

专家洞察

IBM 商业价值
研究院

将 Data Fabric 融入混合多云

使用互联数据大展身手

IBM

主题专家



Sandipan Sarkar 博士

IBM 全球企业咨询服务部混合云转型服务线, 全球数据 CTO, IBM 杰出工程师

sandipan.sarkar@in.ibm.com
linkedin.com/in/sandipansarkar

Sandipan 主要负责帮助大型跨国企业成为混合多云环境中的数据驱动型组织。他在人工智能、物联网和区块链等呈指数级发展的技术领域是享誉全球、广受尊敬的家。Sandipan 拥有人工智能博士学位。他经常在科学期刊和商业杂志上发表文章, 并在行业论坛和学术机构发表演讲。



Varun Bijlani

IBM 全球企业咨询服务部
混合云转型服务
全球管理合伙人

varun.bijlani@uk.ibm.com
linkedin.com/in/varunbijlani

Varun 负责领导 IBM 的全球混合云转型服务业务, 帮助客户设计云战略和架构, 并通过迁移、现代化和新的云原生能力来执行战略。他善于将领域知识和全球项目管理专业知识与咨询和行业方面的战略和运营领导能力相结合, 在这方面拥有超过 26 年的丰富经验。他拥有英国西英格兰大学的商业领导力研究生文凭和印度孟买大学的电子工程学士 (BE) 学位。



Richard Warrick

IBM 商业价值研究院云计算
全球研究负责人

rwarric@us.ibm.com
linkedin.com/in/richardwarrick

Richard 负责领导 IBM 商业价值研究院 (IBV) 的云研究计划, 探索云如何作为一种新的工作方式, 推动组织逐步发展成为完全数字化的企业。Richard 在咨询行业拥有超过二十年的经验, 为全球 1000 家客户在业务和技术的交叉领域指引方向。他的工作聚焦于业务 + IT 战略、运营模式设计和实现, 以及大规模的转型计划。他拥有乔治梅森大学组织学习科学硕士学位。

本文贡献者

Subhendu Dey

IBM 高管架构师/副合伙人、
数据和 AI 云顾问

Sujay Nandi

IBM 高管 IT 架构师, 互联解决方案、
物联网和云顾问

要点

显而易见的优点

虽然 Data fabric 技术看起来很复杂，但它的优点显而易见：无论应用在混合多云环境中的何处运行，始终能够在适当的时间为适当的应用提供适当的数据。

边界至关重要

Data fabric 管理着三个关键边界，这些边界对于在分布式混合云环境中运行智能数据非常重要：由数据平台共享的边界、由云和云提供商共享的边界，以及由企业的事务性和分析性数据运营和社区共享的边界。

让所有数据能力发挥作用

投资于人工智能 (AI)、机器学习、物联网 (IoT) 和边缘计算的企业可使用 Data fabric，将数据能力的事务和分析方面“编织”在一起。

扫清价值实现途径中的障碍

要通过实施 Data fabric 以获得商业价值，在很大程度上取决于企业在多大程度上扫清了所面临的非技术性障碍。

Data Fabric 的重要意义不仅限于数据工程师

参与技术驱动的业务转型的每个人都必须理解 Data fabric 的概念。随着大型企业不断与时俱进，以应对持续的竞争压力和新冠疫情等不可预测的事件，Data fabric 可帮助他们“将企业基因传递”到组织未来的发展中。如今，企业在不断扩大的数据和数据分析能力方面持续进行强劲投资，对混合多云架构也是如此。软件正在“占领”整个世界，它们依靠的是智能数据与混合云的强强联合。

IBM 商业价值研究院的研究表明，从数据中获得的越来越多的价值有助于推动实现强有力的数字战略。表现出众的企业对数据人才中心 (CoE)、数据科学家和数据分析工具进行企业规模的投资。¹目前，大多数大型企业都制定了战略，将信息注入面向客户的应用（数字产品），以改善客户体验、支持客户转型之旅以及推出新服务。这些战略适用于 B2B 和 B2C 业务模式，也是新兴平台业务模式的核心。

技术 + 规模 + 人员 = 复杂性。我们往往通过划定事物的边界来管控复杂性，这样可以一次只专注于复杂系统中的一个部分。这种方法在一定程度上是可行的，但我们经常忽略了管理所划分的边界之间空白区域的关键业务。因为 Data fabric 提供了新的方法，用于管理分隔应用、数据、云以及设计和创建它们的人员的边界，所以必须理解这个概念。成功取决于能否有效管理这些边界。

