

## 2023年中国数据分析和人工智能技术成熟度曲线

Published 17 August 2023 - ID G00799755 - 15 min read

By Analyst(s): Julian Sun, Ben Yan, Xingyu Gu, Fay Fei, Mike Fang, Tong Zhang

Initiatives: [Digital Technology Leadership for CIOs in China](#)

数据和人工智能对于中国的数字经济和国家安全至关重要，在监管框架支持下具有独特地位和增值效应。中国的数据和分析领导者必须了解该地区数据分析以及人工智能生态系统的现状和前景，才能推动业务成果的实现。

### 战略规划假设

到2026年，中国超过30%的白领岗位将被重新定义，使用和管理生成式AI的技能将大受欢迎。

## 分析

### 企业需要了解什么

中国的数据分析（D&A）和人工智能（AI）与全球市场有很多相似之处，但在组织结构、技术重点和价值主张方面也具有独特的差异性。在数字经济时代，数据分析和人工智能是每个企业机构推动成果优先型投资战略的基础。如今，D&A行业的最佳实践需要符合中国的数据、隐私和人工智能相关法规——这比美国或欧盟的规定还要严格。

尽管这一领域热度高涨，但能达到技术成熟期的D&A和AI创新数量很少。本技术成熟度曲线关注的热点主要围绕商品化程度各异的新兴D&A和AI技术，并探讨如何应用这些技术创建超越带有D&A和AI功能的普通系统，以及这些创新对企业内外部人员和流程的影响。

中国的D&A领导者须参考本文，了解和利用当前具有显著影响的技术，并为未来发展做好战略准备。

除了本篇报告，D&A领导者还应参考如下技术成熟度曲线报告，了解周边领域发展趋势：

- [Hype Cycle for ICT in China, 2023](#)
- [Hype Cycle for Smart City and Sustainability in China, 2023](#)
- [Hype Cycle for Security in China, 2022](#)

这四篇技术成熟度曲线报告的分析内容各有侧重，可以帮助中国技术CIO全面了解数据和分析生态系统。

### 技术成熟度曲线

数据分析和人工智能一直是中国CIO明确的首要投资重点，也是中国国家战略的重点关注领域。与此同时，中国出台的有关数据、隐私和人工智能的法律要求也引发了各界技术高管的担忧。

本技术成熟度曲线中，即将进入期望膨胀期的技术数量最多。创新往往被吹捧为传统瓶颈问题的解决方案，有望解决中国CIO共同担忧的问题——如硬件资源短缺、可扩展性、可持续运营、安全风险缓解、技术自主可控和AI模型的多域适用性问题，从而带来清晰的业务价值。然而，终端用户更重视有形的影响，而不是抽象的战略概念。

在所有热点技术中：

- 数据中台——作为曾经备受中国企业机构推崇的实践，已落入泡沫破裂低谷期。许多企业机构和供应商都不愿意在企业内部采用这一概念，或者干脆将其从宣传中移除。

- 数据资产管理是一种变革性的、与众不同的实践，为中国企业机构的数字经济提供了支撑，使其能够有效地阐明和交付D&A项目的价值。
- 中国市场对生成式AI和大模型（LLM）青睐有加。然而，终端用户对使用以模型为中心的方法自建本土LLM的信心相对较低，同时倾向于采用以数据为中心或以应用为中心的方法使用生成式AI。这主要是因为中国尚不具备快速生产国产AI芯片的能力，而且与NVIDIA等全球供应商相比，中国企业仍处于AI芯片设计的早期阶段。

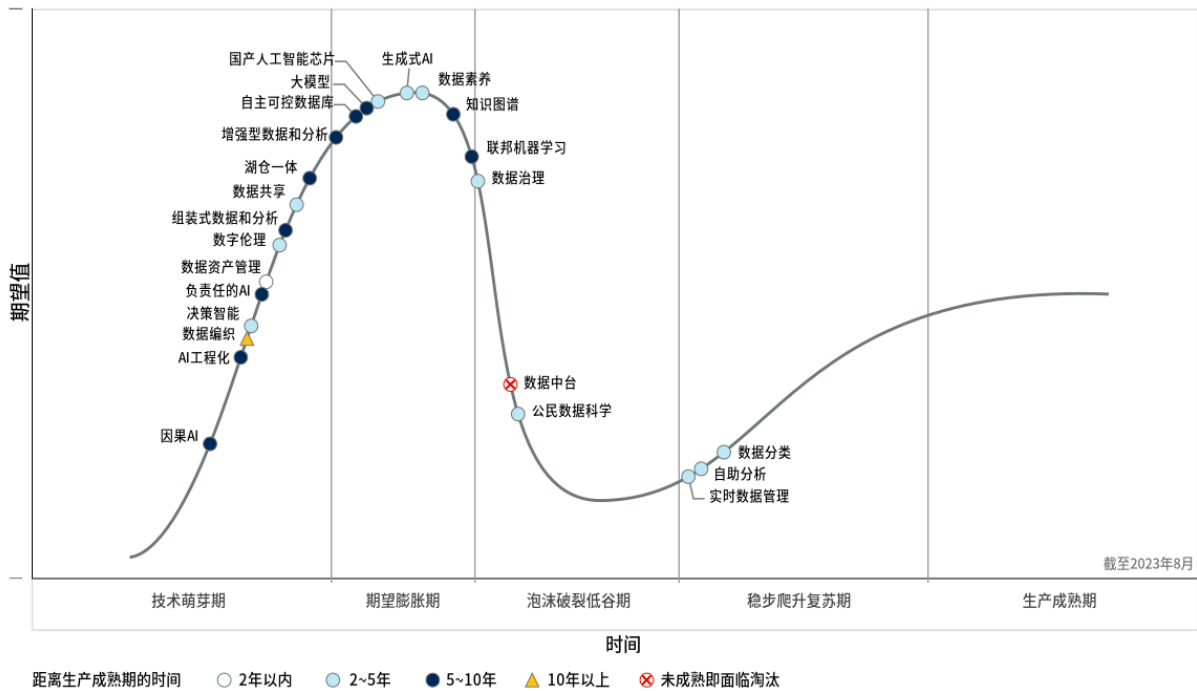
本技术成熟度曲线中的创新技术体现了中国的数据与分析和AI生态系统向融合化、可组装式发展的趋势，主要涉及以下四个领域：

- 分析和数据科学，包括增强型数据和分析、公民数据科学、组装式D&A、数据中台、联邦机器学习、知识图谱和自助分析。
- 人工智能（AI），包括AI工程化、因果AI、决策智能、大模型、生成式AI、国产AI芯片和负责任的AI。
- 数据管理，包括数据分类、数据编织、自主可控数据库、湖仓一体和实时数据管理。
- 数据和分析项目和实践，包括数据资产管理、数据治理、数据素养、数据共享和数字伦理。

今年是Gartner中国数据分析和人工智能技术成熟度曲线的首次发布，预计很快将出现大量创新和失败的案例，这有可能产生巨大影响。其中，中国的关键创新包括数据中台、国产AI芯片、自主可控数据库和数据资产管理。

图1：2023年中国数据分析和人工智能技术成熟度曲线

2023年中国数据分析和人工智能技术成熟度曲线



Gartner

## 优先级矩阵

数据资产管理是本文中唯一一项影响力评级为颠覆，而且能够在不到两年内实现主流采用的创新。中国政府的数字经济目标和相应政策——包括设立国家数据局，为这一战略提供了支持。数据已被认定为第五大生产要素，这将对中国的D&A领导者、CIO和其他技术高管产生影响。<sup>1</sup>自此，中国企业将具有独特的优势，因为世界上还没有其他国家为业务成果驱动因素而设立政府监管。

在未来两到五年内，值得特别关注、将进入主流采用阶段的创新技术包括公民数据科学、决策智能、生成式AI和实时数据管理。较早地采用这些创新技术将带来巨大的竞争优势，缓解业务技术人员匮乏所带来的问题，解决为实现成果而获取业务背景和业务价值的挑战。

部分创新技术需要5至10年时间才能进入主流采用阶段，但其应用可以从小规模项目开始，以在短期内产生效果。这些创新技术包括组装式D&A、数据共享、数据编织、基础模型和负责任的AI。这些创新的实施对于破除数据中台的神话、建立运营性AI系统、实现从创新到生产的转变至关重要。根据Gartner 2022年中国AI用例调研，<sup>2</sup>人才或服务提供商短缺是阻碍AI采用的最大障碍。因此，D&A领导者应加速实施数据分类、提高数据素养，并将其作为长期战略，为实现AI目标奠定坚实的基础。

Table 1: 中国数据分析和人工智能技术优先级矩阵

影响力	距离主流采用的时间			
↓	2年以内 ↓	2~5年 ↓	5~10年 ↓	10年以上 ↓
颠覆	数据资产管理	公民数据科学 决策智能 实时数据管理 数据共享 生成式AI	大模型 组装式数据和分析 负责任的AI	数据编织
较高		国产人工智能 芯片 数字伦理 数据分类 数据治理 数据素养	AI工程化 因果AI 知识图谱 联邦机器学习 自主可控数据库	
中等		自助分析	增强型数据和分析 湖仓一体	
较低				

来源：Gartner (2023年7月)

## 萌芽期技术

### 因果AI

分析师: Ben Yan, Julian Sun

影响力评级: 较高

市场渗透率: 目标受众覆盖率低于1%

成熟度: 发展阶段

#### 定义:

因果人工智能（AI）可识别和利用因果关系，超越基于相关性的预测模型，逐步打造能更有效地指导行动和更自主地采取行动的AI系统。因果AI会采用不同的技术，如因果图和仿真模拟，以便发掘因果关系并改善决策。

#### 为何重要

AI的终极价值在于帮助人类采取更有效的行动。机器学习技术基于统计关系（相关性）进行预测，但并不探究这些关系是否具有因果联系。依靠统计相关性虽然可以预测某一结果，但并不能获知结果背后的原因以及改进结果的方法。在决定应采取何种最佳行动来影响特定成果时，因果AI至关重要。因果AI技术提高了AI的自主性、可解释性、鲁棒性和效率。

#### 业务影响

因果AI具有以下作用：

- 估计干预效果，提高决策增强效果和AI系统自主性
- 添加领域知识，在较小的数据集中启动因果AI模型，提高效率
- 捕捉易于解释的因果关系，提高可解释性
- 利用在变化环境中仍有效的因果关系，提高鲁棒性和适应性
- 能够通过更高效的实验来提取因果信息，节省时间和成本

#### 推动因素

- 预测性分析（可能发生什么）逐渐转向规范性分析（应该做什么）。虽然做出准确预测仍有关键意义，但一个越发重要的需求是理解因果关系、了解如何影响预测结果。

- AI系统需要通过自主行动来创造业务价值，尤其是因时效性强、复杂度高而无法进行人工干预的使用场景。实现这一点，需要AI理解各项行动的效果以及如何实施有效干预。
- 某些使用场景的数据可用性有限，促使企业机构转向因果AI等更有效的数据技术。因果AI可利用人类的因果关系领域知识，在仅有少量数据的情况下构建AI模型。
- 运用AI的用例与环境日益复杂，需要采用更具鲁棒性的AI技术。与统计相关性相比，因果关系的改变速度慢得多，所以因果AI在快速变化的环境中表现更强大、更灵活。两类AI的训练环境不同，而过去几年的变动显示，各行业基于相关性构建的AI模型表现较为脆弱、适应性不强。
- 人们希望AI模型能提高可信度和可解释性，因而更直观的AI模型引起了更大关注。因果图等因果AI技术能帮助用户更准确地发掘原因，并且可用易于人类理解的语言对模型进行解释。

#### 阻碍因素

- 因果关系不易把握，并非每种现象都能从因果联系的角度简单完成建模。无论是否使用AI，因果关系可能都是未知的。
- 因果AI模型的质量取决于其所依赖的因果关系假设和所用数据，而数据可能存在偏见和失衡问题。使用因果关系的模型，不一定优于使用相关性的模型。
- 只有具备技术和领域专业知识，才能建立因果AI，进而正确地估计因果效应。这往往比构建基于相关性的预测模型更有难度，要求领域专家和AI专家积极地展开合作。
- AI专家可能并不了解因果分析方法，过于依赖机器学习等数据驱动模型，这可能成为部署因果AI的阻力。
- 厂商开发尚处于起步阶段，而且当前企业使用率较低，对用户启动因果AI初期试点和确定最相关用例形成了挑战。

#### 使用建议

- 了解当前流行的AI和ML预测方法的局限性。这些方法主要利用关联关系进行预测，但往往会忽略因果关系。
- 利用因果AI增强决策智能，提升其自动化水平，确保不但能生成预测，而且能了解如何改变预测结果。运用实例包括客户留存计划、营销活动分配和财务组合优化。
- 根据每个用例的复杂度选择不同的因果AI技术，包括因果规则、因果图和贝叶斯网络、仿真模拟，以及利用机器学习进行因果学习。

- 研究并判断因果AI相比于传统AI具有独特优势的场景，优先实施可发挥最大影响力的用例。
- 向数据科学团队普及因果AI的相关知识，包括与基于相关性的AI技术之间的区别，以及可引入因果关系的各项技术。

## 厂商示例

Causality Link、九章云极、华为、IBM、零犀科技、微软

## Gartner相关推荐阅读

[Innovation Insight for Causal AI](#)

[Case Study: Causal AI to Maximize the Efficiency of Business Investments \(HDFC Bank\)](#)

[Innovation Insight for Composite AI](#)

## AI工程化

分析师: Mike Fang, Tong Zhang

影响力评级: 较高

市场渗透率: 目标受众覆盖率为1%~5%

成熟度: 孵化阶段

## 定义:

人工智能工程化（AI工程化）是企业规模化交付AI解决方案的基础。该学科结合了DataOps、MLOps和DevOps管道，创建以AI为基础的企业开发、交付（混合环境、多云、边缘）和运营（流媒体、批量）的一致系统。

## 为何重要

完全工业化的AI的价值在于，能够在企业内的不同环境中快速开发、部署、调整和维护各种AI模型（统计模型、机器学习模型、深度学习模型、图模型、语言模型和基于规则的模型）。为了扩展AI项目的规模，企业机构必须采用AI工程化实践来部署和架设管道。鉴于AI工程化的复杂性，以及缩短上市时间这一需求，这些实践至关重要。



## 业务影响

很多企业机构尽管已经进行了数百次AI试点，但还是难以将AI模型从试点阶段推进至生产阶段。由于流程存在重大瓶颈，多数企业机构仅将少量AI模型投入生产。AI工程化为企业机构提供了一个解决瓶颈的框架。通过AI工程化方法，即DataOps、ModelOps和DevOps，能够基于结构化、可重复的工厂模型框架，轻而易举地将模型部署到生产环境中，实现重要价值。

