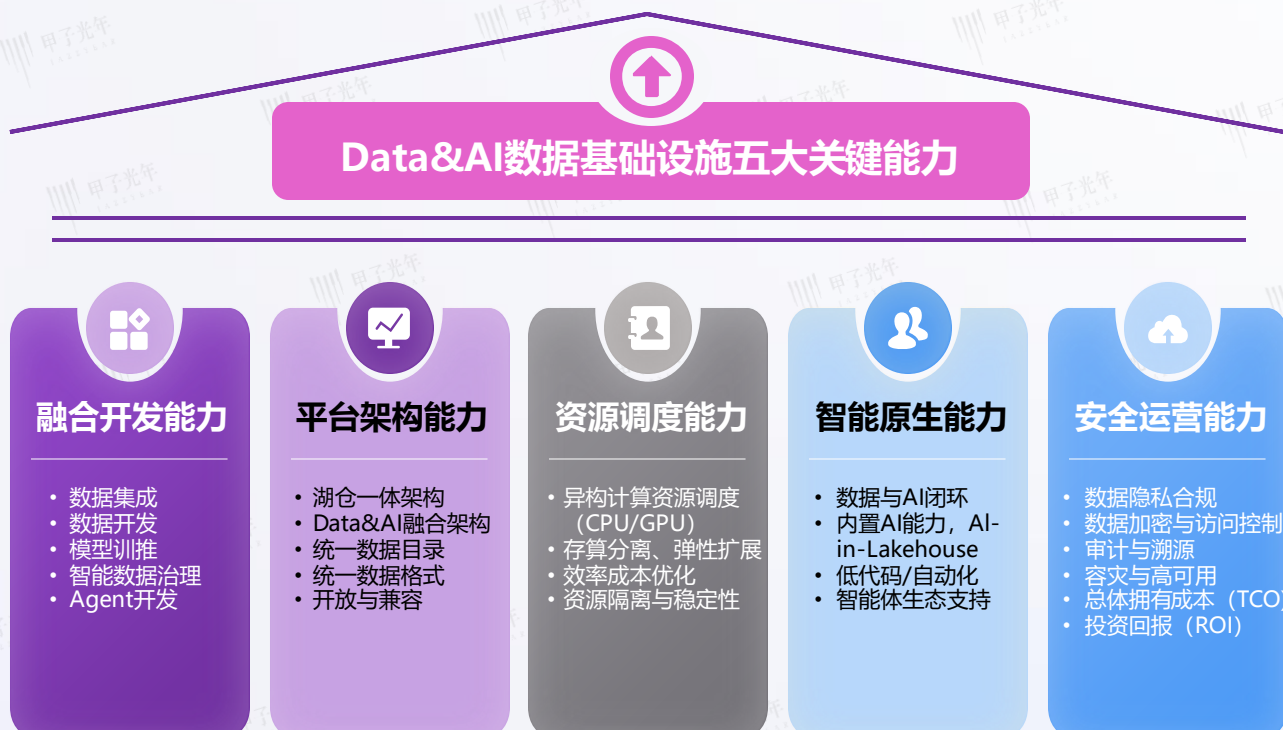
3.5 行业图谱：两类厂商主导行业趋势

Data&AI数据基础设施的五大关键能力要素

Data&AI数据基础设施让数据平台和AI智能工具融合成为统一系统性底座，其关键能力主要体现于以下五个方面：

- **融合开发能力**：指系统对多源、多模态数据进行集成、处理、智能治理、模型训推、Agent开发，以及系统化建设Data&AI的融合工程化能力，这是AI应用的基石。
- **平台架构能力**：指平台是否采用湖仓一体、Data&AI一体化新型架构，例如平台是否支持多模态数据Unity Catalog、Unity Format、Unity Processing、存算分离、Data Fabric等先进技术能力。
- **资源调度能力**：指系统对底层计算资源（包括CPU和GPU）的调度效率和成本控制能力。这直接影响AI模型开发和应用的效率和成本。
- **智能原生能力**：指系统原生支持AI开发与应用的深度和广度，以及能否实现数据与AI的闭环互动。这是“Data&AI一体化”的核心体现。
- **安全运营能力**：指系统在企业级部署中所需的高可用性、安全合规性和运营运维效率。它确保了数据的安全性、可用性、可靠性，是AI在企业侧落地的保障。