定义：Data Fabric

Data fabric 不仅仅是一种具体的技术，更是一种概念性的方法。可将其视为一个环境，包含架构和一系列统一的数据服务。这些服务共同支持覆盖企业自身网络（本地和多个云环境）的一致数据能力。

“经纬”这个词既有误导性，也有启发性：虽然 Data fabric 本身并不是一个事物，但它的组成部分可以视为具有织物的编制性质，相互连接和维系从而成为整体。

本文探讨了三种边界。首先是数据平台之间的边界。其次是云和云提供商之间的边界。第三是企业的 *事务性*和 *分析性* 数据运营和社区之间的边界。对于从业人员来说，第三种边界可能不像前两种那样熟悉，但是其重要性并不低于前两种。

Data Fabric — 不仅仅是另一种复杂的技术

与数字化业务转型的许多要素一样，Data fabric 技术和架构非常重要，但也相当复杂。通过谷歌搜索“Data fabric architecture”，就可以对其概念有所了解（请参阅边栏“定义：Data fabric”）。这种复杂性与混合云、AI、边缘计算和物联网所推动的变化速度交织在一起，让人们觉得，除了数据科学家、数据工程师、AI 建模师或相关从业人员外，其他人无法掌握 Data fabric 的概念。这意味着我们大多数人可能会对 Data fabric 感到困惑不解。

不要害怕。对于本报告而言，只要理解 Data fabric 是针对反复出现的长期问题的新解决方案就足够了。尽管数据很有价值，而且大型企业拥有海量数据，但这些数据并没有得到妥善管理。数据存储在大小的容器中，分布在企业各处：包括数据中心、数据仓库和数据湖等。需要使用不同的方法以不同的格式存储和检索数据。需要将数据从一处移到另一处，有时这样做速度非常慢，而且难度非常大。数据难以管理，存储成本高昂，并不始终可信，而且只能在限制条件下通过中介访问。

还好，现在我们有了云计算。云计算解决了这一难题，通过将数据转移到云端，降低了存储成本，创造了新的空间，使数据能够在更多的孤岛数据源、端点和由不同云服务提供商创建的封闭环境中猛增。云以新的云端应用的形式鼓励创新，所有这些应用都可能产生更多的数据孤岛。

如果说数据是新的石油，那么大多数大型企业就是喷薄的油井。喷薄而出的数据到处都是，但目前未经“加工”的数据都没有多大价值。Data fabric 就是要解决这个问题。

如果说数据是新的石油, 那么大多数大型企业就是喷薄的油井。Data Fabric 就是要解决这个问题。

Data Fabric：超越数据库

Data fabric 的基本概念是不再将数据库、数据湖、数据仓库或数据集市视为固定的数据储存。相反，它将数据看作在“随时可用”的广泛网络中流动，可在适当的时间为适当的位置和适当的应用提供适当的数据。从用户的角度来看，这个概念确实行之有效。

超越数据库是一种应对“数据重力”的方式——数据在一个地方聚集得越多，就越难移动。Data fabric 可使数据移动更加方便，也可以让无法移动的数据（例如必须保留在本地的数据）更容易访问。

Data Fabric 和因陀罗网

可以借用古老的东方教义因陀罗网来比喻 Data fabric。² 因陀罗是众神之王，他有一张广阔无边的网格，在格子的线上四处都缀有珍贵的珠宝。这张网把无数的珠宝串连在一起，每

一颗珠宝的每一面都反映着其余每一颗珠宝的每一面。万物互联，因果在广阔无垠的网中传播（见图 1，“因陀罗网：Data fabric 的隐喻”）。

在我们的隐喻中，因陀罗网中的珠宝代表了企业中数以千计应用，以及由客户、业务合作伙伴、行业生态系统参与和其他参与方运行的应用。每一颗珠宝都和其他珠宝相连。尽管每一颗珠宝都是独特的，但它们也由这张网（也就是经纬）连接起来。该经纬提供了每颗珠宝所需的一切：计算能力、存储、连接和数据。无论每颗珠宝出现在网中哪个位置，它都能够从最合适、最接近的任意数据源获得所需的数据。

在介绍了 Data fabric 及其要解决的问题后，接下来我们将探讨企业在数据存储中遇到的三个边界，以及类似于因陀罗网的方法如何提供帮助。

图 1

因陀罗网

Data fabric 的隐喻



构建客户的端到端视图所需的数据不太可能位于任何单一数据平台上。

数据平台之间的边界

数据从业人员常说：“一旦添加了第二个数据库，就会出现集成问题。”自从“大数据”成为IT领域的热词王起，通过更有效地使用数据而获利的机遇也越来越多，但与此同时，大规模地为适当的应用提供适当数据方面的挑战也与日俱增。甚至在云出现之前，企业就已经开始构建数据平台：用于集成不同数据库中数据的技术解决方案。数据平台以服务的形式提供。按照适当的规则，需要数据的人可以访问数据，或者将数据提供给用户、应用或其他技术。