## 推动因素

- 推广“中台”概念。“中台”是在中国使用的一个独特术语，指用于构建可复用能力的组织战略。这一战略的目标是实现纵向扩展和业务敏捷性。“AI中台”从众多解决方案中脱颖而出，用于支持AI工程化的采用。
- 淘汰传统孤岛式数据管理方法。
- 减少数据摄取、处理、模型工程和部署中的阻抗失谐，防止AI模型投入生产后发生偏移。
- 通过数据、模型和部署管道的标准化、治理和自动化，加速AI的产品化。
- 采用DataOps、ModelOps、DevOps和其他最佳实践，开发和部署企业的AI平台。
- 在企业的各类场景中实现可发现、可组装和可复用的AI工件（数据目录、特征存储、模型存储）。
- 推动跨混合环境、多云、边缘AI或物联网的编排，以扩展AI项目。
- 整合数据工程、数据科学、应用开发、安全和平台基础设施团队，实现AI架构的运营化。
- 拓展利用现有数据、分析和治理框架的基础性平台的使用范围，以扩展AI项目的生产规模。

## 阻碍因素

- 中国企业往往寻求“独角兽”企业的专家来推动AI平台的产品化。与全球供应商相比，中国鲜少有供应商能够提供强大的AI工程化能力。企业通常需要通过开源架构来构建和支持这些环境。虽然这一方法可以快速见效，但从长期来看可能导致扩展性和维护性问题。
- AI工程化需要同步开发跨领域的管道。
- AI工程化要求将功能全面的解决方案与特定工具（包括开源技术）集成，以尽可能少的功能重叠来填补企业架构的缺口，包括提取、转换和加载（ETL）存储、特征存储、模型存储、模型监控、管道可观测性和治理。
- AI工程化需要的云成熟度较高，还可能需重新架构，或者跨部署环境集成数据和AI模型管道的能力。分析和AI工作负载可能较为复杂，管理难度较高，同时成本较高，这些都可能导致处于AI项目初始阶段的企业机构望而却步。

## 使用建议

- 采用“AI中台”平台，从AI项目中获得最大业务价值。遵循AI工程化的实践，简化数据、模型和实施管道。
- 明确端到端AI平台运营所需的关键能力，简化数据和分析管道。
- 避免构建零散的AI工程化管道来集成零散的功能。尽量不使用单点解决方案，而是仅填补全功能DataOps、MLOps、ModelOps和PlatformOps工具中的功能/能力缺口。
- 发展AI模型管理和治理实践，使模型性能、人类行为和业务价值交付保持一致。将利益相关者的信任度和价值实现速度作为模型设计的主要考虑因素，使业务用户更容易采用和执行AI模型。
- 利用云服务提供商环境构建AI工程化。同时，在迁移到云端时，合理调整数据、分析和AI组合。
- 提升数据工程 and 平台工程团队的技能，采用可推动AI工件持续集成/持续交付（CI/CD）的工具和流程。

## 厂商示例

第四范式、亚马逊云科技、百度、IBM、微软

## Gartner相关推荐阅读

[Market Guide for AI Software, China](#)

## Top Strategic Technology Trends for 2022: AI Engineering

### Chinese AI Survey Analysis: AI Trends Wave 3.0 — From Operational to Strategic

#### Go Beyond Machine Learning and Leverage Other AI Approaches

##### 决策智能

分析师: Ben Yan, Mike Fang

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为5%~20%

成熟度: 发展阶段

##### 定义:

决策智能 (DI) 是一门实用学科, 通过明确理解和设计如何制定决策, 以及如何通过反馈来评估、管理和改进结果, 从而改进决策的制定。

##### 为何重要

在人工智能 (AI) 决策技术的助力下, 自动化决策和增强智能的热度正盛, 推动着决策智能向期望膨胀期发展。近期的危机暴露了业务流程的脆弱性。重建业务流程, 使其具备韧性、适应性和灵活性, 需要决策智能方法和技术的加持。围绕各类软件领域快速崛起的决策智能市场, 开始为决策者提供答案。

##### 业务影响

决策智能对企业机构会产生以下业务影响:

- 在企业机构内各层级实现数据驱动型决策。利用决策支持、增强和自动化, 优化运营/策略/战略决策成果。
- 提高决策的可见性, 降低不可预测性。实现企业决策模型可视化, 提高决策透明度和可审计性。明确业务环境中的不确定性因素。

## 推动因素

- 业务环境复杂且不断变化，业务发展节奏难以预测、不易确定。各类AI技术的结合以及围绕复合型AI、智能业务流程、决策管理和高级个性化平台的若干技术集群的融合，推动着围绕支持决策智能学科的决策系统平台创造出一个新市场。
- 需要减少孤立、脱节的非结构化临时决策。此类决策往往缺乏协调性，在实现局部优化时，会以牺牲整体效率为代价，无论从IT还是业务角度都是如此。
- 人类与机器之间的协作不断扩大，AI等技术接管的任务日益增多，引发了人类的不安情绪，人类对这些技术缺乏信任。决策智能实践提升了决策模型的透明度、可解释性、公平性、可靠性和责任机制，这对于实现业务差异化的技术部署至关重要。
- 更严格的法规使风险管理更加普遍。随着隐私和道德准则相继出台，以及新法规和政府规定不断颁布，企业机构越来越难以充分了解其决策的风险影响，而决策智能有助于决策模型的明确表述，从而降低这种风险。
- 难以确定整个企业机构决策的一致性。决策缺乏明确的表述，不利于适当协调集体决策成果。决策智能可对此进行纠正。
- 决策智能平台（DIP）软件工具的出现，将帮助企业机构切实落地决策智能项目和战略。

## 阻碍因素

- 碎片化：决策孤岛导致数据、能力和技术集群难以协调，可能减缓决策模型的实施。
- 运营结构：围绕先进技术的组织结构不完善，例如，AI卓越中心的缺失可能会阻碍决策智能的进展。
- 自主决策模型（从嵌入式分析资产到独立的机器代理）可能会带来损害商誉的结果，并且企业机构对此整体影响缺乏了解，这会阻碍决策智能的采用。
- 业务部门之间缺乏适当的协调，并且无法公正地重新考虑部门内部和部门之间的关键决策流，这会削弱决策智能早期工作的有效性。
- 缺乏更宏观的建模：企业机构几乎仅关注技术技能，人类决策的心理、社会、经济和组织等其他关键因素并未得到重视。

## 使用建议

- 构建决策模型时，遵循以提升可追踪性、可复制性、相关性和可靠性为目标的原则，提高跨部门决策的弹性和可持续性。
- 通过模拟决策代理的集体行为，同时对比评估其全局贡献与局部优化，提高决策代理的可预测性和一致性。
- 在传统和新兴的决策增强和决策自动化技术方面，拓展员工的专业技能，包括描述性、诊断性、预测性和规范性（优化、业务规则）分析。
- 与行业专家、AI专家和业务流程分析师合作，根据每个决策情况的特定要求来调整决策技术的选择。

## 厂商示例

第四范式、阿里云、杉数科技、九章云极、百分点科技、顺丰科技、芯盾时代、悠桦林、中科闻歌

## Gartner相关推荐阅读

[Innovation Insight for Decision Intelligence Platforms](#)

[Reengineer Your Decision-Making Processes for More Relevant, Transparent and Resilient Outcomes](#)

[How to Choose Your Best-Fit Decision Management Suite Vendor](#)

## 数据编织

分析师: Xingyu Gu, Fay Fei

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率低于1%

成熟度: 发展阶段

## 定义:

数据编织是一种设计框架，用于获得灵活而且可复用的数据管道、服务和语义，涉及数据集成、主动元数据、知识图谱、数据剖析、机器学习和数据分类。数据编织颠覆了现有的数据管理主导方法，不再针对数据和用例“量身定制”，而是“先观察再使用”。

## 为何重要

数据、分析和AI用例的出现，以及快速变化的数据安全法规，导致了中国数据管理的复杂性和不确定性。数据编织能充分利用沉没成本，同时也能为数据管理基础设施方面的新支出提供优先级排序和成本控制指导。这项技术提高了基础设施的灵活性以及规模和功能的扩展空间，确保数据可在多个场景和环境之间使用（无论是通过人员还是机器来使用）。

## 业务影响

数据编织的优势包括：

- 为数据工程师提供洞察，最终实现数据集成、质量、交付和访问相关可重复任务的自动化
- 为多云和混合基础设施提供支持，提供更便利的数据集成、查找、治理和共享
- 为上下文和业务含义注入语义知识，丰富数据模型
- 可演进成为自我学习模型，识别以不同形式和结构存在的相似数据内容，拥有连接更多类型新资产的能力
- 监控数据资产，以实现优化和成本控制

## 推动因素

- 很多企业机构发现，目前的数据获取和集成方法并不完善。数据编织可通过多种方式提供集成数据，包括提取、转换和加载（ETL）批量化、数据虚拟化、消息队列、应用编程接口（API）和微服务。
- 在横跨本地、多云和混合生态系统的异构环境中，中国企业机构很难提升数据管理的效率。数据编织可确保在不破坏和替换现有基础设施的情况下集成、治理和共享数据。
- 在多数中国企业机构内，不同类型数据的积累以及AI技术的采用，导致了由一系列数据孤岛构成的数据和分析（D&A）复杂架构，因此越来越多的企业机构需要自动化程度更高的D&A架构运营方式。
- 跨部门沟通效率低下，仍是D&A投资持续创造业务价值的主要障碍。数据编织的主动元数据功能，提供了针对数据使用和模式的更多洞察，改善了跨数据孤岛的交流、洞察和数据共享。
- 数据跟踪、审计、监控、报告和评估的需求不断增加，导致人工任务负载增加，影响了成本效益和D&A的敏捷性，因此需要采用自动化程度更高的数据管理概念。
- 新的技术进步使数据编织能够辅助图数据建模，对于保留数据上下文及其复杂关系具有重要意义。这些进步也为业务用户提供了更便利的访问，使其能够利用多方商定的语义丰富分析模型。

## 阻碍因素

- 不被共享的专有元数据阻碍了数据编织从各种应用和数据源（如日志文件、跟踪文件、应用日志和脚本）获取和共享元数据。尽管数据编织可用于设计时和运行时元数据的分析和机器学习，但这些功能在初始阶段非常容易出错。
- 只有少量的中国本土数据和分析厂商在产品中嵌入了主动元数据和知识图谱等必要技术，导致数据编织的概念和成功实践之间存在鸿沟。
- 构建数据编织的技能和平台种类繁多，带来了技术和文化障碍。数据管理的基础，需要从分析、需求 and 设计转向发现、响应和推荐。
- 对于如何协调数据编织和传统的数据和分析治理项目，存在误解和知识缺口，导致实施难度较高。

## 使用建议

- 对所有数据管理项目中的“主动元数据”工作进行梳理，确定其优先级。主动元数据是数据编织的核心能力和差异化优势。
- 从资源效率、性能、安全、合规与可用性的角度，评估供应商如何通过内外部元数据的共享和利用来实现相邻应用之间的自动化运营。严重依赖人工（如驻场技术人员服务）来完成上述任务的供应商，在选型时应降低其优先级。
- 如果某个解决方案与其相邻平台即服务（PaaS）/软件即服务（SaaS）解决方案的元数据访问相隔离，则降低其优先级。
- 投资增强数据目录，协助创建灵活的数据模型。通过语义和本体来丰富模型，以便业务部门使用目录。
- 部署充分利用知识图谱的数据编织。
- 利用业务语义丰富知识图谱能力，确保业务流程专家为数据编织提供支持。

## 厂商示例

阿里云、数语科技、Denodo、亿信华辰、IBM、精鲲

## Gartner相关推荐阅读

[Quick Answer: What Is Data Fabric Design?](#)

[Emerging Technologies: Critical Insights on Data Fabric](#)

[Case Study: An Active Metadata Augmented Data Classification System to Boost Analytics Efficiency](#)

[Quick Answer: What Is Active Metadata?](#)

[From Logical Data Warehouse to Data Fabric](#)

## 数据资产管理

分析师: Tong Zhang, Julian Sun

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为5%~20%



## 成熟度: 发展阶段

### 定义:

数据资产管理是指管理、处理和利用对业务运营来说具有宝贵资产价值的数据的过程。数据资产管理适用于多种数据形态——例如，系统中的图像、视频、文件、资料和交易数据，并涵盖从数据获取到销毁的整个数据生命周期，目的是以管理资产的方式管理数据，并从中创造价值。

### 为何重要

数据作为一种新的生产要素，已成为企业机构的竞争优势。数据具有快速、多样、大量和描述事实的特点，因此企业机构必须整合流程来生成数据洞察。数据资产不仅能提升运营质量和决策水平，更可以创造更多业务价值，还能够产生新的业务模式和利用数据直接变现。然而，尽管价值创造在加速，数据资产仍存在潜在风险。企业机构必须谨慎管理数据资产，避免监管违规和数据意外泄露。

### 业务影响

数据被认为是无形资产，可惠及业务决策、业务模式设计、运营效率和成本削减等各领域。打造数据驱动型企业，并非只是鼓励将数据用于决策这么简单。企业机构还必须发展数据能力，并使数据工作与企业创造价值的目标保持一致，从而建立数据资产管理。如果数据在资产负债表上被确认为无形资产，就可能对财务报表产生直接影响。这也是数据资产管理至关重要的原因之一。

### 推动因素

- 直接数据变现要求数据资产管理为价值创造奠定良好的基础。
- 高级分析和人工智能的成功，离不开数据资产提供的大量优质数据输入。
- 数据和分析（D&A）用例，包括客户体验、客户留存率和销售转化率改进等，依赖于历史交易数据、服务记录和交互行为数据。所有这些数据都属于数据资产管理的对象。

## 阻碍因素

- 数据资产往往孤立存在，并受到各个部门的保护，导致管理难以统一。
- 根据现行会计规则，将数据确认为无形资产的做法会受到限制。企业无法采用会计模型来衡量这类资产并为其分配价值。
- 企业机构很难量化数据资产管理的发展情况。这项工作需要花费数小时进行数据收集、清洗和增强，还要计算人员招募、数据存储和计算的成本，以及其他成本。
- 数据在很大程度上是情境化的。原始数据本质上是字母数字，随着其在价值链中的移动而增加价值。根据数据在不同场景下的使用为其分配价值之所以困难，原因在于未来收益的不确定性。随着监管和数据隐私要求的提高，不必要的数据收集可能会给企业机构带来合规风险。