Data&AI数据基础设施能力模型



Data&AI数据基础设施为4+1架构，下接硬件和数据源，上承业务

- 资源接入层：此层是“入口”，负责从各种来源摄取数据，并进行初步的流式或批量处理，为上层提供可用的数据资源。
- 数据管理层：此层是“核心仓库”，对接入的原始数据进行存储、治理、加工和组织，形成干净、可信、易用的数据资产。
- Data&AI工程化层：此层是“智能引擎”，直接利用治理好的高质量数据集，进行机器学习模型、大模型的训练、推理与运维。
- 智能应用层：此层是“价值出口”，通过低代码、智能体等方式，将数据和AI能力封装成最终用户可直接使用的应用或工具。
- 安全运营层：此层是“安全底线”，保障数据与AI资产的全生命周期安全，满足合规要求；确保平台稳定、高效、经济地运行，持续实现业务价值。

Data&AI数据基础设施架构示意



资料来源：甲子光年智库整理

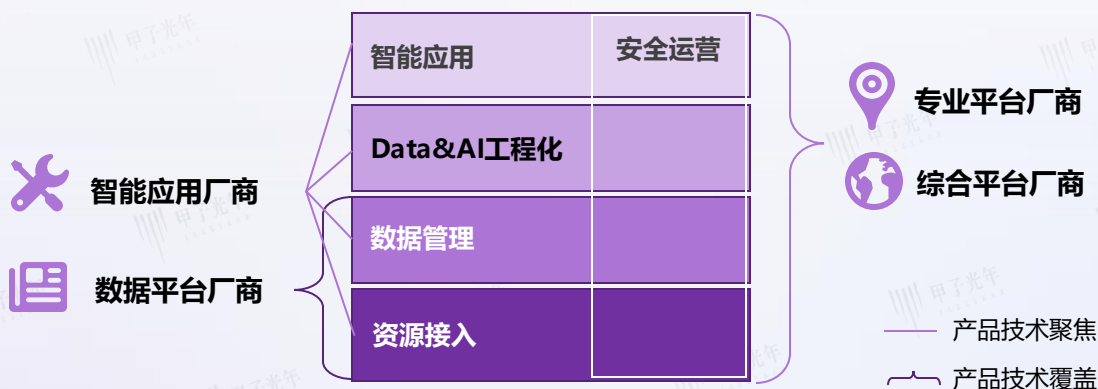
数据基础设施厂商主要有四类，各自专注不同的产品服务层次

- 从全球和中国市场整体智能化需求来看，企业需要的是一体化、低代码、AI与数据原生融合的一体化基础软件平台工具，允许企业开发者以更加灵活的方式来进行，对存储在数据库、数据湖中的多模态数据进行实时提取、治理、标注，以及高质量数据集的管理，进行面向业务场景的模型训练和智能体的开发，以服务AI应用的全面落地。
- 从整个数据基础设施行业来看，AI原生的Data&AI技术融合基础设施软件产品处于早期阶段，更多企业仍在围绕企业原有赛道和自有产品功能来提升效率。

提供Data&AI数据基础设施产品及服务的厂商主要有四类

- 智能应用厂商**，以围绕数据或AI场景做定制化服务、工具的厂商，如爱数、数新智能、数势、Dify等，这些厂商没有体系化的数据基础设施技术产品和服务经验，主要业务是服务客户数据或AI场景的快速落地，长期会成为Data&AI数据基础体系中聚集某个细分领域的专业类厂商。
- 数据平台厂商**，以传统大数据平台厂商，如Snowflake、星环、明略、东方金信为代表。这些厂商有较多的大数据平台项目和客户积累，目前处于融合AI能力的转型期，尚未形成Data&AI一体化产品，缺少基础设施一体化建设的实施经验和案例，在Data&AI数据基础体系中处于探索期。
- 专业平台厂商**，值得一提的是以Databricks、科杰科技为代表的专业平台厂商基于长期的在这个方向上头部客户的落地实践和技术积累，率先实现了面向AI原生的Data&AI一体化基础设施软件成熟产品，加之多个行业的标杆案例和长期实践形成了一套成熟的方法论，在AI时代Data&AI一体化基础设施赛道取得先发优势，随着在客户侧的落地这个优势会逐步扩大。
- 综合平台厂商**，以阿里、华为、腾讯为代表的综合型平台厂商，有较好的品牌力，基于其丰富的产品线整合形成Data&AI基础设施产品体系化服务能力。但由于产品各模组分属不同团队，在内部支撑不同业务，因此对外的架构上仍然割裂。落地服务往往还依赖三方，整体性上产生进一步偏差，且常对客户选型有垄断性要求。

