

深度解读 Gartner 技术成熟度曲线 Hype Cycle for Data Management, 2023



深度数据云
关注深度数据云，把握数据、云计算、AI智能的技术与行业脉动

+ 关注她

8 人赞同了该文章

在今年 7 月中旬，Gartner 发布了最新的数据管理技术成熟度曲线——“Hype Cycle for Data Management, 2023”（如需获取 PDF 原文件，关注深度数据云微信公众号，首页对话框输入【Data】），数据管理技术成熟度曲线是 Gartner Hype Cycle 技术成熟度曲线家族中和数据库、数据仓库⁺、数据平台最为相关的。近年来，随着云采用的迅猛增长，企业将大量的数据迁移上云，而数据迁移给数据管理的发展和创新带来了巨大的市场空间，但是也面临了来自各方面的挑战，比如数字业务的变化以及日益复杂的生态系统。

为了应对这些挑战，一些新的技术进入了今年的曲线、一些已经完全市场成熟、一些迅猛发展中。在今天的文章中，小瑞将和大家一起看看今年的技术成熟度曲线⁺上是如何反应出这些变化的吧。

01. 两图概览

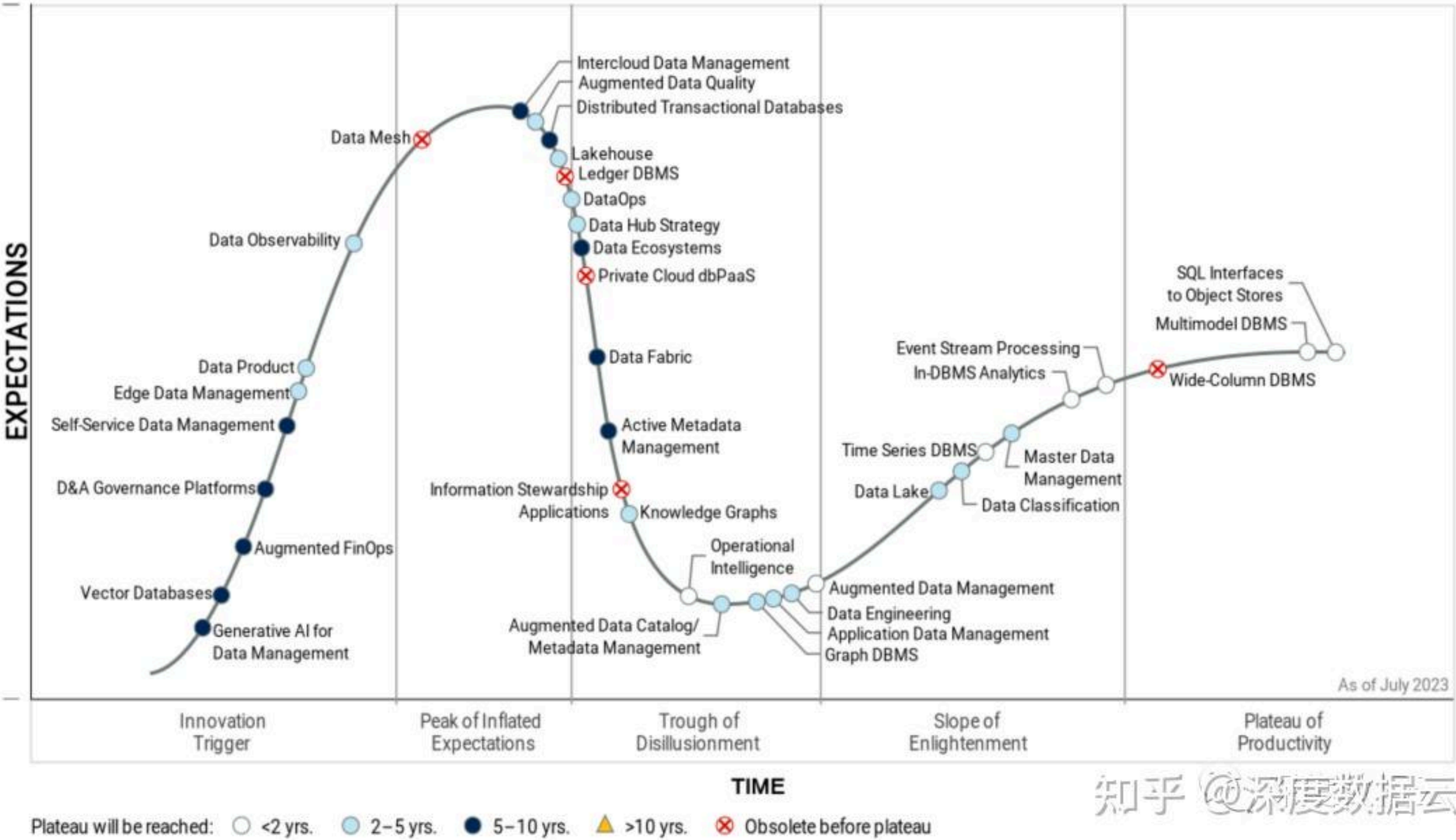
Gartner的Hype Cycle报告核心包括两张图：

- 技术成熟度曲线
- 优先级矩阵⁺

下面分别描述一下这两张图的含义。

技术成熟度曲线

Hype Cycle for Data Management, 2023



技术成熟度曲线从两个维度来衡量一项创新技术的成熟度：「当前所处阶段」和「主流采用所需时间」。

其中，「当前所处阶段」分为 5 个阶段：

- 技术启动 (Innovation Trigger): 该技术开始获得媒体关注、产生舆论，但是可能没有实际的产品和应用。
- 期望膨胀 (Peak of Inflated Expectations): 由于媒体过度炒作，导致公众对该技术的期望被过度放大。此期间可能会出现一些成功的案例，但是更多的是失败的尝试。
- 失望谷 (Trough of Disillusionment): 当实际效果达不到过度炒作的期望时，工作会开始对技术感到失望。
- 启蒙坡道 (Slope of Enlightenment): 一些企业开始了解如
- 生产高地 (Plateau of Productivity): 该技术已经成熟且被

已赞同 9

添加评论

分享

喜欢

收藏

申请转载

...

今年总共有 37 项技术出现在成熟度曲线中。

优先级矩阵

Table 1: Priority Matrix for Data Management, 2023				