## 使用建议

- 制定可体现高管团队期望的数据驱动型企业愿景。明确对数据资产的期待，例如数据资产变现、建立数据和分析治理，以及通过数据和分析改善决策能力等，并确定实现这些期待的优先级。
- 确保数据和分析的权威和责任与企业的数据驱动目标保持一致。与企业高管共同确定对首席数据和分析官（CDAO）角色的期望及其责任，并消除与其他高管职责相关的歧义。
- 改变数据和分析的运营模式，以弥补相关的能力差距。在整个企业范围内为业务赋能，发展数据驱动型文化，确保成功实现企业愿景。

## 厂商示例

阿里云、数语科技、亿信华辰、华为、普元

## Gartner相关推荐阅读

[Quick Answer: How Should Chinese Enterprises Better Deliver Data Monetization Regarding “20 Data Measures”?](#)

[3 Ways to Promote Your Data Agenda at the Center of the Chinese Digital Economy](#)

## 负责任的AI

分析师: Mike Fang, Tong Zhang, Ben Yan

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为1%~5%

成熟度: 发展阶段

定义:

负责任的AI是一个总括性术语，是指企业机构为采用人工智能（AI）而做出适当的业务和伦理决策时所要考虑的各个方面，包括商业和社会价值、风险、信任、透明度、公平性、偏见减轻、可解释性、可持续性、责任制、安全性、隐私和监管合规性。负责任的AI涵盖了企业机构的职责和实践，旨在确保AI的开发和运营以积极、可靠、符合伦理道德的方式进行。

为何重要

负责任的AI已成为Gartner客户最关心的AI议题。以AI取代人类进行决策时，会出现好坏两极分化的结果。负责任的AI可以解决价值交付与风险容忍的平衡困境，实现正确的结果。这要求供应商和企业采用一套工具和方法，包括行业特定的方法。越来越多的司法辖区提出了新的AI法规，要求企业机构以有意义的方式做出响应。

业务影响

负责任的AI意味着在个人、企业机构和社会层面，对AI的开发和使用逐步实行责任制。如果AI治理由指定的团队执行，那么负责任的AI会涉及到AI流程中的每个人。负责任的AI有助于实现公平，即使数据中难免存在偏见；有助于获取信任，虽然透明度和可解释性方法在不断演进；也有助于确保合规性，虽然AI具有概率性。

推动因素

- 2019年3月，中国科学技术部成立了国家 [新一代人工智能治理专业委员会](#)，由知名学者、企业家和政府官员组成，就AI相关的道德问题提供建议。
- 2019年6月，专业委员会发布了《新一代人工智能治理原则——发展负责任的人工智能》，强调AI使用的“安全可控、责任共担”，提出了和谐友好、公平公正、包容共享、尊重隐私、安全可控、共担责任、开放协作、敏捷治理等八项原则。
- 中国政府于2022年12月发布了 [《中国关于加强人工智能伦理治理的立场文件》](#)。
- 2023年4月11日，中国国家互联网信息办公室发布了 [《生成式人工智能服务管理暂行办法》](#)。
- 依赖于AI模型输出的企业机构希望其模型具有可靠性和稳定性。

## 阻碍因素

- 法规要求使企业将工作重点放在了合规方面，而忽视了负责任的AI的其他驱动因素。
- 快速发展的AI技术，包括用于可解释性、偏见检测、隐私保护及合规性的工具的发展，使企业机构误以为采用这些工具和技术即履行了责任，但其实仅依靠技术是不够的。除了采用技术之外，还有必要采取周密的AI伦理和治理方法，综合考虑其中的各个环节。
- 中国的AI发展议程强调技术开发的“顶层设计”和以国家为中心的开发方式，这可能会使负责任的AI成为纸上谈兵——比如缺乏灵活性，无法在执行层面为企业部署负责任的AI提供指导。

## 使用建议

- 遵守政府规定，根据法规要求来定义和调整AI技术和项目的采用。
- 结合负责任AI的角度，在所有重点关注领域宣传一致的方法。负责任的AI在企业中最典型的应用领域包括用于满足公平性、偏见减轻、伦理、风险管理、隐私和合规性要求。
- 指定一名牵头人，负责每个用例中负责任AI的开发和使用。
- 定义模型设计原则，应对负责任的AI在模型开发各个阶段出现的问题。尝试解决艰难的权衡问题。为员工提供有关负责任AI的培训。
- 成立AI伦理委员会并制定伦理原则，以解决AI相关的矛盾。确保参与者的多元化，并为表达AI相关问题提供便利。

## Gartner相关推荐阅读

[3 AI Priorities to Increase Consumer Trust for Retailers in China](#)

[Quick Answer: How Should Chinese Enterprises Use Privacy-Enhancing Computation in Artificial Intelligence Initiatives?](#)

[A Comprehensive Guide to Responsible AI](#)

## 数字伦理

分析师: Mike Fang, Julian Sun

影响力评级: 较高

市场渗透率: 目标受众覆盖率为20%~50%

成熟度: 主流采用起步阶段

定义:

数字伦理是指人、企业机构和物件之间进行电子交互时应遵循的价值理念和道德准则，涵盖人工智能（AI）、数据和分析以及社交媒体等领域。

为何重要

数字伦理，尤其是隐私和偏见问题，已成为行业和社会热议最多的话题之一。社会舆论的声音越来越得到重视，而负责任地使用AI也迅速成为个人、企业机构和政府关注的焦点。人们逐渐意识到自身信息的价值，并且对信息使用缺乏透明度、信息滥用和违规使用感到沮丧。企业机构正在采取行动，以降低个人数据管理和安全方面的风险。同时，政府也在相关法规的实施方面更加严格。

业务影响

数字伦理有助于企业机构在客户、员工、合作伙伴和社会中产生积极影响，提高商誉。数字伦理对业务的影响涉及创新、产品开发、客户互动、企业战略以及产品上市等诸多环节。关键在于出发点，如果数字伦理只是实现业绩的一种方式，就显得缺乏诚意。在数字伦理方面表现出色的企业，其目标应是为相关各方和整个社会提供服务，从而提升企业的信誉和业绩。

推动因素

- 为加强科技伦理治理，中国发布了相关指导意见。文件明确了科技伦理的原则，认为科技活动应当为人类福祉服务、尊重人的生命权、坚持公平正义、合理控制风险、保持公开透明。
- 董事会成员和其他高管都对使用创新技术可能带来的意外后果表示担忧。
- 有关数据和技术对企业和社会产生影响的媒体热点报道越来越多。
- 随着AI的出现，有关伦理的讨论在技术广泛实施前和实施中都会出现。人们采取AI伦理和其他负责任AI措施，试图扭转围绕AI的负面舆论，并且更负责任地使用AI。
- Gartner预测，到2024年，30%的主流企业机构将采用新的“社会之声”指标来决定其对社会问题的应对措施，并评估相关问题对业绩的影响。在科技伦理方面，社会之声将给政府和公共/私人企业机构带来更大的压力。在社会舆论中，“科技巨擘”在科技伦理方面的负面形象已深入人心。

## 阻碍因素

- 数字伦理的定义模糊，企业机构很难将其投入运作，也难以投入大量精力贯彻最佳实践。
- 不同的人、地区和文化有着不同的伦理要求。即便在较为重视伦理问题的企业机构，内部与外部利益相关者（如客户）有时也难以就此达成共识。
- 由于社会对伦理的理解不甚明晰，企业机构也很难确定数字伦理目标。企业机构的立场和信念甚至可能导致数字伦理与多数人的看法相背离。
- 数字伦理经常只能被动响应，被狭隘地理解为合规性，局限于对隐私保护的技术支持，并且/或仅被视为可解释型AI。
- AI伦理目前是整個数字伦理的主要关注点，其支撑技术的成熟度还需要进一步提升。

## 使用建议

- 明确具体的数字伦理问题和相应的机遇，将意识落实为行动。
- 从多个角度进行伦理论证，讨论伦理困境。
- 确保已将伦理后果纳入考量范围，同时能胸有成竹地为技术应用辩护，包括对意外出现的负面后果做出解释。
- 关注将数字伦理作为提升社会与商业价值的途径，而非仅仅考虑合规与风险，从而提升相关问题的讨论质量。将数字伦理与具体业绩指标挂钩。
- 确保数字伦理在引领而非跟随数字化转型。在设计阶段尽早解决数字伦理问题，制定快速解决伦理困境的方法。
- 组织伦理培训和研讨会，在所有AI项目中树立伦理意识，并且在AI设计和实施中明确责任机制的重要性。

## Gartner相关推荐阅读

[Tool: Assess How You Are Doing With Your Digital Ethics](#)

[Tool: How to Build a Digital Ethics Curriculum](#)

[AI Ethics: Use 5 Common Principles as Your Starting Point](#)

[Every Executive Leader Should Challenge Their Teams on Digital Ethics](#)

## Expert Insight Video: What Is Responsible AI and Why You Should Care About It?

### 组装式数据和分析

分析师: Fay Fei, Julian Sun

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为1%~5%

成熟度: 发展阶段

#### 定义:

组装式数据和分析（D&A）利用基于容器或业务微服务的架构和数据编织理念，将现有资产组装为灵活、模块化和用户友好的数据分析和人工智能（AI）能力。这项技术可在低代码和无代码功能的支持下，综合运用一系列技术将数据管理和分析应用转变为数据分析和AI组件或其他应用模块，并且支持自适应和智能决策。

#### 为何重要

面对瞬息万变的业务环境，中国企业机构需要提高敏捷性，加快洞察产出速度。组装式D&A有助于企业机构使用模块化数据和分析能力，在各项措施中融合多个洞察与参考信息，避免割裂式开发。企业机构可通过组装或重组D&A能力进一步提高交付的灵活性，应对不同使用场景。

#### 业务影响

中国企业机构过去针对独立的D&A项目进行投资，现应关注下列影响：

- 组装式D&A有助于应用/公民/分析开发人员在应用和软件中构建分析能力，实现前所未有的业务价值。
- 日益严重的D&A孤岛问题，限制了投资价值的实现。组装式D&A可在大范围重复使用的情境下正确识别数据对象。
- 这项技术有助于企业机构嵌入与组装AI增强型功能，从而提高先进分析能力，解决更为复杂的问题。

## 推动因素

- 中国新兴的低代码或无代码平台市场，为构建组装式数据和分析应用创造了更多机会。无/低代码解决方案可用于获取数据或以技术难度较低的方式组装更多的数据分析能力，例如交互式可视化、预测建模等方法来丰富嵌入式分析。
- 中国企业机构逐渐增多的分析需求，推动了组装式数据和分析。很多企业机构开始寻找可在多种场景下实现数据分析能力复用和组装的解决方案。
- AI在中国的使用率逐渐增高，企业机构可通过组装来打通商业智能（BI）和人工智能（AI），实现BI能力的扩展，在无需调用多种应用的前提下为用户提供全面、灵活甚至是个性化的解决方案。
- 很多企业机构需要动态结合描述性、诊断性、预测性和指示性分析能力，在决策流程的不同阶段提供参考，以互联、连续、情境化的方式为决策提供分析支持，推动有效的行动。
- 企业机构期待着政府引导的“数据共享机构”、“行业云”等项目能成为分发和共享数据和分析应用的有效渠道，而组装式数据和分析有助于企业机构轻松共享和寻找所需的组件，更好地完成数据分析变现。

## 阻碍因素

- 组装式数据和分析在中国仍处于起步阶段，市场中很少有专门提供组装式功能的成熟解决方案。
- 中国的数据分析团队缺乏与业务和财务团队合作的经验，在应对日益严重的管理和治理问题时也缺少经验指导。
- 基于云的架构推动了组装式数据和分析的发展，毕竟许多新兴功能来自云优先或纯云解决方案。不过，中国企业的领导者可能对云采用持相对保守的态度，这或许会影响通过组装式数据和分析获得价值的速度。
- 在许多中国企业机构中，开发团队和数据分析团队之间的紧密合作并不常见，同时也缺乏标准流程和最佳实践来支持实现组装式数据和分析的工作模式。
- 组装式数据和分析还需要供应商之间更加紧密的合作。如果没有技术方面的深度合作，中国现有数据和分析产品的可组装性还不够完善。



## 使用建议

- 与业务利益相关者合作，从现有分析组合中引入可组装、模块化、可复用的数据分析功能，提高决策质量和业务影响。
- 利用组装式分析，将先进的数据科学和机器学习功能整合到分析应用中，从而推动创新。
- 建立联合团队，并且确保应用开发人员和业务分析师之间的持续协作，探索为应用增加分析能力的机会。重新思考组织架构、流程和技能，以支持分析服务的敏捷组装和重组。
- 在云端进行组装式分析试点，据此建立分析交易平台，推动并支持协作和共享。
- 寻找在技术和业务模式方面支持组装式数据和分析的供应商，例如应用编程接口封装、治理封装等。

## 厂商示例

滴普科技、衡石科技、科杰科技、跬智信息

## Gartner相关推荐阅读

[Use Gartner's Reference Model to Deliver Intelligent Composable Business Applications](#)

[Predicts 2023: Analytics, BI and Data Science Composability and Consolidation](#)

[Client Question Video: What Is Composable Data and Analytics?](#)

[Adopt Cloud Analytics to Drive Innovation](#)

## 数据共享

分析师: Mike Fang, Xingyu Gu

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为1%~5%

成熟度: 主流采用起步阶段

## 定义:

数据共享是提供企业价值的一项关键业务流程，可提高效率，有助于不同用户之间的数据复用。数据共享对内发生于共有实体之间，例如母公司/子公司/姐妹公司；对外发生于共同控制实体之外的各方，例如跨行业的计划或跨公共部门的项目。数据共享有助于改善协作。

## 为何重要

发展数字经济是中国企业的一项战略举措，而数据共享在其中发挥着至关重要的作用。数据共享为企业提供了机会，使其能够利用内外部数据来支持更有效的业务决策，通过数据变现获得额外的收入流，并与合作伙伴建立协作生态系统，而仅使用内部数据则无法做到这些。高绩效企业机构中的数据和分析（D&A）领导者可破除数据访问障碍，或打破数据孤岛；推动数据共享，或增加与业务论证相匹配的数据访问。

## 业务影响

数据共享通过以下方式提供社会和业务价值：

- 增加更相关数据的访问机会，以满足社会或业务目标，在推动业务成果的同时实现利益相关者的价值
- 利用内外部共享数据来支持更高效的业务决策
- 通过数据变现获取额外收入流
- 通过合作伙伴建立协作数据生态系统
- 提高已部署/新建模型的准确性
- 推动卓越运营和环境可持续性