各类数据基础设施厂商服务聚焦范围示意



资料来源：甲子光年智库整理

四类厂商能力各有侧重，专业平台厂商Data&AI一体化能力突出

四类厂商的技术路线、市场侧重各有不同，因此在能力表现上也各有侧重：

- 综合平台厂商能力最为均衡，其中资源调度能力最强。集团旗下各产品线整合为综合平台产品在客户侧深度场景应用服务上能力略显不足。
- 专业平台厂商在AI原生架构、融合开发（Data工程和AI工程融合开发）能力上最突出。其能力最能体现Data&AI数据基础设施的发展重点和趋势。
- 数据平台厂商在融合开发和资源调度上相对较强。其能力体现出了客户对大数据平台服务的核心需求。
- 智能应用厂商以AI Agent场景应用服务厂商为主，智能原生能力表现相对强。该类厂商主要解决客户场景侧的Agent开发、AI数据来源和数据质量的需求，因此能力建设就更聚焦于应用端。



资料来源：甲子光年智库整理

Data&AI数据基础设施的行业图谱

- Data&AI数据基础设施行业厂商主要集中于资源接入层和数据管理层；Data&AI数据基础设施的核心也在于前三层。
- 以阿里云为代表的综合平台厂商业务涉及全行业每个层次，以Databricks、科杰科技为代表的专业平台厂商在Data&AI数据基础设施的核心域布局最深；这两类厂商主导了Data&AI数据基础设施行业发展大势。

Data&AI数据基础设施产品及代表厂商



资料来源：甲子光年智库整理

更多免费行业报告

并购家

www.ipoiipo.cn

目录 >>>>

Part 04 新应用 Data&AI数据基础设施典型应用

- 4.1 应用场景：Data&AI数据基础设施推动大型组织智能化转型
- 4.2 大型国央企：破解数据孤岛与业务智能化诉求
- 4.3 大型金融企业：释放多模态数据价值，响应实时决策需求
- 4.4 跨国外企：契合本地化与国际化的统一
- 4.5 城市政府：打通数字化产业生态的关键堵点

Data&AI数据基础设施推动大型组织智能化转型

核心痛点

- 案例将展示Data&AI数据基础设施如何支撑不同组织解决各自的痛点。

- 数据孤岛
- 安全合规

大型国央企

破解数据孤岛与安全合规诉求

- 大型国央企，拥有并持续产出海量、高价值的数据，但同时也面临着最严格的数据安全合规（如等保、国资监管、司库）要求。
- 数据孤岛问题突出。大型国央企发展历史长，老信息系统多，跨多时期系统、跨部门、跨子公司、跨地域的数据整合需求和困难尤为突出。

大型金融企业

响应快速变化，不断提升投入产出比

- 监管要求高
- 实时响应

- 监管要求高，数据系统厂商复杂，维护和切换升级复杂。
- 数据质量要求高，业务部门用数流程复杂成本高。
- 场景细分，最适合智能化应用落地。
- 对数据的实时性有极高要求；跨区域业务协同需求高，用户规模大。

跨国外企

契合本地化与国际化的统一

- 异构协同
- 跨国治理

- 跨国外企业务遍布全球，IT架构复杂，常采用多云或混合云策略，整合多种异构数据源和AI工具链，实现全球数据的流动与协同分析。技术整合、生态互通困难。
- 跨国外企还面临特殊的数据治理之困。在国际化背景下，如何实现全球统一治理框架下的本地化合规，既需满足全球统一标准，又要符合各国数据主权法规（如GDPR）。

城市政府

打通数字化产业生态的关键堵点

- 数据流通
- 数据成本

- 城市政府是既是城市数据的最大拥有者和使用者，也是数据应用生态的关键建设者之一。但行业标准缺失，生态难以互通，很多中小微企业“望智兴叹”在城市层面集中体现。提供公共数据开放服务、降低企业（尤其是中小企业）使用数据和AI技术的门槛，促进产业生态繁荣是城市政府推动AI时代产业和社会发展的关键任务。

