

# Data Fabric 架构的三大即 时优势

整合并使用贵企业所有数据,以获得  
更好的业务成果

## 目录

3 目前的数据挑战

5 Data Fabric的商业收益

6 IBM 如何提供Data Fabric

7 运行中的Data Fabric

7 Data Fabric是贵企业的下一步转型

# 目前的数据挑战

数据是企业数字化转型中不可或缺的元素。但当企业设法利用自己的数据时，遇到来自多种数据源、类型、结构、环境和平台的挑战。当企业采用混合多云架构时，这种多维数据困境的复杂性进一步增加。目前，对于许多企业而言，运营数据在很大程度上仍是孤立和隐藏的，导致出现大量暗数据。

例如某北美能源企业试图转型为数据驱动型组织 — 想要跨多个业务部门随时获得数据科学能力。他们很快意识到，企业的数字化转型受到数据孤岛、工具不一致和技能水平参差不齐的阻碍，导致数据能力之间出现巨大差距。

该企业面临的问题并非其独有；事实上，这是那些数据环境的发展超出数据管理架构的常见后果。

最终，一个新的数据架构概念 — Data Fabric将该能源企业带回到数字化转型的成功道路上。

那么究竟什么是Data Fabric，它与以前的架构有何不同，它可以帮助企业实现什么，以及 IBM 在实施过程中起到什么作用？本白皮书将给出答案。

## 什么是Data Fabric？

过去，企业试图通过点对点集成或引入数据中心来解决数据访问问题。这两种方法不适用于数据高度分散和数据孤岛。点对点集成给需要连接的额外终端增加呈指数级增长的成本，意味着这种方法不可扩展。数据中心支持更简便地集成应用和数据源，但却加剧在中心内部维护质量和信任的成本与复杂性。

Data Fabric 是一个新兴的架构，旨在解决混合数据环境中出现的数据挑战。其基础理念是充当数据端点之间的虚拟连接组织，在去中心化和全球化之间取得平衡（参见第 4 页）。

通过自动化和增强集成、联合治理以及激活元数据等技术，Data Fabric 架构能够在分布式环境中实现动态和智能数据统筹，创建一个即时可用的信息网络，为企业提供动力。

Data Fabric 独立于部署平台、数据流程、地理位置和架构方法。它有助于将数据用作企业资产。Data Fabric 确保您高效且有效地成功组合、访问和治理各种数据。

## Data Fabric的能力和原则

Data Fabric架构的核心是数据管理平台，它支持全面的综合数据管理能力，包括发现、治理、整理和统筹。

然而，Data Fabric从传统的数据管理概念（例如只关注构建实践的 DataOps）推进和发展而来，旨在提高数据运营水平。它基于分布式架构和先进技术构建，能够解决数据资产的极端多样性和分散所产生的需求。