Benefit	Years to Mainstream Adoption			
	Less Than 2 Years	2 - 5 Years	5 - 10 Years	More Than 10 Years
Transformational	Event Stream Processing	Augmented Data Quality Data Product	Active Metadata Management Augmented FinOps Data Fabric Generative AI for Data Management	
High	Augmented Data Management In-DBMS Analytics Operational Intelligence	Augmented Data Catalog/Metadata Management Data Classification Data Engineering Data Hub Strategy Data Observability DataOps Edge Data Management Graph DBMS Knowledge Graphs Lakehouse Master Data Management	D&A Governance Platforms Data Ecosystems Intercloud Data Management Vector Databases	
Moderate	Multimodel DBMS SQL Interfaces to Object Stores Time Series DBMS	Application Data Management Data Lake	Distributed Transactional Databases Self-Service Data Management	
Low				

优先级矩阵包含两个维度：

- **效益评估** (Benefit)：表示技术项带来的收益的高低，分为 4 种，变革性 (Transformational)、高 (High)、中 (Moderate)、低 (Low)
- **主流采用所需时间** (Years to Mainstream Adoption)：表示技术项被主流市场接受的时间预估，包括 Less Than 2 Years、2~5 Years、5~10 Years、More Than 10 Years。这个维度在上面技术成熟度曲线图中也有所体现，对应 “Plateau will be reached” 指标