## 推动因素

- 自2020年以来，中国政府就将数据视为继土地、劳动力、资本和技术之后的第五大生产要素（请参阅 [China Unveils Guideline on Improving Market-Based Allocation of Production Factors](#)）。作为新生产要素，数据的共享对于经济发展和增长至关重要：可促进资源的优化配置、提高生产率并增加经济效益。
- 政府推动数据共享，深化公共数据资源的开发和数据驱动型用例的发展。例如，最近推出的各种政府支持的数据交换计划，旨在实现不同企业之间数据资产交易的标准化管理，以提升业务和社会效益。
- 最近出台的数据法规和数据分类分级指引（金融、政府、医疗、汽车、工业和IT行业）通过监管和标准化的方式，为数据共享计划的可持续发展提供支持。
- 中国新兴技术的快速发展，如区块链和隐私增强计算（PEC），有助于增强数据共享流程的可靠性和合规性。
- 由于存在隐私法规，真实数据可能会十分昂贵、无法获取或不可用。因此，合成数据的数量出现增长，而这也说明数据共享需求有所提升。
- 人工智能（AI）训练缺乏所需的相关数据，在处理大量AI训练数据（例如大语言模型的数据语料库）时也面临着可持续性和成本压力。
- 模型环境的不断演变，也需要持续共享近实时数据。
- 由于客户行为的相移，专门使用内部或第一方数据训练的模型出现了模型漂移。因此，外部数据在支持预测性模型方面的相关性有所提高。
- Gartner观察到，企业机构对利用更多样化数据源生成更强大的预测性分析的需求日益增加，以推动通过相关、独特或难以获取的洞察实现数据驱动型创新。

## 阻碍因素

- 在数据所有权、定价、流通和治理方面缺乏成熟实践，阻碍了数据共享
- 对如何实现规模化、可信的数据共享缺乏了解
- 由于惧怕、数据管理和治理策略过时，以及缺乏相应工具/技术，利益相关者存在抵触情绪
- 认为数据滥用的风险超过了成本节约和营收增长等数据共享的业务收益，这一想法虽然已经过时，但仍然存在
- 存在内部数据囤积、外部数据劫持、机密性武器化和隐私羞辱等问题
- 对技术赋能因素过于依赖，可能会掩盖成功实现数据共享所必需的数据素养要求、流程/策略和协作的重要性
- 未能对数据共享进行适当的投资，阻碍了实现价值创造型高效数字业务成果所需的发现、复用和再共享

## 使用建议

- 实现数据管理的现代化，采用基于信任的D&A治理，培育数据共享文化，而非数据所有权文化。
- 评估潜在的AI和分析用例，利用数据共享实现更高/新的业务价值。
- 通过数据资产的变现、管理和估量，从原始数据开始搭建D&A产品，以打造新的收入来源，吸引更多合作伙伴。
- 明确与数据产品相关的价值和风险。通过适当的数据交换渠道，弥合数据提供商和消费者之间的交付能力差距。
- 利用不同的技术，如区块链和隐私增强计算，以更可靠、更合规的方式编排数据共享流程。
- 打破企业机构之间的边界，培养外部相关方的数据素养，以加速数据共享，提升业务成果。

## 厂商示例

翼方健数翼数坊、中国移动数联网、华为交换数据空间、星环科技Transwarp Navier

## Gartner相关推荐阅读

[Top 3 Priorities for Chinese Enterprises to Promote Effective Data Sharing](#)

Quick Answer: How Should Chinese Enterprises Use Privacy-Enhancing Computation in Artificial Intelligence Initiatives?

Quick Answer: How Should Chinese Enterprises Better Deliver Data Monetization Regarding “20 Data Measures”?

3 Ways to Promote Your Data Agenda at the Center of the Chinese Digital Economy

湖仓一体

分析师: Xingyu Gu, Julian Sun

影响力评级: 中等

市场渗透率: 目标受众覆盖率为5%~20%

成熟度: 发展阶段

定义:

湖仓一体是一种融合式的数据基础设施环境，结合了数据湖的语义灵活性与数据仓库的生产优化和交付能力。通过对原始数据进行精细化处理，最终将优化后的数据交付给消费者使用。

为何重要

中国的企业机构仍在努力开展集成数据仓库与数据湖等系统的实践探索。采用基于Apache Hadoop的大数据技术后，架构变得更加复杂。湖仓一体可以统一分析型数据库、简化架构、提高效率，同时最大程度地减少系统之间数据的迁移。因此，一个运营足迹更小、效率更高的环境成为可能。

业务影响

能为企业带来收益的不仅是简化交付和快速访问数据，还有既能为技术娴熟的数据科学家、工程师和分析师提供支持，也能为通过内置报表或仪表板使用数据的普通用户赋能的综合性数据管理平台。湖仓一体提供了一条定义明确的路径，从以探索为导向的分析和模型开发（经数据湖）发展为向最终用户交付分析洞察和显著价值（经数据仓库）。

## 推动因素

- 由于分析架构复杂，中国企业机构的数据科学/AI项目上线投产面临挑战，而最近出现的生成式AI用例加剧了这种困境。湖仓一体可以统一不同的分析用例，降低D&A架构的复杂性。
- 数据湖和数据仓库专注于不同的领域。数据湖支持数据科学和临时分析，可以支持灵活的数据结构和延时需求，还可以对容器进行管理。数据仓库擅长处理需要审计追踪以及对质量和准确度有高要求的提炼后数据，或者特殊的数据结构（如维度、时间序列、层次结构等）。湖仓一体融合了这两种架构，有望满足所有这些数据要求和业务用例。
- 许多云数据仓库和几乎所有云数据湖都已使用语义灵活的云对象存储来保存记录。这就自然而然地需要对这些存储环境进行统一，从而减少分散和重复的基础设施。
- 湖仓一体概念正迅速走向成熟，是得益于市场对数据仓库、数据湖及两者利弊的充分理解。
- 湖仓一体以一体化平台为基础，提供各种用户技能。数据工程师可以创建管道、摄取原始数据并将其传输到湖仓中的数据湖区。然后，数据科学家、超级用户和管理员可以在数据仓库区对数据进行细化和建模。分析师可使用此数据生成报表，或者完成其它重复性分析工作负载需求。

## 阻碍因素

- 由供应商建造的湖仓一体平台仍在发展阶段。一部分供应商拥有强大的数据湖功能，但不支持数据仓库可以满足的各类事务一致性以及工作负载管理功能；另一部分供应商拥有强大的数据仓库功能，但所支持的数据模型有限或缺乏数据湖的数据科学功能。
- 中国的一些企业机构已经对基于Hadoop/Spark的数据仓库和大数据平台进行了大量投资。从投资回报率的角度来看，这些企业机构更愿意改进现有架构，而不是用湖仓一体取而代之。
- 数据管理的整个优化范围包括数据质量、安全、性能，尤其是良好的元数据管理和数据集成。湖仓一体平台很少能解决所有这些问题。
- 企业永远在寻求快速的、无需提炼加工的数据访问，以克服与数据仓库相关的交付延迟。湖仓常被认为是解决这一问题的又一个灵丹妙药，但这是不切实际的期待。

## 使用建议

- 采用以使用场景为基础的思考方式，解决特定问题，并在此基础上进行扩展，以实现长期成功。随着时间推移，湖仓一体会像数据湖和数据仓库一样，发展出更多的使用场景。
- 全面测试候选解决方案，确保湖仓一体产品能够提供可靠的、高性能的工作负载，避免“承诺过高、交付不足”。
- 仔细评估投资回报率，确认更换（相较于改进）是正确的选择。多数基于大规模并行处理（MPP）的数据仓库和基于Hadoop的数据湖都可以演进，以处理更多种类的分析用例。
- 在处理更大范围的数据和分析问题时，应选择逻辑数据仓库（LDW）。湖仓一体是机缘巧合下从逻辑数据仓库中衍生的一个子概念，而逻辑数据仓库仍然是一个成熟的最佳实践。
- 评估安全和治理能力，确保湖仓一体符合企业标准和数据要求。

## 厂商示例

阿里云、滴普科技、华为云、偶数科技、巨杉数据库、顺丰科技、腾讯云、星环科技

## Gartner相关推荐阅读

[Market Guide for Analytics Query Accelerators](#)

[Does My Organization Need a Data Lakehouse?](#)

[Exploring Lakehouse Architecture and Use Cases](#)

[Cool Vendors in Data Management, China](#)

## 膨胀期技术

### 增强型数据和分析

分析师: Fay Fei, Julian Sun

影响力评级: 中等

市场渗透率: 目标受众覆盖率为1%~5%

成熟度: 发展阶段

#### 定义:

增强型数据和分析（D&A）整合了机器学习（ML）/人工智能（AI），以支持数据管理和分析能力。这项技术可实现数据管理和分析工作流的自动化，推动数据和分析在企业内的采用，同时通过ML/AI提高工作效率。与人工分析相比，该技术允许分析用户更快地为企业提供更深入的洞察，同时也可降低技术要求、减少分析中的偏见。

#### 为何重要

很多企业机构在寻求以更简便、更可靠的方式完成数据和分析的创建和管理。中国企业机构的整体数据素养仍然较低，因此采用增强能力可以减少专家的人工任务量，弥补人员技术能力的不足。此外，随着ML/AI技术的发展，增强分析也可提供深入的情境化洞察，揭示额外的业务洞察并发现机会。

#### 业务影响

业务分析师、决策人员和运营人员创造的先进分析功能，使增强型数据和分析补充和扩展了现有的数据和分析平台：

- 这项技术缩短了用户管理和运营数据的时间，使用户有更多时间针对最符合业务需求的数据洞察开展协作和采取行动。
- 通过自动化洞察，一线员工可以访问情景化的分析和指导建议，从而改进决策，更快速地采取正确行动。

#### 推动因素

- 生成式AI及其代表性解决方案OpenAI ChatGPT的相关炒作，使增强型数据和分析的关注度进一步上升。这些技术尽管仍处于早期阶段，尤其是在数据管理方面，但仍然鼓励中国企业进行了大量的实验和探索。



- 随着中国超大规模云厂商的迅速发展，云部署带来的弹性和高性能计算工作负载为实现增强型数据和分析能力提供了技术基础。
- 从数据管理的角度来看，企业机构往往缺乏构建高效数据和分析平台的技术和专家。使用增强型数据和分析能更充分地运用D&A平台，可以推荐和实施新的设计、模式和查询，推断数据的语义和关联，从而提出数据结构改进建议。
- 从分析的角度来看，企业机构并不满足于描述性分析，而是希望获得高附加值的诊断性、预测性和指示性分析。这类分析可以通过增强型数据和分析功能来实现，例如围绕高影响力人士提供数据、识别异常值、进行预测，以及自动生成具有深度分析洞察的叙述和故事。

## 阻碍因素

- 成熟度和数据素养水平较低的企业机构，可能对增强技术的期待过高，不切实际。
- 虽然为增强型数据管理和分析找到统一解决方案的炒作热度很高，但很多企业机构对此并没有明确需求。
- AI支持的数据和分析能力可信度和透明度较低，自动化洞察（包括来自中国供应商解决方案的洞察）的可解释性不足。
- 增强型数据和分析并不能抵消对企业员工数据素养需求，而很多企业机构缺乏培训计划，需要在专家的帮助下解读增强型分析结果。

## 使用建议

- 将数据管理和分析方面的增强能力作为数字化转型战略的一部分，以惠及更多用户。首先要明确从中获益最多的业务和用户，通过试运行来证明价值、建立信任。
- 将增强能力作为数据和分析产品未来采购和升级时的“必备”选择标准。审核增强型数据和分析部署的结果，评估扩大使用面时产生错误和降低性能的风险。
- 投资于图形技术和可解释型AI等新兴技术，充分运用增强型数据和分析的力量。
- 组织培训和研讨会，提高数据素养；实施更多的内部激励，以鼓励员工使用增强型能力。

## 厂商示例

阿里云、数语科技、广州思迈特软件、海致科技、杭州观远数据、跬智信息、网易数帆

## Gartner相关推荐阅读

[Adopt Cloud Analytics to Drive Innovation](#)

[Market Guide for Augmented Analytics](#)

[Cool Vendors for Analytics Platforms in China](#)

[Market Guide for Analytics Platforms, China](#)

国产人工智能芯片

分析师: Roger Sheng, Mike Fang, Julian Sun

影响力评级: 较高

市场渗透率: 目标受众覆盖率低于1%

成熟度: 发展阶段

定义:

人工智能（AI）芯片是一种半导体器件，为运行AI算法（如深度神经网络[DNN]）而进行了优化，通常用于加速AI工作流的特定模块，例如训练和推断。AI芯片是大幅提升AI处理能力的重要组成部分，也可以减轻数据中心、边缘和设备的中央处理器（CPU）工作负载。与NVIDIA等全球厂商相比，中国的AI芯片设计企业仍处于初期发展阶段。

为何重要

中国企业正在大举投资于AI技术领域，而AI加速器芯片是其中的关键组成部分。然而，从2022年起，美国出口管制条例开始限制向中国公司出口先进的AI芯片，迫使中国企业寻找替代的AI芯片资源。成熟的IT巨头与获得风投的初创公司都已开始开发AI芯片，以迎接AI领域持续的全球竞争。

业务影响

对于基于AI的工作负载，AI芯片的作用比传统芯片架构更加重要。大多数超大规模云服务提供商使用高性能的AI加速器芯片来提高海量数据的处理效率。最近备受热捧的生成式AI需要数千个AI加速器来训练基础模型。这些芯片通常采用先进的半导体工艺技术制造，后续的生产成本和能耗十分高昂。

## 推动因素

- 对AI创新用例的需求。最近备受热捧的生成式AI基于基础模型，需要更强大的AI加速器芯片，尤其是在训练阶段。由于美国政府的出口限制，中国企业不得不将目光转向本土芯片供应商，将其作为后备替代供应商。中国的大型成熟企业已建立了内部芯片开发团队，并投资了本土新兴厂商。
- 政府的芯片国产化计划。中国政府积极支持在IT领域打造国产生态系统。考虑到中国对高科技产业发展和网络安全的需求，国产化计划在战略性高性能计算领域尤为重要。中国政府鼓励在政府主导的数据中心项目中更多地使用国产芯片。例如，寒武纪AI芯片业务的主要客户，目前就是政府客户。
- 系统运行效率。当前基于图形处理器（GPU）架构的AI加速器芯片产品供应主要由全球供应商主导，这些供应商的产品价格高昂且产量有限。虽然GPU在通用AI数据训练中表现卓越，但专门为AI应用定制的专用集成电路（ASIC）在能效方面具有优势。这也是百度决定在其搜索业务中使用内部自行开发的AI芯片的原因。
- AI物联网（IoT）应用的需求呈碎片化。AI和IoT的结合是创建数字业务流程的重要因素之一。AI物联网的应用和用例类型多种多样，因此系统需求也各有不同。这为中国的AI芯片厂商提供了绝佳的机会，可以针对各类AI物联网需求开发优化的解决方案。例如，中国制造的监控系统所使用的支持AI计算视觉功能的AI芯片，多数是由中国芯片厂商开发的。