一个Data Fabric在逻辑上可分为四种能力(或组件)。

数据来源和用户

### 知识、洞察和语义

- 提供一个数据市场和购物体验
- 用知识和语义自动丰富已发现的数据资产,帮助消费者找到并理解数据。

### 统一治理与合规性

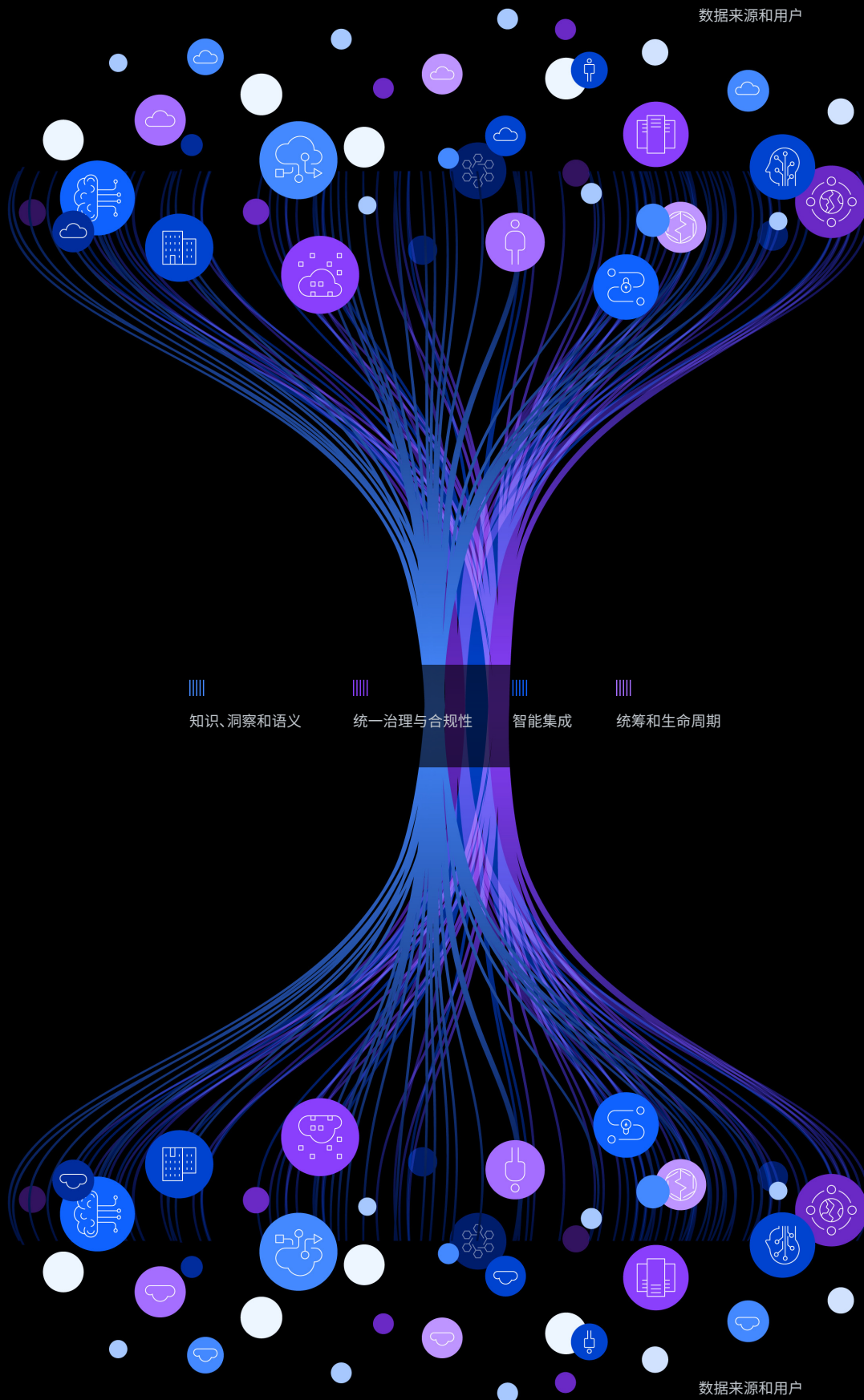
- 允许本地管理和治理元数据,但支持全球统一的视图和策略执行
- 根据全球和当地规则,自动针对数据资产应用策略
- 利用先进的能力,自动进行数据资产分类与整理
- 为所有编录资产自动建立可查询的访问路线,以增强数据激活能力

### 智能集成

- 通过跨分布式数据源的自动化流程和管道创建,帮助数据工程师快速完成任务
- 在当地和全球深入执行数据保护策略,实现对任何数据的自助式接入和访问
- 通过优化的工作负载分配和自我调整以及纠正架构漂移,自动确定最佳执行方式

### 统筹和生命周期

- 支持数据管道的构成、测试、运营和监测
- 在数据生命周期中融入 AI 能力,以实现任务自动化、自我调整、自动修复和检测源数据变化,最终促进自动更新。



知识、洞察和语义

统一治理与合规性

智能集成

统筹和生命周期

数据来源和用户

# Data Fabric的商业收益

数据只有处于背景中并支持企业中任何用户或应用访问时,才能带来商业价值。如果实施得当,Data Fabric能够确保在整个企业中以最有效和最自动化的方式展现这些价值。因此,该结构拥有三大优势:

1. 支持自助式数据使用和协作
2. 使用现有元数据实现自动治理、保护和保障
3. 自动执行数据工程任务,增强跨混合云资源的数据集成能力

## 支持自助式数据使用和协作

通过集成多个数据源的数据并针对每天生成的海量数据进行大量分析,企业能够获得更好的洞察,并加速响应不断变化的业务需求。Data Fabric能够快速将数据送到需要的人手中。自助式服务支持整个企业更快地找到适当的数据,并将更多时间用于使用数据提供切实可行的洞察。

Data Fabric为自助式数据使用带来的好处:

- 企业用户拥有单一访问点,可用于寻找、理解、塑造和使用整个企业的数据。
  - 集中式数据治理和世系有助于用户了解数据的含义、来源及其与其他资产的关系。
  - 大规模、定制元数据管理可轻松扩展,并可通过 API 访问。
  - 自助访问可信和受管理的数据,有助于业务线与其他用户协作。
- 一项 Forrester 总经济影响研究<sup>1</sup> 显示,这些能力可能意味着:



## 使用现有元数据实现自动治理、数据保护和安全保障

为所有数据计划建立分布式主动治理层通过提供信任和透明度,来减少合规性和监管风险。它能够对任何数据访问自动执行策略,提供高水平的数据保护和合规性。

Data Fabric用户通过使用 AI 和机器学习技术,可以提高自动化水平,例如,根据监管文档中的语言和定义自动提取数据治理规则。这使企业能够在几分钟内应用特定于行业的治理规则,帮助避免昂贵的罚款,并确保无论数据存储在哪里都能以符合道德规范的方式使用。

为实现治理虚拟化的Data Fabric的好处:

- 数据工程师、数据科学家和商业分析师的敏捷性、安全性和生产力得到提高。
- 集成多个全球数据源。
- 可能大规模发现新的行业领先的个人信息 (PII) 和关键数据元素。

这些能力可能意味着:



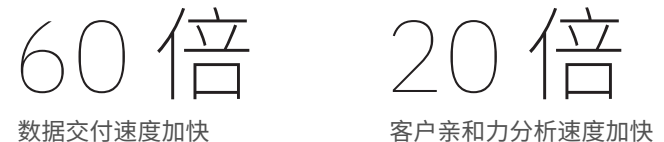
## 实现数据工程任务自动化,并增强数据集成能力

先进的数据工程意味着几乎任何数据访问或交付过程都实现自动化,不需要任何繁琐或容易出错的编码过程。增强集成利用元数据数据来优化数据交付和访问。

Data Fabric对数据工程和集成的好处:

- 自动优化的数据集成有助于加速数据交付。
- 自动化工作负载平衡和弹性扩展意味着随时准备在任何环境和对任何数据量进行作业。
- 内置弹性和 CI/CD 自动化。
- 实时捕捉变化的自动化流程支持为业务流程提供高质量数据。
- 机器学习可以自动化和扩展定制的数据发现、分类和整理流程,从而加快价值实现的速度。
- 无论数据存储在哪里,都可以实时自动进行连续分析。

某知名零售商的成果<sup>3</sup>:



# IBM 如何提供Data Fabric

IBM Cloud Pak® for Data 让Data Fabric的概念成为现实。IBM Cloud Pak for Data 是一个洞察平台，能够简化和自动化数据收集、组织和分析，并加速在整个企业融入 AI。

IBM Cloud Pak for Data 能够随处连接数据、运行工作负载以及在混合云环境中大规模构建、部署和管理 AI，它是企业数字化转型的助推器。

该平台为混合型企业提供无缝集成，帮助他们获得：

- IBM Cloud Pak for Data 服务
- 外部应用和数据源
- 进行数据管理和治理的基于 AI 的先进能力

这种智能基础使消费者既可以在成本、性能、合规性之间取得最佳平衡，又能够获得经过整理的数据，还可以根据工作负载、数据位置和数据策略来统筹和优化数据处理。

IBM Cloud Pak for Data 利用以下能力来提供Data Fabric所需的业务就绪数据。这些能力在支持Data Fabric架构方面发挥作用。

## 1. 基于元数据的知识核心

便于发现数据源和目录；丰富数据资产；并进行分析以提取洞察，利用 AI 实现更多自动化。知识核心使用语义搜索为市场提供动力。

## 2. 自助式数据市场

下一代数据目录，帮助业务分析师等数据消费者从企业的数据环境中检索数据。

## 3. 智能集成

通过提取、虚拟化、转换和流式传输数据，支持数据使用。它与知识核心集成，以实现数据集成自动化，还能够根据工作负载和数据策略智能选择最适合的集成方法。它也可用于数据准备，作为数据工程工作负载的一部分，或用于创建数据产品。最后，它还能够向数据产品发布更新。

## 4. 治理

编录和整理元数据，定义数据政策以保护隐私，整理数据，捕捉数据世界，并执行其他与安全性和合规性有关的任务。这一层可以判断数据格式（例如，结构化与非结构化）和数据重要性（例如，公开与受保护），并将正确的政策应用于所有数据和潜在用户。无需通过人工将标准和规则应用于数据，这种集成的平台能力意味着可以在企业层面应用，并根据需要在所种数据资源中传播。不同工具中的分析模型可以沟通；细化层面的数据策略执行可以在很大程度上实现自动化。

## 5. 统一的开发和运营

实现统一的生命周期，便于配置和运行运转中的数据平台的所有方面。



图 1. IBM Cloud Pak for Data 平台的能力—支持 AI 开发和Data Fabric

# 运行中的Data Fabric

了解Data Fabric价值的最佳方式是看到组织通过在洞察平台上实施Data Fabric而获得的实际商业收益。

## 能源企业

本白皮书开篇的挑战部分介绍了某北美能源企业的数字化转型案例。

该客户与 IBM 合作,在 IBM Cloud Pak for Data 的基础上实施Data Fabric 架构。

在灵活且综合的Data Fabric的帮助下,该企业能够在不同部门实施一系列重要的数据项目,包括:

- 电动汽车
- 天然气业务文档发现,包括笔迹提取
- 用电客户细分和负荷预测
- 资产管理
- 新冠病毒疫情对负荷的影响
- 返岗复工风险模型

有了Data Fabric,该企业可以:

- 通过一个自助式洞察平台为多个业务部门提供直接的数据访问。
- 使用适合的计算能力,在数十亿的海量数据基础上有效运行模型。
- 协作开发模型,并轻松部署,在整个企业融入洞察。

## 减少迁移,改善洞察

某大型工业服务提供商改善数据治理和监管合规性。

该企业试图将海量数据迁移到他们基于云的数据湖。他们有两个要求,一是对每一个数据访问实例应用强大的治理层,二是在提供业务用户访问之前确定数据质量。

通过实施建立在 IBM Cloud Pak for Data 上的Data Fabric,该企业实现数据治理、数据合规性和数据迁移过程的巨大改进。有了Data Fabric,他们可以为业务用户提供对基于云的数据湖和本地 SAP 数据源中数百个数据源的简单、安全的访问。

他们还利用行业监管加速器工具扫描数据源,寻找可能出现的将受到 GDPR 和 CCPA 监管的 PII 数据。自动数据标记有助于识别任何需要管理的数据源,以处理本地和云端 PII。

更好的访问和更快的 PII 筛选帮助企业用户挖掘相关数据,以获得重要的洞察,无需长时间等待数据访问或冒着暴露受保护数据的风险。

有了Data Fabric,该企业可以:

- 确保适当的数据治理,同时利用整个企业的数据。
- 获得可信的数据,减少数据准备的数量。
- 协助遵守隐私法规,如 GDPR 和 CCPA。

## Data Fabric是贵企业的下一步转型

使用建立在 IBM Cloud Pak for Data 技术上的Data Fabric,您可以在混合和多云数据环境中实现数据发现、数据治理和数据使用超自动化。采用Data Fabric使业务用户更快地实现价值,提高数据工程和运营的生产力,并提高治理和合规准确性。

[免费试用 IBM Cloud Pak for Data,开始您的转型之旅。](#)

[安排与 IBM 数据与人工智能代表进行一对一的免费咨询。](#)



© Copyright IBM Corporation 2021

国际商业机器中国有限公司  
北京市朝阳区北四环中路27号  
盘古大观写字楼25层  
邮编: 100101  
美国出品 2020年5月

IBM、IBM 徽标和 IBM Cloud Pak 是 International Business Machines Corporation 在美国和/或其他国家或地区的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。Web 地址 [ibm.com/trademark](http://ibm.com/trademark) 上提供了 IBM 商标的最新列表。

本文档为自最初公布日期起的最新版本, IBM 可随时对其进行修改。IBM 并不一定在开展业务的所有国家或地区提供所有产品或服务。

引用的性能数据和客户示例仅用于演示目的。实际性能结果可能因具体配置和运行条件而异。用户负责评估并验证与 IBM 产品和程序配合使用的任何其他产品或程序的运行。本文档中的信息“按现状”提供, 不附有任何种类的(无论是明示的还是默示的)保证, 不包含任何有关适销、适用于某种特定用途的保证以及有关非侵权的任何保证或条件。IBM 产品根据其提供时所依据协议的条款和条件获得保证。

良好安全实践声明: IT 系统安全性涉及通过防御、检测和响应来自企业内部和外部的不正当访问来保护系统和信息。不正当的访问可能导致信息被篡改、破坏或盗用, 或者导致您的系统遭到误用而攻击别人。任何 IT 系统或产品都不应被认为是完全安全的, 而且没有任何单一产品、服务或安全措施在防止不正当的使用或访问方面是完全有效的。IBM 系统、产品和服务旨在成为合法、全面的安全方法的一部分, 它必定涉及额外的操作程序, 并且可能需要其他系统、产品或服务配合才能获得最好的效果。IBM 不保证任何系统、产品或服务免受任何一方的恶意或非法行为侵扰, 或帮助您的企业免受任何一方恶意或非法行为的攻击。

客户负责确保对适用的法律和法规的合规性。IBM 不提供任何法律咨询, 也不声明或保证其服务或产品将确保客户遵循任何法律或法规。

- 1 Forrester 新技术: 预测总经济影响™研究, 由 IBM 委托进行, 2020年2月
- 2 在IBM硅谷实验室的受控测试环境中, 使用 IBM 数据虚拟化对多种 100TB 数据源进行性能测量。将 2019年5月的测量数据和性能提升与 IBM 其他公司进行比较。
- 3 基于一个 IBM 客户体验。