Data&AI数据基础设施助力大型国央企探索“数据+平台+应用”的数智化变革新模式，推动人工智能与实体经济深度融合

核心痛点：三大挑战，数据治理难，系统整合难，数据统计难

数据孤岛与集成挑战

- 数据资源体系不完善
- 数据治理机制缺失
- 数据共享模式缺乏

技术架构与整合挑战

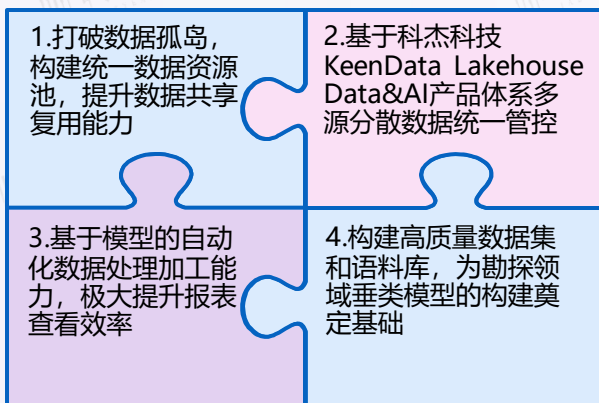
- 历史遗留系统整合
- 历史数据仓库性能低
- 新架构建设扩展难，兼融难

数据统计难，时效性差

- 集团整体经营分析报表人工汇总统计数据错误频出
- 从提出报表需求到形成可查看的报表周期长

解决问题：AI赋能数据管控

数据资产化运营保障机制及支撑体系



落地效果：支撑勘探业务数字化、智能化

打造高质量数据资源池，赋能业务应用

- 围绕勘探9大核心业务，构建勘探开发数据资源目录，支撑集团61套业务系统用数服务需求。

支撑国内上游数据资源管理模式变革

- 构建了新型数据资源管理体系，实现了“数据管理制度化、管理组织齐全化、治理规则标准化、数据架构一体化”。

搭建国内上游Data&AI平台，筑牢数据支撑底座

- 实现了面向结构化数据、图形文档、实时数据、音视频、体数据等海量数据管理应用，支撑勘探开发构建“便捷接入、快速流通、实时处理、统一汇聚”能力。



中国石油化工集团有限公司是成立于1983年的特大型国有独资企业。是世界第一大炼油公司、第二大化工公司，其加油站总数位居全球第二，连续多年位列《财富》世界500强前列。集团为响应国家数字经济发展政策号召，加快国有企业数字化转型，推动人工智能与实体经济深度融合。

资料来源：甲子光年智库整理

Data&AI数据基础设施支撑金融机构将数据转化为“支撑亿级客户服务、保障万亿资产安全”的核心资产

核心痛点：决策实时性不足，制约规模创新

战略决策层

- 数据标准不统一
- 数字化能力难以规模化推广
- AI驱动的应用预测准确率低
- “数据驱动战略”在零售、对公领域难落地

技术支撑层

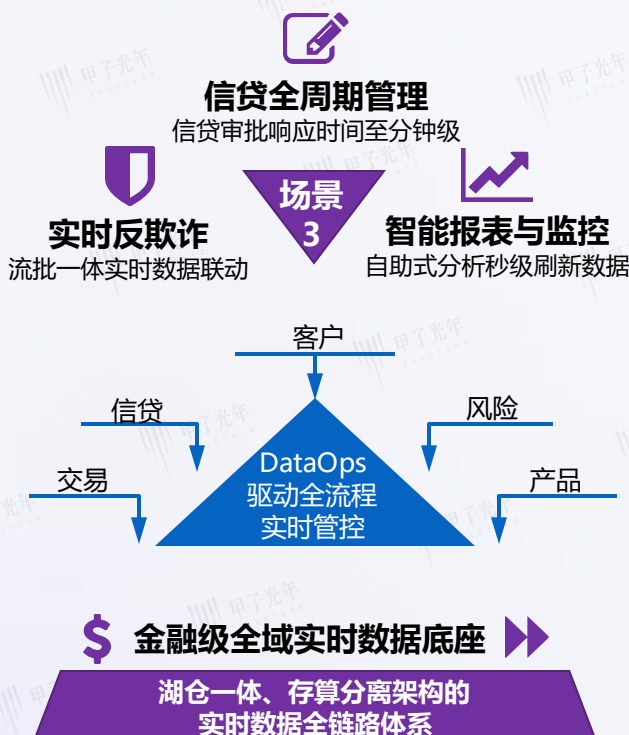
- 缺乏统一的实时处理架构
- 缺乏全域数据标准与治理机制
- 高可信的实时经营与风险数据不足
- 数据开发依赖专业技术人员

业务应用层

- 核心场景（如信贷审批、风险交易拦截）响应延迟高
- 数据孤岛问题突出
- 会员精准服务、实时资产监控等AI驱动场景难以落地。

解决方案：构建场景融合实时数据体系

解决问题：提升实时性和智能化水平



打通数据壁垒

构建适配规模需求的实时数据体系

提速核心场景响应

匹配金融业务“低延迟、高可信”诉求

降低智能应用门槛

推动企业内数智化能力的规模化应用

落地效果：全域实时智能支撑服务转型

数据底座能力：

实现从“分散处理”到“全域实时”的质变

业务价值释放：

实现效率与风控的双向突破

转型支撑力：

基于科杰科技Data&AI产品体系，构建“数据-服务-应用”的数字化生态

资料来源：甲子光年智库整理

大型跨国零售集团通过构建统一、实时、智能的Data&AI数据基础设施，成功将数据转化为核心生产资料，支撑AI应用场景的规模化落地

关键驱动：“Digital AEON” 战略



数字化能力

- To C (顾客)
- ➔ • To B (内部经营)
- To P (合作伙伴)

AI应用场景



智能定价



个性化推荐



会员运营

.....