## 阻碍因素

- 缺乏通用的AI工具包支持。NVIDIA的关键成功因素之一就是其CUDA平台。该平台基于NVIDIA的GPU架构，可支持大多数AI开发工具。中国的AI芯片企业必须加大对软件的投资，以支持主流的AI开发工具包。
- 开发者社区尚未成熟。中国企业缺乏构建广泛、跨企业的软件开发生态系统的经验，而这对于AI应用开发至关重要。
- 美国对先进半导体制造的限制。美国禁止中国企业向全球代工厂商购买用于高性能计算的先进芯片（例如，5纳米和3纳米芯片），还禁止将用于14纳米和更先进的半导体芯片制造设备运往中国。

## 使用建议

- 与成熟的中国半导体企业合作，在AI基础设施方面采取整体解决方案，在专注于开发定制化的AI芯片的同时积累专业知识。
- 关注和参与不需要先进工艺节点的新半导体技术的研发，例如异构集成、RISC-V CPU和芯片间光互连。
- 对软件优化和专用应用的定制芯片设计给予更多关注，以提高计算系统性能。
- 做好用于AI基础设施的全球供应商先进AI芯片的合理库存，并对此进行投资。
- 统一行业标准与开发平台，与行业同行合作，打造符合中国企业需要的AI生态系统。

## 厂商示例

百度、壁仞科技、寒武纪、燧原科技、华为、海光信息

## Gartner相关推荐阅读

[Market Trend: U.S. Restrictions Force China ICT Industry to Seek Alternative Semiconductor Solutions](#)

[Market Trends: China's New 'Secure and Trustable' Initiative for the IT System Will Accelerate the Growth of Domestic Chip Vendors](#)

## 大模型

分析师: Tong Zhang, Mike Fang

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为1%~5%

成熟度: 成型阶段

## 定义:

大模型是在大范围的数据集上以自监督方式训练的大参数模型，其中大多数都是基于Transformer架构或扩散型深度神经网络架构，并且在不久的将来可能会成为多模态。大模型这一名称是源于其对于多种下游使用场景的重要性和广泛的适配性。这种适应多种场景的能力得益于模型充分和广泛的预训练。

## 为何重要

大模型拥有大量的预训练数据集，适用于十分广泛的用例，因此是推动AI领域发展的重要一步。相较于前几代的大模型，大模型具有更高的效率，可提供更先进的功能。大模型现在已经成为了自然语言处理的首选架构，并已应用于计算机视觉、音视频处理、软件工程、化学、金融和法律领域。大模型衍生出的一个热门子概念就是基于文本训练的大语言模型。ChatGPT就使用了这样的大语言模型（GPT-3）。

## 业务影响

大模型有潜力为各类自然语言用例中的应用提供增强效果，因此将在垂直行业和业务职能中产生深远影响。它们可以提高员工生产力、实现客户体验自动化和增强，并能以经济高效的方式创建新产品和服务，从而加速数字化转型。大模型可以处理多模型输入，因此其用途将超越自然语言处理，并且未来多个大模型可以协作生成更多多样性的输出物。

## 推动因素

### 大模型：

- 只需少量定制化工作即可提供有效成果。预构建应用编程接口（API）、智能提示语工程和精调功能可以让大模型高效地交付价值。其中微调因其定制化能力而能够产生最高价值，而另外两项功能的复杂度较低。Gartner客户表示，他们可以直接使用Google的开源BERT模型，不需要对整个网络进行训练，只需要对最上面几层进行再训练精调，即可针对所需的语言域进行定制。
- 提供卓越的自然语言处理分类。大模型与以前的神经网络解决方案之间存在明显差异。大型预训练模型能以空前的规模和准确度生成具有逻辑的文本、代码、图像、语音和视频。Transformer架构模型不是基于前面的单词预测下一个单词，而是将文本划分为相对较大的文本块进行预测。这一进步极大地改善了语音、语言和文本应用。Google Translate便是一个突出的例子。
- 支持低损耗的创造。大模型能够以最少的输入创建格式规范的文本段落。在过去的一年中，大模型大量涌现，同时带动产生了一批在此基础上创建的特定领域小预训练模型。其中大部分可作为云API或开源项目使用，进一步降低了实验时间和成本。GPT-3和GPT-4是OpenAI开发、Microsoft授权的大模型。这类Transformer模型可从简短的片段中创建段落或长篇文本，原理是根据其吸收掌握的累积训练结果，预测句子中最可能的下一个单词。
- 以巨大的参数规模加速了AI创新。例如，OpenAI的GPT-3和Meta AI的OPT都拥有1750亿个参数。Google DeepMind的Gopher有2800亿个参数，而Google AI的PaLM有5400亿个参数。NVIDIA和微软联合推出的Megatron-Turing NLG有5300亿个参数。阿里巴巴的M6，参数量达到十万亿。此外，Meta和谷歌也开源了各自的部分模型。

## 阻碍因素

### 大模型：

- 无法提供完美的结果。大模型需要精心训练和监管。由于训练方法和黑盒性质，其生成的结果可能出现错误或幻觉，甚至使数据集的偏见或版权问题倍增。
- 需要适当的技能和人才。最终结果取决于训练人员的技能、知识和能力，尤其是提示语工程和精调方面的能力。
- 会扩展到不切实际的规模。大型模型包含的参数多达数十亿甚至数万亿。对多数企业机构而言，这些模型已大得不切实际。训练模型所需的计算资源太多，会推高模型的成本，并且造成负面的生态影响。
- 会导致权力集中。这些模型大多由技术巨头创建，因此权力会集中在少数财力雄厚的实体手中。这种情况可能会在未来造成重大失衡。
- 可能会威胁到人类。大模型凭借其新兴的推理能力，可能会超越人类智能，使AI成为一种威胁。

## 使用建议

- 在现有语音、文本或编码程序中引入大模型。如果所用语言处理系统版本较旧，则迁移到基于Transformer的模型可以显著提高性能。
- 从开源的现有模型着手。这些模型具有卓越的生态系统支持，其数据科学和机器学习平台对安全和隐私拥有足够的企业护栏，并且部署范围更广。
- 探索新的用例，例如自然语言推理、情感分析或基于自然语言的企业搜索。在这些用例中，大模型可以显著提高准确度，缩短上线时间。
- 指定一个孵化团队，用于追踪行业发展、沟通可能性、在业务单元进行实验，并在全公司范围内分享宝贵的经验教训。
- 探索新方法，例如轻量级精调和提示语工程，将业务知识赋予大模型。
- 遵循与AI的价值一致性。评估大模型的风险，并探索设计方法，防止AI系统无意中以有悖于人类价值观的方式行事。

## 厂商示例

阿里云、百度、华为、科大讯飞、腾讯、智谱AI

## 自主可控数据库

分析师: Xingyu Gu, Julian Sun

影响力评级: 较高

市场渗透率: 目标受众覆盖率为20%~50%

成熟度: 成型阶段

### 定义:

数据库管理系统（DBMS，以下简称数据库）自主可控，是中国数据库市场的特点，该领域供应商主要为在中国运营的国内和国际企业提供数据库产品。数据库是用于存储和组织数据的产品，通常是针对具有已定义格式和结构的数据。技术自主可控能使本地企业机构在不断变化的全球环境保持韧性，因此是中国数据库市场的主要驱动力。

### 为何重要

随着中国各行各业数字化转型的不断推进，数据库作为数据基础设施的关键组成部分，需要具备优秀的性能、弹性、强大的AI兼容性和成本效益。与此同时，近年来地缘政治环境高度动荡，也促使中国企业确保其数据库技术能够实现自主可控。这两个因素导致中国迫切需要开发现代化的数据库。

### 业务影响

数据库的自主可控具有战略意义，符合本地企业的数字化转型路线图，其用途在于：

- 使企业机构的做法符合“信息技术自主可控”的国家战略。
- 降低主机硬件昂贵的维护成本。
- 解决本土商业环境中出现极高并发数据交易（例如双11购物节）时的性能瓶颈。
- 确保数据科学和实时分析场景能够进一步扩展，嵌入到生产环境中。

## 推动因素

- 中国对技术基础设施的需求非常旺盛。一方面是因为人口众多，有大量的互联网用户；另一方面是现代化技术的需求日益增多，以便维持智慧城市和电商直播等在中国独具规模的数字化项目。
- 传统数据库解决方案（主要依靠非中国供应商）趋于老化，难以满足这些需求。
- 近期地缘政治环境促使中国企业从进一步投资于非本土技术转向实现自主可控，这推高了中国企业对本土供应商的需求。

## 阻碍因素

- 不成熟的迁移工具和服务，阻碍了企业机构将现有业务系统迁移到现代数据库。由于关键业务系统中的数据库迁移失败会产生巨大业务影响，中国企业需要更成熟的迁移工具和服务来实现无缝迁移。
- 由于传统软硬件存在深度捆绑的技术债务，将数据库从旧系统迁移到现代系统的工作量和风险仍然很高。大规模迁移通常需要一年以上的時間。
- 出于对数据安全、数据主权和公有云供应商锁定等问题的担忧，金融、政府和公共服务等企业机构更倾向于将关键数据库升级至混合云环境。本地数据库产品的混合云能力有待改进，难以充分发挥现代数据库的优势。

## 使用建议

- 针对弹性、灵活性、成本效益或合规性等方面的价值偏好，评估最适合自身用例的部署模式（公有云/本地/混合）。
- 观察中国云数据生态系统兴起的方向。从行业、主要业务区域、企业规模和业务组合的角度，选择最适合企业的生态系统。
- 从产品功能和售后服务等级协议的角度评估本地数据库供应商，不仅要评估其产品能力，还要评估其对本地其他技术供应商生态系统的开放性（如云平台、硬件、业务应用和数据安全）。
- 利用Gartner对云数据库管理系统的研究和相关的需求建议书工具包，评估中国的新兴数据库能力（例如分布式交易数据库、湖仓一体和增强交易）。了解每种能力的优缺点，并且与业务利益相关者沟通，找准合适的用例。

## 厂商示例

阿里云、华为云、OceanBase数据库、PingCAP、巨杉数据库、腾讯云、星环科技、武汉达梦数据库



## Gartner相关推荐阅读

[Market Guide for DBMS, China](#)

[中国数据库管理系统市场指南](#)

[Magic Quadrant for Cloud Database Management Systems](#)

[Exploiting the Evolving Database Management System Trends in China](#)

## 生成式AI

分析师: Ben Yan, Mike Fang, Tong Zhang, Tracy Tsai

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为1%~5%

成熟度: 发展阶段

### 定义:

生成式人工智能（生成式AI）技术可以从大型原始内容存储库的资料中学习，从而生成内容、策略、设计和方法的新衍生版本。该技术可产生深远的业务影响，范围包括：内容发现、创建、真实性验证和监管，人工任务自动化，改进客户和员工体验等等。

### 为何重要

随着Stable Diffusion、Midjourney、OpenAI（ChatGPT）等各类厂商的模型和生成式AI应用的普及，业界对生成式AI的探索速度与日俱增。多数行业的终端用户企业机构都在积极尝试这一技术。中国的技术厂商，如百度、阿里巴巴和商汤科技，都把提供生成式AI支持的应用和工具作为优先任务。2023年，利用生成式AI进行创新的初创公司纷纷涌现，Gartner预计这一趋势将会继续发展。包括中国在内的一些政府已开始评估生成式AI的影响，并制定相关法规。

### 业务影响

在未来12个月内，领先的技术产品和服务将采用生成式AI功能，采用会话式技术进行创建和沟通，从而实现生成式AI技术的全民化。生成式AI将在垂直行业、科学发现和技术商业化方面取得快速进展。令人遗憾的是，该技术如果被用于不法目的，也将成为安全和社会的威胁。负责任的AI、信任和安全是安全利用生成式AI的必要条件。

## 推动因素

- ChatGPT的流行使围绕生成式AI的炒作愈演愈烈，将其推升至技术成熟度曲线的顶峰。这项技术获得了高管的大量关注，激发了更广泛地采用生成式AI的热情。
- 基础模型在很大程度上提高了AI生成内容的质量，并且加快了包括语言、代码、图像、语音和多模式数据在内的各种工件的生成速度。
- 采用零/少量样本学习的提示工程一直在快速改进生成式建模，同时减少对训练数据和微调的需求。
- 机器学习（ML）和自然语言处理（NLP）平台在添加生成式AI和迁移学习功能，以实现生成式模型的复用性，使AI团队能够定制和访问生成式模型。
- 与其他模型或应用（如视觉生成模型、科学计算模型或Web应用）集成的生成式AI模型，可解锁更多潜在的业务用例。
- 更重要的是，AI满足了行业提高生产力的紧迫要求，这是任何经济体增长过程中都必不可少的。很少有其他技术能提供这样一个重塑工作方式的机会。
- 合成数据协助补充稀缺数据、减轻偏见和保护数据隐私的功能，吸引了企业的注意。将此用于医疗行业，可以提高脑肿瘤手术的准确性。

## 阻碍因素

- 生成式AI的基础模型还不够成熟。完全依赖基础模型结果具有风险，因为模型可能会产生错误的结果，需要人工介入（human in the loop）。
- 生成式AI可能被用于不法目的，其产生的深度伪造内容会在政治、商业和社会等各个方面造成危险。对生成的内容进行全面、准确的检测在未来几年仍将是一个挑战，而且可能无法完全实现。
- 碎片化和专业化的技术产品（例如仅生成图像或仅生成文本）涉及一系列工具，目前尚未形成单一解决方案。
- 训练大模型需要消耗巨量的计算资源，对多数企业而言难以承受。大多数企业可以利用现有模型，但无法开发自己的模型。
- 随着技术相关法规的推出，生成式AI供应商需要调整其策略。
- 中国的国产生成式AI模型与ChatGPT/GPT-4相比，在性能上仍存在差距。企业机构如希望采用具备ChatGPT/GTP-4性能的本土替代方案，需要等待或利用其他专业解决方案。

## 使用建议

- 研究行业或部门可从生成式AI中获益的方式，从而确定利用直接购买的功能或与研究人员合作可实现的初始用例。
- 探索如何利用合成数据来缩短开发周期、减轻监管担忧、缓解数据偏见、促进数据变现并降低数据获取成本——尤其是在缺乏罕见事件数据的情况下。
- 审查并量化对于生成式AI优势和局限性。首先将其用于改进现有流程，而在生成式AI可以带来突破的领域则需要先制定指导和策略，因为这些突破需要技能、资金和严谨的决策。需要在技术能力与伦理因素之间进行权衡。
- 为减轻深度伪造影响、规避其可能导致的严重风险做好准备。缓解风险的方法，例如应用算法检测和内容来源验证，仍在不断发展。因此，为了打击深度伪造，相关的技术、制度和政府干预是必要的。
- Gartner预计生成式AI技术很快就会得到采用，因此需要对这些技术保持密切关注。