02.今年出现的新创新技术

生成式 AI 的冲击、不断增长的数据量、层出不穷的新用例等激发了数据管理领域的技术创新。在今年的「数据管理成熟度曲线」上，出现了 4 个新面孔：

自助式数据管理 (Self-Service Data Management)、数据产品 (Data Product)、用于数据管理的生成式 AI (Generative AI for Data Management)、[向量数据库⁺](#) (Vector Database)

其中，「[数据产品⁺](#)」和「用于数据管理的生成式 AI」评级为变革性。关于这两项技术的详细分析可以在本文最后一个部分找到。

03.一些创新技术移除

其中，「[数据集成工具⁺](#)」和「[数据预处理工具](#)」已完全市场

而还有一些技术则在成熟前被淘汰，包括：

- **数据网格⁺** (Data Mesh)：虽然市场预期值仍在增长，但 Gartner 预期其主要的核心功能将被**数据编织⁺** (Data Fabric) 所取代。
- 作为独立技术的私有云数据库平台即服务 (Private Cloud dbPaaS)：主要的供应商和终端用户已经或正在将**业务往公有云⁺上迁移**。
- 分类账数据库 (Ledger Database)：在市场上，不具有数据库功能的不可变对象存储和分布式分类账的使用已获得市场接受且已成熟。
- 宽列数据库 (Wide-Column Database)：该技术已经进化为多模数据库的一个**特性**，而非一项独立的技术。

04.其他技术的发展

今年，有多个产品进入了「期望膨胀」期，包括增强型数据质量 (Augmented Data Quality)、Lakehouse、**分布式事务数据库⁺** (Distributed Transactional Database)、Intercloud 数据管理等。Lakehouse 在去年才刚刚进入「期望膨胀」期并快速的获得了显著的市场渗透率。

而围绕未来数据架构的几种技术和学科目前正在进入或通过低谷：DataOps、数据编织、活动元数据管理 (Active Metadata Management)，以及增强型数据目录/**元数据管理⁺** (Augmented Data Cataloging/Metadata Management) 解决方案。

一些技术，如运营智能 (Operational Intelligence)、**图数据库管理系统⁺** (graph DBMS)、**知识图谱⁺** (Knowledge Graph)，数据工程和增强型数据管理 (Data Engineering and Augmented Data Management)，预计在未来五年内将进入高原阶段。

还有部分技术已经抵达了「生产高地」：

对象存储的 SQL 查询接口 (SQL Interfaces to Object Stores)：该技术发展极其迅速，它将标准 SQL 接口和便宜的对象存储/文件存储结合起来，使企业可以使用 SQL 和存储在对象存储或文件存储中的数据进行交互，得到了市场的喜爱。

多模数据库⁺：将多种数据引擎（包括关系型或非关系型）结合到一个数据库中，可以针对不同的数据，分门别类地进行优化。

05.各类创新技术详细解析

接下来小瑞从 Gartner 关于每项技术详细分析中摘取了部分，供大家参考。如果您感兴趣的技术不在本文中，可以关注深度数据云微信公众号，首页对话框输入【Data】获取报告原文哦~

用于数据管理的生成式 AI

效益评估：变革性 | 成熟度：萌芽期

市场渗透率：目标受众不足 1%

定义：生成型 AI 通过学习可以派生出新的内容、策略、设计和方法。生成型 AI 的出现对企业业务产生了深远的影响，这些影响体现在方方面面，包括内容发现与创作、工作自动化、用户体验等。在数据管理领域，生成式 AI 的应用使用户可以使用**自然语言进行数据管理**，简化了数据管理活动，降低了数据管理的门槛。

重要性：生成式 AI 产品的利用可能会改变数据管理学科，使**更多的人能够参与到自助式数据管理活动**中，使通过对话式界面进行数据管理任务成为可能。

商业影响：Gartner 2023 年的 CDAO 调查再次将技能的可用性确定为数据和分析的反复出现的问题。与此同时，商业界要求更大的灵活性和敏捷性来访问数据，这**推动了数据管理活动的分散化⁺**。使用对话式界面将减轻像数据工程师这样的高技能角色的压力，并使数据管理员甚至 DBA 的工作变得更轻松。这可能导致组织模式、角色和数据管理实践的转变。