亚洲领先的综合零售及服务企业集团，布局日本、中国、东南亚三大总部。全球拥有约300家企业、19094家店铺（中国近500家）。

核心痛点：数据孤岛与智能不足制约创新

战略决策层

- 缺全局数据支撑
- 标准不统一
- 数字化能力难规模化复制
- 高质量、全维度数据不足

技术支撑层

- 算力瓶颈
- 架构缺陷
- 实时性不足
- 数据管理低效

业务应用层

- 报表响应速度慢
- 业务创新缺乏数据支撑
- 会员运营、智能营销等AI场景难以落地。

解决方案：构建一体化Data&AI基础设施

数据体系

- 采购、门店、供应商等十大主题域
- 近5年300+表、100亿+数据
- 128个核心指标与19类标签
- 数据质量管理实现数据标准化稽核

技术架构

- 分布式、高可用、可扩展存储架构
- “数据源-采集-处理-存储-分析-运维”全流程
- “离线+实时+AI”三位一体

场景赋能

- 客户360°画像
- 库存自动化预警
- 智能化商品推荐
- 智能定价、CDP会员运营等

落地效果：数据智能驱动业务价值释放

- 基础设施提升：• 基于科杰科技KeenData Lakehouse数据智能平台，数据全生命周期管理，支撑集团19,000+店铺及多业态业务需求。
- 业务效率提升：• 核心报表响应速度提升10倍，业务决策周期缩短50%。
 - 智能定价，KVI商品销量提升9%；
- AI场景落地：• CDP会员运营，会员复购率提升8.45%；
 - 库存自动化预警，缺货率下降1.2%

资料来源：甲子光年智库整理

Data&AI数据基础设施支撑建设城市数据流通平台，推动数据智能、高效、安全流通，释放数据要素价值，带动城市数字经济高质量发展



关键驱动：国家战略+国家任务

解决方案：AI原生设计，与数据双向赋能



落地效果：构建城市智能的数据底座

构建新范式

- 践行了“供得出、流得动、用得好、保安全”为核心的新业务范式

注入新动能

- 推动了数据从“资源”向“生产要素”的跃迁

夯实新底座

- 完善技术生态，为某市各行业提供了支撑数据要素市场化配置的关键底座



特别鸣谢

感谢对本报告撰写提供支持的众多科技媒体、智库机构、大数据厂商、大型组织和专家学者，感谢为报告的顺利完成贡献了大量数据、观点、案例等一手资料，在此特别鸣谢！

法律声明

版权声明

本报告由甲子光年智库制作完成，报告内容的版权及相关知识产权均归北京甲子光年科技服务有限公司所有。任何单位或个人在引用本报告内容时，须保持内容的原始性，不得进行歪曲、删改或误导性引用，并须注明报告出处为“甲子光年智库”。否则，由此引发的一切后果由引用方自行承担，甲子光年保留追责权利。

免责条款

本报告中的行业数据、市场预测和相关分析主要来源于甲子光年研究团队通过桌面研究、专家访谈、问卷调查、公开数据整理及甲子光年产品数据等方式获得，部分数据通过甲子光年自主统计预测模型进行估算。我们已尽合理努力确保数据的准确性、完整性与可靠性，但甲子光年不对其作出任何明示或暗示的保证。在任何情况下，本报告中包含的内容、数据或观点均不构成对任何单位或个人的投资建议、法律建议或其他形式的专业意见，相关决策应由读者自行判断并承担风险。由于调研方法和样本范围的限制，报告中发布的数据结果仅反映特定时间段、特定对象的调研情况，具有一定的局限性，仅供读者作参考用途。甲子光年对因使用本报告内容而产生的任何直接或间接损失不承担任何法律责任。

THANKS

谢谢

北京甲子光年科技服务有限公司是一家科技智库，包含智库、媒体、社群、企业服务版块，立足于中国科技创新前沿阵地，动态跟踪头部科技企业发展和传统产业技术升级案例，致力于推动人工智能、大数据、物联网、云计算、AR/VR交互技术、信息安全、金融科技、大健康等科技创新在产业之中的应用与落地



扫码关注甲子光年公众号

智库院长

宋涛微信 (stgg_6406)

商务合作

郑爽： 18600502376 (微信同号)