## 厂商示例

aiXcoder、阿里云、亚马逊云科技、百度、华为、微软、商汤科技、腾讯、智谱AI

## Gartner相关推荐阅读

[Innovation Insight for Generative AI](#)

[Top Strategic Technology Trends for 2022: Generative AI](#)

[Predicts 2022: Generative AI Is Poised to Revolutionize Digital Product Development](#)

[Quick Answer: China Perspective — Frequently Asked Questions on ChatGPT and Large Language Models](#)

## 数据素养

分析师: Fay Fei, Mike Fang

影响力评级: 较高

市场渗透率: 目标受众覆盖率为20%~50%

成熟度: 主流采用起步阶段

## 定义:

数据素养是在特定场景下解读、应用和交流数据，同时了解数据源及其结构、分析方法以及采用的相关技术的能力。具备数据素养的个人能够识别、理解、解读业务环境中的数据并采取相应行动，从而对由此产生的业务价值或成果产生影响。

## 为何重要

数字体验渗透于人们个人和职业生活的方方面面，因此数字化转型在国家和行业层面都变得越来越重要。作为一种理解、解读和处理数据的能力，数据素养已成为数字经济和社会的基础，可帮助利益相关者：

- 提升知识型员工的业务敏锐度
- 从与特定业务用例相关的数据中获取洞察
- 建立对数据的信任，提升决策自信

## 业务影响

为了利用数据获取竞争优势，并呼应 [中国政府的顶层数字战略](#)，企业必须向数据驱动型发展。要完成这一转型，企业需要进行明晰、持久的组织变革，在影响业务成败的关键领域促进和协调文化变革。成功的关键就是通过加强数据素养和数据驱动型文化，获得为员工提供指导的能力。

## 推动因素

- 中国政府已展示出推动数字经济和数字业务的强烈意愿。近年来，政府发布了大量鼓励数字化转型的政策和计划，提高了社会的整体数据意识。
- 员工的数据素养在企业机构整体数字灵活性中的作用正得到越来越多的认可。在中国，数据素养水平较高的行业已呈现出大幅业务增长。
- 数据和分析（D&A）职能已发生变化。如今，它已发展为企业机构业务模式和数字平台的核心，使每个员工都成为信息工作者。因此，数据和分析在业务中的使用比以往更为广泛。
- 多年来，数据和分析的价值一直局限于可视化领域。但企业机构现在开始意识到，数据素养的提升将推动价值导向型D&A用例的增加。
- 很多中国人，尤其是年轻一代，自然而然地对数字化持欢迎态度，并且热衷于培养数据素养。

## 阻碍因素

- 各行业缺乏通用的数据素养模型、框架、标准和术语。
- “数据素养水平”和“业务优化”之间的关联较难梳理清晰，目前仍然很难对数据素养水平提升带来的业务价值进行明确量化。
- 企业机构在提高数据素养方面缺乏指导和实践，培训和认证方法零散且缺乏一致性。
- 数据和分析领域的人才和技能短缺，导致当前员工队伍的数据素养较差，甚至连着手培养都不易。
- 数据素养未得到高管层的足够重视。因此，企业机构仍然缺乏在战略和项目中应对文化和数据素养挑战的举措。
- 由于提升全体员工的技能并非易事，数据素养的全面推广仍需数年。

## 使用建议

- 梳理 [中国的数字政策](#)，并与企业内部数据素养计划相协调，以便在不断变化的环境中获得竞争优势。
- 吸引利益相关者的参与，提高数据素养意识，并创建关于数据素养预期业务成果的强大共同愿景，特别是在数据驱动创新方面。
- 与人力资源和业务领导者合作，将数据素养融入到员工价值主张中。列出“高水平”和“低水平”数据素养的实例，强化期待发生的行为。
- 评估不同职位和角色所需的数据素养水平、学习目标和成果，并将数据素养纳入职业路径和发展规划中。
- 不局限于厂商的产品培训，根据企业机构所处行业、业务领域和角色来调整培训课程。综合利用多种培训方法（课堂、线上、社群、在职）来提升整体学习效果。
- 从D&A着手快速建立数据素养的知识基础，但为数据素养提升的不同阶段设定合理的预期。持续变革需要时间，因为这需要人们学习新的技能，改变行为方式。

## 厂商示例

Coursera、和鲸科技、爱课程、慕课网、Udacity、Udemy、学堂在线

## Gartner相关推荐阅读

[How CDAOs Must Lead Data Literacy and Data-Driven Culture](#)

Address Both 'Skill' and 'Will' to Deliver Data-Driven Business Change

Drive Business Outcomes by Measuring the Value of Data Literacy

Tackle Data Literacy Head-On to Avoid Data and Analytics Program Failure

知识图谱

分析师: Tong Zhang, Mike Fang, Ben Yan

影响力评级: 较高

市场渗透率: 目标受众覆盖率为5%~20%

成熟度: 成型阶段

定义:

知识图谱是一种机器可读的数据结构，通过图形数据模型——即由节点（顶点）和边（或弧线）组成的关系网络，呈现物理世界和数字世界中各个实体（人、公司、数字资产）的概念及其相互关系。

为何重要

与其他类型的数据模型相比，知识图谱对信息的展现更为直观，更便于理解、操作和使用。亚马逊、Facebook和谷歌等技术公司的许多产品和服务都以图数据作为支柱，因为可以直接从数据源中对不同的数据进行编码和关联。知识图谱可支持协作和共享、搜索和发现，以及通过分析提取见解。

业务影响

知识图谱可以在以下多个应用场景提升业务影响：

- 调查分析（例如执法、网络安全或金融交易）
- 数字商务（例如产品信息管理和推荐）
- 数据管理（例如元数据管理、数据目录和数据编织）
- 展示人类知识和商业洞察，作为一种组件，供多种机器学习和人工智能模型（包括大语言模型）使用。

## 推动因素

- 由于数字化和全球化举措层出不穷、变化多端，静态、孤立的方法需要被具有自适应性的一体化方法取代，而知识图谱正是提供了这种方法。
- 知识图谱在消费产品和服务中的应用越来越多，比如智能设备和语音助手、聊天机器人、搜索引擎、建议引擎和路线规划等。
- Web3应用的新兴格局和对跨信任网络访问数据的需求，催生了去中心化知识图谱，以创建不可更改但可查询的数据结构。
- 图数据库管理系统技术的改进，可以大规模处理图数据结构的存储和操作。比如平台即服务（PaaS）产品可以降低硬件和基础设施配置及优化的复杂性。
- 追求通过使用具有可关联性和可管理性的标准化元数据，更好地利用文档、图像和视频中保存的非结构化数据。
- 越来越多的数据孤岛需要管理，而其中的数据通常是重复的且没有明确的定义、用途和使用方式。
- 图算法和机器学习被用于识别复杂网络中的影响者、客户群、欺诈活动和关键瓶颈。
- 大语言模型的提示语工程和模型精调需要知识图谱的协助。

## 阻碍因素

- 人们认识到知识图谱可以用于更多的场景，但其商业价值在实施初期很难体现出来。
- 将知识图谱模型从原型发展到生产应用，需要工程化和系统性集成的专业知识。随着知识图谱的规模不断增加，需要对其进行维护，以确保性能稳定、能够处理重复数据和保持数据质量，而维护方法仍旧不够成熟。
- 碎片化的图数据库管理系统市场——多种不同的图数据模型（资源描述框架[RDF]或属性）、实施架构（原生或多模态）和最佳工作负载，使使用者感到困惑和犹疑。
- 企业内部数据与外部知识图谱的融合面临挑战，无法摄取、验证和共享本体以及有关实体（比如位置、人和时间）的数据。

## 使用建议

- 通过试点项目来确定需要定制化知识图谱的使用场景。这不仅能为业务带来实际价值，也是数据和分析人员学习的机会。
- 采取敏捷方法开发知识图谱，缩短价值实现时间。对构建知识图所需的结构化和非结构化数据进行评估，创建适用于业务场景的最小可行子集。
- 利用供应商和服务提供商的专业经验，为企业负责人部署和验证知识图谱验证的价值。
- 将知识图谱纳入数据和分析治理和管理的范围内。防止数据孤岛的存在，研究如何打通多个知识图谱并将其扩展到数据编制架构。

## 厂商示例

阿里云、海致星图、腾讯云、星环科技、杭州悦数科技、创邻科技

## Gartner相关推荐阅读

[Top Trends in Data and Analytics, 2022](#)

[Graph Steps Onto the Main Stage of Data and Analytics: A Gartner Trend Insight Report](#)

[Working With Graph Data Stores](#)

[Quick Answer: What Is Data Fabric Design? AI Design Patterns for Knowledge Graphs and Generative AI](#)

## 联邦机器学习

分析师: Ben Yan, Mike Fang

影响力评级: 较高

市场渗透率: 目标受众覆盖率低于1%

成熟度: 发展阶段

## 定义:

联邦机器学习的目标是利用不同本地节点中的多个数据集来训练机器学习（ML）算法，而无需明确共享数据样本。联邦机器学习有助于保护隐私，帮助机器学习尤其是深度神经网络（DNN）使用更多数据，解决数据传输瓶颈，并赋能协同学习以提高模型准确性。



## 为何重要

2021年，中国出台了《数据安全法》和《个人信息保护法》等法规，而公众对于数据和隐私的保护意识也在逐渐提高。联邦机器学习（FedML）是在不泄露敏感业务信息的情况下，在去中心化环境中训练ML算法的一项重要创新。FedML通过在智能手机、软件机器人、自动驾驶汽车或物联网边缘设备中进行本地学习，实现更个性化的体验，并帮助企业机构跨越数据孤岛构建协同学习模型。

## 业务影响

FedML在本地存储数据的同时，通过中心化共享本地模型参数更新来支持协同机器学习，尤其有利于物联网、网络安全、隐私、数据变现和受监管行业的数据共享。对于中国的企业和公共部门来说，将个人、企业机构和行业数据用于为社会提供价值是一个颇具吸引力的概念。

## 推动因素

- 隐私法规的密集出台，要求在机器学习中对本地数据进行保护。
- 随着边缘人工智能的热度升高，数据将分散于多个异构边缘设备和云中。FedML能帮助企业机构妥善管理数据。
- 数据量持续快速增长，集中收集和存储大数据变得更加困难，这在物联网场景中尤为明显。传感器数据是在设备上收集的，通常没有时间或理由将其集中收集。由于存在可扩展性问题、功耗过高、连接和时延问题，数据的收集逐渐以联邦计算架构的形式转移到边缘基础设施。
- 供应链上下游合作伙伴需要FedML在不同数据孤岛之间统一利用数据。
- 随着大语言模型（LLM）的发展和联邦LLM的研究的出现，一些企业机构可以协同训练LLM。
- 在区块链技术的支持下，集群（联邦）学习将边缘计算、对等网络与协同相结合，正在成为去中心化机器学习中一种颇具前景的方法。
- FedML通常与其他隐私增强计算技术相结合，形成完整的安全计算解决方案。

## 阻碍因素

- 要打造协同学习模型，需要花时间在企业机构之间建立信任。
- 需要让各方参与进来并商定激励机制，以保持参与者的积极性，使FedML团队保持长期稳定。
- 系统和数据的异构性要求系统之间进行大量协调和标准化，以充分发挥作用。
- 实现FedML需要完整的端到端基础设施堆栈，集成DataOps、ModelOps、部署和持续跟踪/再训练功能，并且需要较高的实施成熟度。
- 由于本地的学习主体类型多样或存在重叠，其数据可能很难评估，并且差异巨大，因此通过本地模型参数更新来创建更准确且无偏见的新中心化模型并非易事。
- 解决安全和隐私验证问题需要额外的安全相关技术。

## 使用建议

- 在创建和维护去中心化智能服务或产品时应用FedML，同时保护用户隐私，避免集中收集大量数据。
- 与上下游合作伙伴共同探索FedML用例，寻找机会提高整体运营效率。
- 部署通用的、集中预训练的模型，率先采用去中心化机器学习应用。同时，基于本地数据和反馈对模型进行本地再训练，以实现个性化和情境化。
- 通过协同学习持续改进去中心化机器学习应用。不断收集本地模型参数更新情况，创建卓越的中心化新模型，然后重新部署新模型，并以去中心化的方式进行使用和微调。

## 厂商示例

第四范式、蚂蚁集团、百度、翼方健数、铭崴科技、腾讯、微众银行

## Gartner相关推荐阅读

[Innovation Insight for Federated Machine Learning](#)

[Quick Answer: Why Is Federated Learning Prominent in China?](#)

[Top 3 Priorities for Chinese Enterprises to Promote Effective Data Sharing](#)

## 低谷期技术

### 数据治理

分析师: Tong Zhang, Julian Sun

影响力评级: 较高

市场渗透率: 目标受众覆盖率为20%~50%

成熟度: 成型阶段

#### 定义:

数据治理是对决策权的规范，也构成了问责制框架，用以确保在对数据和分析进行评估、创建、使用和控制时采取适当的行为，涉及对象包括管理、保护和利用数据资产过程中所需的人员、流程和技术。

#### 为何重要

数据治理为数据资产设立了护栏。这是一种收集、管理、保护和存储数据的综合方法，由适当的人员通过适当的流程使用相关技术，确保数据产生价值。中国企业机构正在快速实施以实现业务价值为目标的数据和分析（D&A）项目，却未能有效应对数据治理风险，对企业机构的数据和分析以及数据驱动转型战略的扩展造成了阻碍。

#### 业务影响

- 业务决策和数据驱动洞察的质量会受到数据准确性、及时性和可靠性的影响。数据治理有助于企业机构避免因数据质量低劣而产生的返工、数据清洗和数据集成等成本。
- 有效的数据治理有助于建立利益相关者（包括客户、投资者和员工）的信任和信心，使其相信企业机构在负责任地管理数据。
- 在中国，数据被定义为一个生产要素。数字经济要求所有企业机构都拥有良好的治理护栏和框架。

#### 推动因素

- 遵守《中华人民共和国网络安全法》、《中华人民共和国数据安全法》和《中华人民共和国个人信息保护法》等法规是实施数据治理的关键驱动力，目的是确保数据得到妥善的管理、保护、审计和使用。
- 数据作为一种新的生产要素，已成为企业的一项关键差异化特征，可为企业提供竞争优势。从数据中提取业务价值依赖于良好的数据质量以及由数据治理赋能的各种数据的多样性。