驱动因素：加快企业对数据和分析用例创新和落地、降低数据管理的门槛使更多角色可以参与进来、**提高数据管理专家的效率、使用自然语言进行数据管理活动**等

阻碍因素：萌芽期效果还不明显、**安全问题和**访问控制问题需要专业人士处理、生成式 AI 需要**专门的人力来进行监管**等

提供商：AWS、Databricks、谷歌、Informatica、微软、PingCAP 等

向量数据库

效益评估：高 | 成熟度：萌芽期

市场渗透率：不足 1% 的目标受众

定义：向量数据库存储数据的数值表示。在这样的数据库中，每个点都由具有固定数量维度的向量表示，这些向量可以通过数学运算（如**距离度量⁺**）进行比较。向量数据库通常用于 ML 解决方案中，使用向量来表示数据的特征/属性，如文本嵌入。向量数据库可以**极大地降低用户在搜索相似的数据点时的延时**。

重要性：向量数据库适用于**相似性搜索**和**产品推荐**等用例。生成式 AI 的快速创新和 AI 基础模型的采用推动了向量数据库市场的发展。当客户采用生成式 AI 模型时，向量数据库被用于存储模型训练产生的嵌入 (Embedding)，从而可以进行相似性搜索，将提示与特定或相似的向量嵌入进行匹配。

商业影响：企业通过提供差异化的客户体验 (CX) 蓬勃发展。生成型 AI 越来越多地嵌入到应用中，以**增强人机共生**。在此背景下，企业迫切需要可扩展的方式来加速这类应用的长期构建与支持。向量数据库是一种重要的后端服务，允许企业对其启用生成式 AI 能力的应用进行**面向未来验证和扩展**。

驱动因素：向量嵌入的流行、性能和扩展性需求、检索和生成式模型的混合实现等

阻碍因素：对向量数据库能力缺乏了解、向量数据库在数据迁移、数据集成以及用力扩展方面的局限性、自建涉及众多**技术栈⁺**、价格不便宜等

提供商：Couchbase、Croma、Elastic、谷歌、Pinecone Systems 等

事件流处理

效益评估：变革性 | 成熟度：早期主流

市场渗透率：目标受众的 20% 至 50%

定义：事件流处理 (ESP) 是对流数据进行计算，从而进行流分析或流数据集成的一项技术。ESP 支持动态数据 (Data in Motion)，使得对威胁和机会的情境感知和**近实时响应**成为可能，或者存储数据流以供后续应用使用。

重要性：ESP 使持续智能和数字业务的实时性要求成为可能。ESP 的动态数据架构是**对传统的静态数据方法的根本性替代**，这也使得 ESP 平台从一个小众创新技术发展至今，并被大部分用户所应用。Gartner 预计 ESP 将在**两年内达到生产力高峰**。

商业影响：ESP 改变了金融市场，成为电信网络、智能电网，以及一些物联网、供应链、车队管理和其他交通运营的必需品。然而，在接下来的 10 年中，ESP 的大部分增长将来自已经确立的领域，特别是物联网。ESP 平台的流分析**通过仪表板和告警提供态势感知**，并支持**异常检测**等。

驱动因素：企业流数据的爆发式增长、流数据处理成本和难度的下降、实时数据处理的需求、产品的广泛可获得性等

阻碍因素：应用场景的局限性、许多架构师和软件工程师不了解 ESP 产品的使用等

提供商：Couchbase、Croma、Elastic、谷歌、Pinecone Systems 等

增强型数据质量

效益评估：变革性 | 成熟度：早期主流

市场渗透率：目标受众的 20% 至 50%

定义：增强的数据质量 (ADQ) 解决方案提供了增强体验的功能，旨在改善洞察发现；提出下一个最佳行动建议；并通过利用 AI/ML 特性、图形分析和元数据分析实现自动化。每种技术都可以独立工作，也可以合作创建网络效应，然后可以用来增强在各种数据质量用例中的**数据质量自动化和效果**。

重要性：确保高质量的数据对于数据和分析工作至关重要。基于当代数据环境的快速扩展，众多的数据类型和商业的迫切需求，组织正在寻找快速、经济、可扩展且易于实施的创新方法来解决数据质量问题。ADQ 技术通过增加自动化和增强洞察力，**革新了传统的、耗时的手动程序**。

商业影响：**提高了数据质量，减少了人工**并提高了效率；多元化用户可用性使非技术用户能够通过自然语言运行过程，消除了技能障碍；AI/ML 技术和元数据分析增强了多个数据质量过程；语义连接、血统追踪和领域数据映射使知识图谱能够识别影响/ 解决方案；对数据工程师的支持包括在复杂的场景中进行监控/可观察性。

驱动因素：传统的数据质量实践严重依赖于人工、对于新兴和未来生态至关重要、ADQ 可以利用 ML、NLP 等先进技术来提升企业数据管理能力等

阻碍因素：企业对 ADQ 的认识不足、和现有的数据基础设施的集成度过低、扩展性上的局限、对 AI/ML 算法的解释能力和跟踪能力不足等

提供商：Ataccama、Collibra、DQLabs、Experian、IBM 等

数据产品

效益评估：变革性 | 成熟度：萌芽期

市场渗透率：不足 1% 的目标受众

定义：数据产品是一组精选数据、元数据、语义和模板的自包含组合。它包括用于处理特定业务场景和重用的访问和实施逻辑。数据产品必须**被消费者信任，时刻保持与时俱进，并且可用性已经过验证**。数据产品使很多数据和分析用例变为可能，包括数据共享和数据货币化等。

重要性：当今的商业环境要求更快更好的数据集成和治理。而数据产品自带的可信赖、自包含、好治理、经过重用验证等特性，非常适合应用于数据和分析领域中有针对性的商业用例。
商业影响：数据产品使得基于数据的自助服务、响应加速得以实现；数据产品可以编目、发现、预集成，**使企业能更多的关注于结果**；成功的数据产品可被重用。

驱动因素：更快的数据分析、数据产品重用、数据产品使领域团队离数据更近、带来了新的收入流、支持 DataOps 等

阻碍因素：数据产品的泛滥、要求运营模型的转变、疏于计划等

增强型 FinOps

效益评估：变革性 | 成熟度：萌芽期

市场渗透率：不足 1% 的目标受众

定义：FinOps 将传统的 DevOps 概念，如敏捷性、持续集成和部署以及终端用户反馈，应用于**财务治理、预算和成本优化**工作。增强型 FinOps 则是通过 AI 和 ML 应用将 FinOps 的这个过程（主要是在云上）自动化，根据客户使用自然语言描述的商业目标来自动完成成本的优化。

重要性：虽然现目前在云环境中已经可以完成特定项目的特定工作负载的成本评估，但是由于底层云基础设施和服务提供的选择的复杂性和多样性，以及统一的价格模型缺失，价格/性能比（云效率的主要衡量指标）难以进行有效评估。增强型 FinOps 可以通过**应用 AI/ML 技术来自动化**这个过程。

商业影响：云预算规划和财务运营的自动化在理想状态下，允许企业用自然语言来表达他们的商业目标，同时允许企业的云生态系统自动优化底层云资源来满足这些目标。从而实现资源的优化利用，节省开销。