- 企业依靠数据治理来避免数据泄露、数据丢失、未经授权的访问等风险，以及数据管理不善产生的成本。

## 阻碍因素

- 难以让业务伙伴认识到，数据是一个业务问题，而非IT问题。业务合作伙伴虽然理论上认可数据质量、一致性和访问的重要性，但还是会认为数据是“IT的责任”，不愿对数据决策承担起直接责任。
- 难以定义和采用一致的数据治理流程和策略。如果不设立一个同时包含业务部门和IT部门代表，并且限定其目标范围的正式治理实体，治理流程将无法开展，收益也会减少。在中国，相对不够精简和清晰的业务流程放大了这一问题。
- 未能就常见的企业数据定义达成共识。业务部门负责人倾向于认为自身的数据需求是独特的，而不是与企业其他部门的数据需求保持一致。
- 不同项目的数据处理方法不一致。整个企业的数据处理方法不一致，使项目解决方案产出面临相互矛盾的风险，且无法实现可扩展性。由于需要为每个项目重新设计任务和文档，项目的进度因此放慢。缺乏成熟的数据治理供应商加剧了这一问题。
- 难以定义和维持通往目标数据能力状态的路径。实现数据的高质量、准确和一致是一个宏大而模糊的目标。企业机构很难确定数据能力的实际目标状态，以及近期和中期要达到的过渡阶段里程碑。
- 监管政策变化的不确定性，使数据治理更倾向于以合规为驱动，而非以价值为驱动。数据治理的目标是为业务带来价值，而不是规避风险，业务部门需要在对风险进行管理的同时创造价值。

## 使用建议

- 明确数据治理任务和目标的范围。明确目标有助于将稀缺资源用于最重要的治理活动。
- 根据数据资产的差异，实施不同程度的治理。优先考虑业务价值最高、在企业中用途最广泛的数据资产。
- 促进业务部门承担数据管家的职责和任务。数据是一项企业资产，需要业务负责人据此对其进行管理，因此要确保管委会中有业务部门的代表。
- 在开展数据标准化工作时要善于抓住有利时机。寻求业务环境中可引发新数据服务要求和需求的事件（如合并或高管新上任）带来的数据标准化和集成机会。

- 根据数据对于企业机构重要性所存在的质量差距，确定数据质量改进的优先级。从多个用户的角度和场景出发，找到关键数据的数据质量问题，确保有针对性地改进数据质量。
- 研究国家数据局法规和最新的中国数据政策。探索新的机会是成功实施数据治理的关键，而新数据用例的发展方向则由国家数据局确定。

## 厂商示例

阿里巴巴、东方国信、数语科技、亿信华辰、华为、精鲲

## Gartner相关推荐阅读

[Playbook: Building a Modern Data Governance Program](#)

[Distilling Data Governance Essentials for FP&A Leaders](#)

[Market Guide for Data and Analytics Governance Platforms](#)

[Data and Analytics Governance as a Business Capability: A Gartner Trend Insight Report](#)

[3 Ways to Promote Your Data Agenda at the Center of the Chinese Digital Economy](#)

## 数据中台

分析师: Fay Fei, Xingyu Gu

影响力评级: 中等

市场渗透率: 目标受众覆盖率为5%~20%

成熟度: 主流采用起步阶段

## 定义:

数据中台（DMO）是一种组织战略和技术的实践。通过数据中台，不同业务线的用户能够依据单一事实源，高效地使用企业数据进行决策。创建数据中台可以被视为企业构建可组装和可复用的数据和分析（D&A）能力的一种方式。这些能力可以提供独特的数字运营，并通过技术栈将数字运营贯穿到整条价值链中。

## 为何重要

很多中国企业之所以采用数据中台实践，是为了减少其数据和分析架构的技术冗余，打通不同系统的数据孤岛，并推动可复用的数据和分析能力。

## 业务影响

数据中台通过以下方式优化数据的使用并消除D&A孤岛：

- 促使首席信息官（CIO）和D&A领导者重新思考如何利用组装式架构来建立业务驱动型D&A框架，使业务用户得以灵活、快速性地访问和利用企业的信息资源。
- 引导企业机构利用和丰富现有资产，而不是重建所有资产。
- 持续构建可复用数据分析能力，将企业机构的D&A举措由以项目为导向转变为以平台为导向。

## 推动因素

- 数据中台在一些行业和企业机构中得到成功应用的案例广为流传，使其成为各行各业数字化转型的关键。
- 本土供应商的能力不断增强，特别是在基于开源技术的元数据管理、数据集成和数据准备方面，使数据中台的实际部署实施得以改善。
- 数据中台已被供应商和市场宣传为最先进、最全面的D&A平台，因此对所有企业机构都至关重要。

## 阻碍因素

- 数据中台在许多情况下未能兑现其组装式敏捷D&A能力的承诺，因此在市场中的地位被削弱。非数字原生企业机构并未为在自身业务环境中构建可复用的D&A能力做好准备。
- 数据中台的目标和价值主张往往不清晰，导致项目范围不够明确。业务用户低估了数据中台部署所面临的挑战以及所需的资源，因此，在随后的实施过程中会面临巨大的困难，使数据中台成为一个看似无法实现的目标。
- 对于数据管理，数据中台采用的是将数据集中起来的“收集”实践，但缺乏将不同系统数据连接起来的“打通”实践，导致价值实现时间较长，无法对快节奏的变化做出反应。
- 传统的项目交付方法并不关注长期业务成果，而是关注项目的上线时间，因此无法留住用户，也无法创造可持续价值。
- 在中国，多数企业机构缺乏数据素养，无法吸引更多的公民数据用户使用这些工具。

## 使用建议

- 将重点从“数据中台”这个术语本身转向该术语所蕴含的数据和分析核心能力的培养上。
- 创建长期的数据素养项目。这对于数据中台发挥“力量倍增器”功能至关重要。
- 将组装式D&A能力（如对封装业务能力和数据编织的分析）作为数据中台愿景和架构的指导原则。
- 创建混合的组织模式，使中心化团队能够与去中心化的业务团队合作，共同对业务进行探索，从而确定数据和分析能力的可复用性。

## 厂商示例

滴普科技、袋鼠云、科杰科技、明略科技、数澜科技、奇点云

## Gartner相关推荐阅读

[Demystify Data Middle Office by Nurturing Core D&A Capabilities](#)

[Video: Demystifying the Data Middle Office](#)

[From Logical Data Warehouse to Data Fabric](#)

## Data and Analytics Essentials: Architect an Analytics Platform

### Market Guide for Analytics Platforms, China

#### 公民数据科学

分析师: Fay Fei, Julian Sun

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为5%~20%

成熟度: 成型阶段

#### 定义:

公民数据科学，指非专家员工或非数据和分析（D&A）领域员工所掌握的一系列分析洞察能力。公民数据科学家是企业机构内的一种角色，而非头衔或职位。

#### 为何重要

- 公民数据科学家可以创建和交付分析模型，提供业务角度洞察，从而提高数据和分析对企业机构的影响。
- 公民数据科学家具备与学术或专业数据科学家不同的职能知识，可提供不同角度的经验，改善解决方案的效率、效果和深度。随着新兴技术的出现，公民数据科学家通常可以在描述性和诊断性洞察之外有新的发现。
- 中国市场人才短缺，专业数据科学家的成本很高。

#### 业务影响

公民数据科学家可通过恰当的数据科学和机器学习（DSML）工具和培训，将其行业知识融入分析生命周期（例如特征生成和选择，以及算法选择）。公民数据科学家发现的业务导向分析模型或洞察，通常有助于业务决策更好地实现预期成果。

#### 推动因素

公民数据科学最重要的驱动因素包括：



- 业务团队对数据科学人才的需求日益增加：在中国，相关人才仍然供不应求，而公民数据科学家有助于填补一部分人才缺口。过去，构建DSML模型需要专业数据科学家，聘用和留住这些科学家的难度和成本非常高，而公民数据科学则有助于克服这个问题。
- 数据科学需要深入的职能知识：公民数据科学家对业务领域有深入的了解，其职能知识、数据科学技能和技术的组合有助于实现成果。
- 供应商产品成熟度日益提高：供应商已认识到这一新增人群中有很多目标客户。因此，许多供应商产品推出了专供公民数据科学家使用的工具和功能。
- 增强分析模糊了商业智能和数据科学之间的界限：例如自动化数据访问和数据工程；通过自动数据可视化和探索实现的增强型用户洞察；建模和模式检测，包括特征工程、模型选择和验证；自动部署和运营，以及协作和共享支持能力。

## 阻碍因素

- 通过公民数据科学创造价值的一个关键，是进一步掌握DSML先进技术和方法。考虑到这一角色同时所需的行业知识，公民数据科学家在中国企业机构中仍然不常见。
- 在许多场景中，尤其是新项目中，仍然需要由专业数据科学家（以统计方式）验证公民数据科学家得出的结果。
- 很多人认为公民数据科学只是一个初始步骤，而不是功能齐全的方法。中国企业机构的领导者往往低估了这一领域的作用，毕竟相应角色很难匹配，而且赋能和教育成本很高。
- 中国市场缺少针对公民数据科学工作模式和原则的标准。公民数据科学家往往独立工作，不受DSML专家和其他利益相关者的监督，也不与之合作。

## 使用建议

- 成功始于领导力：帮助业务领导者了解扩大交付能力范围和规模的潜在影响。与领导层合作，寻找公民数据科学机会，对现有的分析项目和专业数据科学项目进行补充。
- 有吸引力和包容的环境：创建实践社区并提供培训和工具，营造欢迎和支持所有人的环境，以探索公民数据科学家角色的价值。
- 专家数据科学家价值：承认您可能仍然需要专业数据科学家来验证和操作模型、洞察和应用。
- 工具和技术：提供增强型分析工具（包括但不限于增强型DSML学习工具）、平台和流程，以支持并鼓励业务用户、应用开发人员和数据科学团队之间的协作。及时了解现有的商业智能（BI）和数据科学平台，追踪新兴初创公司的能力（技术）和路线图，为增强型能力提供更好的支持。

## 厂商示例

第四范式、百度智能云、九章云极、和鲸科技、星环科技、Profet AI、神策数据

## Gartner相关推荐阅读

[Maximize the Value of Your Data Science Efforts by Empowering Citizen Data Scientists](#)

[Build a Comprehensive Ecosystem for Citizen Data Scientists to Drive Impactful Analytics](#)

[The State of Data Science and Machine Learning](#)

[Market Guide for Multipersona Data Science and Machine Learning Platforms](#)

[Best Practices to Enable Effective Citizen Data Science](#)

## 复苏期技术

### 实时数据管理

分析师: Xingyu Gu

影响力评级: 颠覆

市场渗透率: 目标受众覆盖率为20%~50%

成熟度: 主流采用起步阶段

#### 定义:

实时数据管理是为了对外部机会或威胁等环境变化作出自动化的实时感知或准实时响应的一组数据分析技术。其组成部分包括具备实时功能的数据库管理系统，以及基于事件流处理（ESP）和变化数据捕获（CDC）的数据集成技术，从而为现代的实时数字化场景提供支持。

#### 为何重要

实时数据管理是一组数据管理技术，为需要实时或近实时响应外部变化的现代数字化场景提供支持。主要场景分为两类：一是嵌入业务流程中的增强交易，二是近实时的人工响应。由于数据库管理系统（DBMS）和数据集成（DI）技术日趋成熟，这两类场景都已进入早期主流采用阶段，最终将被应用于各大企业组织的不同部门中。

#### 业务影响

实时数据管理在中国大多数变革型数字化场景和创新中发挥着至关重要的作用，包括元宇宙、智能制造、智慧城市、自动驾驶、量化计算等，其中很多用例已是成熟的概念，并经过了多年的实践探索。随着实时数据管理技术的日趋成熟，这些场景将进一步发展，并最终成为未来十年中国社会的新常态。

#### 推动因素

推动实时数据管理发展的五个因素：

- 中国社会基础设施的快速发展和数字生活方式的普及，使得传感器、摄像头、数字控制系统、公司网站、交易应用、社交计算平台、新闻和天气推送、数据中介、政府机构和业务合作伙伴产生的数据量不断增加。
- 实时数据管理技术——例如时序数据库（TSDB）、内存数据库、事件流处理（ESP）和变化数据捕获（CDC）——正在迅速发展，这使得企业机构能够利用高效架构处理上述实时数据。

- 中国商业环境竞争激烈，本地业务领导者需要更多实时的持续型智能，提高环境感知能力，以做出更快、更精准和更细致的决策。
- 开源技术让更多供应商可以降低实时数据管理功能的价格。同时，部署具有这些功能的云产品也变得更加容易。
- 中国供应商正在提供易于使用的开发接口、预定义的业务功能和行业服务，以实现更快的部署。超级用户可使用低代码技术和现成模板构建实时应用。

#### 阻碍因素

- 本地数据生态系统成熟度较低，导致企业机构需要开展大量集成工作，将实时数据管理技术嵌入到现有数据库和数据中台中。对于同时需要实时和历史数据的复杂用例，为了确保性能和数据的一致性，实施工作量仍旧很大。
- 实时数据管理技术中的编程语言种类繁多，导致技术门槛高、学习难度大。为解决这一问题，部分本地供应商已开始构建集成平台即服务（iPaaS）或数据库内置机器学习（ML）功能。但这些解决方案仍然处于新兴阶段。
- 支持近实时的架构性能提升，为用户提供了更多可选产品。然而，这些“足够好”的产品使客户评估供应商处理实时场景的能力变得更加复杂。
- 许多架构师和软件工程师仍然对使用实时数据的设计技术不甚了解。他们更熟悉如何处理数据库和其他数据存储中的静态数据，因此会默认使用此类技术，除非业务需求迫使他们使用实时数据。

#### 使用建议

- 向业务领导者介绍实时数据管理所支持的用例及其重要性。与领导者集思广益，明确具体机会，重新思考业务流程，并创建传统架构无法部署的应用。
- 对于需要企业级支持和全套功能的主流应用，使用供应商支持的闭源平台，或带有商业化版本的开核产品。只有当开发人员熟悉开源软件，且许可费比员工成本更重要时，才使用社区支持的免费开源技术。
- 选择经过优化的产品，以便与现有主流数据库和数据中台进行流数据集成。

#### 厂商示例

DataPipeline、智舆科技、KaiwuDB、柏睿、飞轮科技、顺丰科技、涛思数据

#### Gartner相关推荐阅读

[Market Guide for Event Stream Processing](#)

## 5 Essential Practices for Real-Time Analytics

### Time Series Database Architectures and Use Cases

### Exploiting the Evolving Database Management System Trends in China

#### 自助分析

分析师: Fay Fei, Julian Sun

影响力评级: 中等

市场渗透率: 目标受众覆盖率为20%~50%

成熟度: 主流采用起步阶段

#### 定义:

自助分析（SSA）技术和流程可以让业务人员在不需要IT参与或支持的情况下，自主完成数据准备和数据可视化、执行查询和生成报表。SSA的功能通常包括低代码/无代码工具，其作用可通过AI持续增强。虽然这些工具提供了复杂的数据准备和分析功能，但也提供了易于理解的数据，简化了数据访问。

#### 为何重要

对于数据和分析（D&A）项目而言，SSA有助于业务用户轻松创建分析原型，而且相较于完全依赖IT，能更快地开始试点。中国企业机构希望缩短获得分析内容和洞察的时间。与传统的中心化方法不同，SSA为创建数据和分析内容提供了一种去中心化的方式，具有自主性和灵活性，可快速适应不断变化的环境。

#### 业务影响

自助式D&A对于扩大数据驱动型决策的影响至关重要。很多中心化的D&A功能难以满足去中心化团队对数据和洞察的需求。新兴公民分析师或公民数据科学家角色在对于数据的业务背景有充分了解的基础上，能够使用强大的无代码/低代码数据准备和分析平台来快速挖掘洞察。

#### 推动因素

- 中国的分析服务厂商正在努力将自助服务能力融入其产品中。随着企业报表工具发展放缓，供应商开始投资自助服务能力，并将其作为向客户宣传的重点。

- 人工智能（AI）、机器学习（ML）和其他增强技术降低了运营分析的门槛。经过基础训练后，不具备技术技能的用户也能够自主创建分析内容。
- 生成式AI最近成为热点，其空前的能力迅速受到分析市场的关注，将成为用户转变数据交互和数据分析方式的助推力，从而使每个人都可以产出自己的分析洞察。
- 业务用户提出了更高的信息探索需求，期望将自助服务的使用扩展到以分析为目的的数据准备和数据整理之中，使更多业务用户能够自主进行端到端的分析。
- 购买D&A工具的预算越来越多地来自业务部门，不再仅仅是来自中心化IT/数据团队。

## 阻碍因素

- 中国企业机构往往缺乏正式的战略和规则来管理流程和宣传自助式D&A洞察。自助服务的权利义务界定不够明晰，增加了人们的担忧，削弱了自助服务的吸引力。
- SSA治理是一个普遍关注的问题。在SSA模型中，分析内容的指数级增长给IT治理带来了巨大的负担。企业机构需要在敏捷性和可控性之间取得平衡。
- SSA要求业务用户具备一定水平的数据素养，可以理解和解释数据，并操控数据准备流程。而多数中国企业机构的整体数据素养水平阻碍了SSA实践的普及。
- 高质量数据的获取对于许多企业机构来说仍然是一个难题。尽管有功能强大的工具，但糟糕的数据质量更容易导致理解偏差、数据滥用和错误决策。

## 使用建议

- 根据用户使用自助服务的能力和意愿对其进行细分，优先为就绪度最高的用户提供自助服务。成功一次就会推动未来更多次的成功，并逐渐提高D&A的成熟度。
- 评估数据目录和自助式数据管理功能，使业务用户能够在数据视图添加特定的或外部的资源。
- 为自助式D&A创建者和消费者打造社区。自助服务不应只是便利个人使用。随着分析使用群体的发展，能够实现共享、协作、教育、项目综述和成功宣传的社区至关重要。
- 建立数据素养和认证计划，确保用户能够从自助服务中获得增值，避免提供低质量或孤立的信息。

## 厂商示例

帆软软件、观远数据、衡石科技、网易数帆、思迈特软件、永洪科技

## Gartner相关推荐阅读

[Data and Analytics Essentials: 3 Steps to Implement Self-Service Analytics](#)

[Quick Answer: 4 Easy Ways to Promote Trust in Self-Service Analytics](#)

[Magic Quadrant for Analytics and Business Intelligence Platforms](#)

[Market Guide for Analytics Platforms, China](#)

[Critical Capabilities for Analytics and Business Intelligence Platforms](#)

## 数据分类

分析师: Anson Chen

影响力评级: 较高

市场渗透率: 目标受众覆盖率为5%~20%

成熟度: 主流采用起步阶段

## 定义:

数据分类是根据约定的编目方法、分类法或本体论来组织信息资产的过程，包括给数据对象进行标记或打上标签，以提高数据使用和治理的便利度，也包括在数据生命周期中采取控制措施或使用数据编织来激活元数据。通常，数据分类会形成一个大型存储库，内含大量有用的元数据，供决策者使用，以做出知情决策。

## 为何重要

数据分类有助于在涉及价值、访问、隐私、存储、伦理、质量和留存的数据治理和数据安全项目中，采用有效且高效的方式确定数据优先级。根据 [中国的数据监管要求](#)，数据分类应成为安全、数据治理和合规项目中的重要一环，用于帮助企业机构区分数据的敏感度，并提高数据保护控制措施的有效性。

## 业务影响

数据分类增强了企业机构对数据集的分析能力，使企业机构能够在存储库中构建数据，并对数据资产的使用进行即时控制。各类安全控制措施，例如数据丢失防护（DLP）和数据访问治理（DAG），都可极大受益于数据分类或数据标签。数据分类使企业机构能够更轻松地查找和验证数据，同时避免过度保护和留存，从而经济高效地履行监管合规义务。

## 推动因素

- 当前的法律和地缘政治局势增加了人们对数据驻留和主权的担忧，尤其是在涉及重要数据和个人信息时。然而，当前这种“出问题后再解决”的安全治理实践十分低效，因此从数据分类入手以简化和自动化这些流程的需求越来越多。
- 数据分类方法日趋成熟，可按类型、所有者、监管要求、敏感度和留存要求进行分类，使企业机构能够将安全、隐私和分析工作的重点放在重要数据集及其分类上。
- 预先定义了行业特定分类（如金融、电信、医疗和政府）的自动化数据分类工具的出现，降低了启动数据分类项目时对业务知识和安全知识的要求。

## 阻碍因素

- 由于缺乏充分的培训并且依赖于用户自行分类，传统的数据分类计划经常以失败告终。
- 数据分类工作的思路往往是以安全为中心，没有使用业务语言向用户解释分类的目的，导致用户参与度不高。
- 尽管许多供应商提供了自动化数据分类工具，可以更准确地进行数据分类，同时最大限度地减少用户工作量，但分类结果的准确性仍未达到预期，特别是需要对模型进行长期训练的机器学习或人工智能算法。
- 从合规角度来看，如果企业机构所在的行业未受严格监管，或行业监管机构未发布分类标准，那么企业机构可能很难衡量或验证数据分类的效果。

## 使用建议

- 全面评估企业机构内部数据类型和敏感度，并与业务部门和数据分析团队合作，识别对数据分类至关重要的具体用例，从而确定企业机构范围内的数据分类用例和需要开展的工作。
- 在数据安全治理项目中，将用户自行分类和自动化数据分类相结合，并安排用户培训。
- 分析行业监管机构或国家标准委员会发布的数据分类指南和标准，制定符合监管要求的数据分类方案。
- 优先考虑能够与其他数据安全技术，如匿名化、加密、DLP和数据安全平台（DSP）等更好地集成和互操作的数据分类工具。同时也要考虑其他方面，例如更丰富的内置分类模板和灵活的自定义标签。



## 厂商示例

安恒信息、观安信息、美创科技、绿盟科技、全知科技、天融信、明朝万达

## Gartner相关推荐阅读

[Still a Moving Target — What to Do With the Chinese Data Security Law](#)

[Building Effective Data Classification and Handling Documents](#)

[Case Study: An Active Metadata Augmented Data Classification System to Boost Analytics Efficiency](#)

[Improving Unstructured Data Security With Classification](#)

[How to Succeed With Data Classification Using Modern Approaches](#)

附录

技术成熟度曲线的各个阶段、影响力评级和成熟度等级

Table 2: 技术成熟度曲线的各个阶段  
(Enlarged table in Appendix)

阶段 ↓	定义 ↓
技术萌芽期	某一创新的突破进展、公开展示、产品发布等事件，引起了媒体与行业的极大兴趣。
期望膨胀期	外界对某一创新寄予过高的热情和不切实际的期待。技术领先企业大力宣传的项目多以失败告终，只有一小部分取得成功。在此过程中，会展公司和媒体是仅有的获利者。
泡沫破裂低谷期	创新未能满足人们的过高期待，迅速褪去热度。媒体报道的兴趣逐渐降低，只余下几个令人警醒的故事。
稳步爬升复苏期	有针对性的试验和扎实的工作，使人们真正了解到一项创新的适用性、风险点和影响力。商业化的现成方法和工具，使开发流程得到简化。
生产成熟期	某一创新的现实影响得到展示和认可，相关工具和方法不断完善，出现第二代、第三代版本，效果日趋稳定，风险亦逐渐降低，因此接受度也得到提高，开启了采用率快速增长的新阶段。大约20%的目标受众在此阶段已采用或开始采用相关技术。
距离主流采用的时间	一项创新进入生产成熟期所需的时间。

来源：Gartner（2023年7月）

Table 3: 影响力评级

影响力评级 ↓	定义 ↓
颠覆	催生出跨行业开展业务的新方式，可引发行业重大转变。
较高	催生出执行横向或纵向流程的新方法，可为企业显著增加营收或大幅降低成本。
中等	逐步改进现有流程，可为企业增加营收或降低成本。
较低	小幅改进部分流程（例如提升用户体验），难以真正增加营收或降低成本。

来源：Gartner（2023年7月）

Table 4: 成熟度等级

成熟度等级 ↓	状态 ↓	产品/厂商 ↓
孵化阶段	实验室阶段	无
发展阶段	商业化阶段 行业领军企业进行试点和部署	第一代 价格高昂 高度定制化
成型阶段	技术能力和流程理解趋向成熟 运用范围扩大，不再局限于早期采用者	第二代 轻度定制化
主流采用起步阶段	技术得到验证 厂商和技术快速发展，采用率快速提高	第三代 开箱即用方法增多
主流采用成熟阶段	技术稳定可靠 厂商和技术鲜有变化	数家厂商占据主导地位
延续阶段	不适用于开发新项目 替换受到迁移成本制约	维护营收成为重点
淘汰阶段	极少使用	仅在二手/转售市场可见

来源：Gartner（2023年7月）

证据

<sup>1</sup>请参阅 [China Unveils Guidelines on Improve Market-Based Allocation of Production Factors](#)。

<sup>2</sup> 2022年Gartner中国AI用例调研：本调研旨在了解中国的AI实施情况，以及企业机构在部署AI用例方面最为成功的领域。本调研于2022年11月14日至12月16日经线上进行，共有300名来自中国企业机构的受访者参与。调研按企业规模（按年营收计算）和行业对样本设定了配额，以确保样本具有良好的代表性。样本包括45家小型企业（低于5000万美元）、105家中型企业（5000万美元至5亿美元以下）、120家大型企业（5亿美元至100亿美元以下），以及30家全球性企业（超过100亿美元）。参与调研的企业机构须开发过AI产品。受访者须担任经理或更高级别的职务，并且高度参与从构思到测量AI用例生命周期中的至少一个阶段。免责声明：这项研究的结果，并不等同于全球调研结果或市场整体情况，仅代表受访者和受访企业的观点。

---

## Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

[Understanding Gartner's Hype Cycles](#)

[Tool: Create Your Own Hype Cycle With Gartner's Hype Cycle Builder](#)

[Predicts 2023: Strategic Growth for China in Times of Uncertainty](#)

[Forecast Analysis: IT Spending, China](#)

[Chinese AI Survey Analysis: AI Trends Wave 3.0 — From Operational to Strategic](#)

[Exploiting the Evolving Database Management System Trends in China](#)

[3 Reasons to Invest in a Metadata-Driven Data Fabric Design in China](#)

[Demystify Data Middle Office by Nurturing Core D&A Capabilities](#)

[Quick Answer: China Perspective — Frequently Asked Questions on ChatGPT and Large Language Models](#)

[3 Ways to Promote Your Data Agenda at the Center of the Chinese Digital Economy](#)

---

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.

Table 1: 中国数据分析和人工智能技术优先级矩阵

影响力	距离主流采用的时间			
↓	2年以内 ↓	2~5年 ↓	5~10年 ↓	10年以上 ↓
颠覆	数据资产管理	公民数据科学 决策智能 实时数据管理 数据共享 生成式AI	大模型 组装式数据和分析 负责任的AI	数据编织
较高		国产人工智能芯片 数字伦理 数据分类 数据治理 数据素养	AI工程化 因果AI 知识图谱 联邦机器学习 自主可控数据库	
中等		自助分析	增强型数据和分析 湖仓一体	
较低				

来源：Gartner (2023年7月)

Table 2: 技术成熟度曲线的各个阶段

阶段 ↓	定义 ↓
技术萌芽期	某一创新的突破进展、公开展示、产品发布等事件，引起了媒体与行业的极大兴趣。
期望膨胀期	外界对某一创新寄予过高的热情和不切实际的期待。技术领先企业大力宣传的项目多以失败告终，只有一小部分取得成功。在此过程中，会展公司和媒体是仅有的获利者。
泡沫破裂低谷期	创新未能满足人们的过高期待，迅速褪去热度。媒体报道的兴趣逐渐降低，只余下几个令人警醒的故事。
稳步爬升复苏期	有针对性的试验和扎实的工作，使人们真正了解到一项创新的适用性、风险点和影响力。商业化的现成方法和工具，使开发流程得到简化。
生产成熟期	某一创新的现实影响得到展示和认可，相关工具和方法不断完善，出现第二代、第三代版本，效果日趋稳定，风险亦逐渐降低，因此接受度也得到提高，开启了采用率快速增长的新阶段。大约20%的目标受众在此阶段已采用或开始采用相关技术。
距离主流采用的时间	一项创新进入生产成熟期所需的时间。

来源：Gartner（2023年7月）



Table 3: 影响力评级

影响力评级 ↓	定义 ↓
颠覆	催生出跨行业开展业务的新方式，可引发行业重大转变。
较高	催生出执行横向或纵向流程的新方法，可为企业显著增加营收或大幅降低成本。
中等	逐步改进现有流程，可为企业增加营收或降低成本。
较低	小幅改进部分流程（例如提升用户体验），难以真正增加营收或降低成本。

来源：Gartner（2023年7月）

Table 4: 成熟度等级

成熟度等级 ↓	状态 ↓	产品/厂商 ↓
孵化阶段	实验室阶段	无
发展阶段	商业化阶段 行业领军企业进行试点和部署	第一代 价格高昂 高度定制化
成型阶段	技术能力和流程理解趋向成熟 运用范围扩大，不再局限于早期采用者	第二代 轻度定制化
主流采用起步阶段	技术得到验证 厂商和技术快速发展，采用率快速提高	第三代 开箱即用方法增多
主流采用成熟阶段	技术稳定可靠 厂商和技术鲜有变化	数家厂商占据主导地位
延续阶段	不适用于开发新项目 替换受到迁移成本制约	维护营收成为重点
淘汰阶段	极少使用	仅在二手/转售市场可见

来源：Gartner（2023年7月）