驱动因素：云采用者希望云支出透明化、供应商将成本作为竞争优势、云环境非常适合 ML 和 AI 应用等

阻碍因素：云服务供应商的定价模型过于复杂、业界尚未形成的广泛采用的统一的云成本、使用和计费标准等

提供商：Acceldata、Anodot、Apptio、Capital One Software、Densify 等

数据编织

效益评估：变革性 | 成熟度：成长期

市场渗透率：目标受众的 1% 至 5%

定义：数据编织是一种设计框架，用于获得**灵活且可重用的数据管道、服务和语义**。数据编织利用了数据整合、活动元数据、知识图谱、分析、ML 和数据目录。数据编织通过“观察和利用 (Observe and Leverage)”的方式颠覆了当前数据管理的主导方式——针对数据和用例的“定制构建 (Build to Suit)”。

重要性：数据编织将传统方式和技术进步融合在一起，避免了“拆除和替换”。它利用了沉没成本，并同时为新的数据管理成本进行优先级排序和成本控制指导，提供了**灵活性和可扩展性**，确保数据能够在本地、多云或混合云中被人 and 机器使用和重用。

商业影响：**提高了数据的识别、部署和可用性**，使大规模重用数据成为可能，无论其来源或结构如何，都能识

别相似的数据内容，从而实现与新资产的连接；在整个数据生态系统中实现可观察性；降低了与管理数据相关的维护、支持和优化成本。

驱动因素：高效数据重用的需求增加；对新数据资产的快速理解的需求急剧上升；对数据跟踪、审计、监控、报告和评估使用和利用以及数据分析的需求增加等

阻碍因素：企业习惯于将预算或员工应用于一次性和点对点的集成解决方案；缺乏统一的设计和语义标准来记录和共享元数据；数据编织需要分析能力和 ML 能力来推断缺失的元数据，该过程易产生错误等

提供商：Cambridge Semantics、Cinchy、CluedIn、Denodo Technologies、IBM 等

湖仓

效益评估：高 | 成熟度：成长期

市场渗透率：目标受众的 5% 至 20%

定义：湖仓是一种融合的基础设施环境，它**结合了数据湖+在语义上的灵活性和数据仓库在生产优化和交付上的优势**。它支持数据从原始、未经提炼的状态，经过提炼的步骤，最终交付优化的数据供消费的完整过程。

重要性：对于企业而言，使用数据湖，并通过将数据湖与其他相关系统集成起来面临不小的挑战。比如，将数据湖与数据仓库一起使用会增加数据和分析的复杂性。而湖仓的出现恰好是为了解决这个问题，它将数据湖和数据仓库的结合起来，简化了架构的同时还提升了效率，并实现了两者之间数据和分析模型移动最小化。

商业影响：湖仓简化了交付、加速了数据获取，提供了一个**统一的数据管理平台**。这个管理平台不仅支持高技能的数据科学家、工程师和分析师，也支持那些使用预制报表和仪表盘进行数据消费的普通用户。湖仓提供了从面向发现的分析和分析模型开发（通过湖仓的湖部分）到向终端用户交付分析洞察和量化（通过仓库部分）的明确路径。

驱动因素：使数据科学项目运营化面临的挑战；企业一直在寻求快速且无阻的数据访问，并且在处理与数据仓库相关的延迟交付的过程和感知上存在困难；数据湖仓的概念快速成熟等
阻碍因素：技术尚未完全成熟；最复杂的数据仓库工作负载仍可能超出大多数未包含已开发功能的湖仓解决方案的范围；用户设计、部署和维护复杂数据架构能力的不成熟等

提供商：AWS、ChaosSearch、Databricks、Dremio、谷歌等

小结

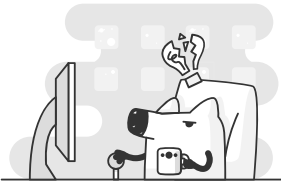
以上就是本期小瑞关于《Hype Cycle for Data Management, 2023》全部内容了，我们下期再见吧。

编辑于 2023-09-18 14:13 · IP 属地浙江

Gartner 数据质量管理



理性发言，友善互动



还没有评论，发表第一个评论吧