每个数据平台的边界通常由所存储数据的类型或使用数据的方式定义。大型企业的特定业务部门可能需要HR数据平台、供应链数据平台或客户数据平台。

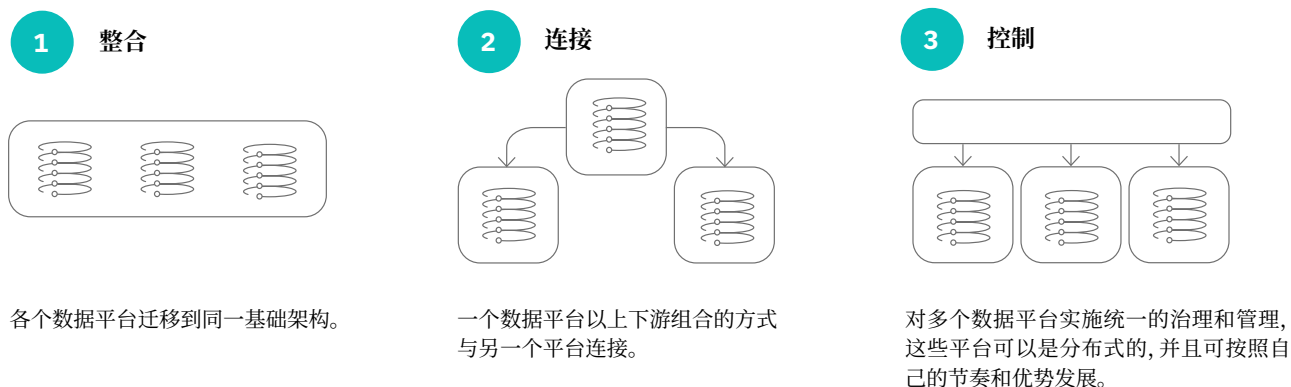
这些边界是从可用数据中获得更多价值的一种权宜之计。但现在，有些变化让事情有所不同。首先，当企业部署新的业务模式并构建客户生命周期互动的单一视图时，按照业务单位划分的客户数据“孤岛”就勉为其难了。其次，随着供应链日益数字化，构建和运行客户价值链的端到端视图所需的数据不太可能位于任何单一数据平台上。第三，随着更出色数据分析的出现，我们可以发现数据平台之间存在连设计师都未曾想到的联系和洞察。

管理数据平台之间的边界有三种方法：整合、连接和控制（见图2：“边界策略：管理数据平台边界的三种方法”）。这些方法都可以是Data fabric的一部分，但最适合分布式混合云环境的是第三个选项。然而，对于每种方法，关键之处在于将平台中的数据与需要数据的应用进行松散耦合。这种松散耦合的概念在整个Data fabric架构中都很重要（请参阅边栏“洞察：向客户进行交叉销售的价值主张”）。

图2

边界策略

管理数据平台边界的三种方法



云和云提供商之间的边界

大型企业采用混合云架构，通常会用到十几个云，涉及各种不同的云服务提供商。这种异构性可能具有优势，但不一定会让共享数据变得更容易。单独的云和云提供商可能是我们在上述数据平台中看到的垂直“孤岛”的另一种形式。

Data fabric 在管理这些边界方面起着关键作用，对于一些架构设计决策也是如此。在这两种情况下，关键是在需要结合的云上应用水平管理层（请参阅边栏“洞察：超越数据进出问题”）。在混合多云环境中，通过 Data fabric 管理这些边界意味着一个云提供商需要成为“同侪之首”，因此该提供商必须采用开放架构。应用 Data fabric 还可能揭示出，我们所管理的一些云可能并不像宣传的那样开放。

事务性和分析性数据运营和社区之间的边界

首先，介绍一些基础知识。大型企业会生成大量 *事务性数据*：也就是在与客户互动过程中创建或交换的数据。目前，我们已经充分了解高效使用和管理事务性数据方面的挑战，尽管这些挑战尚未得到完全解决。例如，大多数企业很难获得客户的单一视图，因为他们使用和交换的事务性数据来自许多不同渠道、应用、产品和服务。

如果不积极进行管理，事务性数据往往会在企业的各个数据孤岛中堆积，妨碍对于提升业务绩效至关重要的数据共享。有多重要？众所周知，亚马逊的杰夫·贝索斯曾威胁要解雇任何不对外公开事务性数据的人。³

第二种数据类型是 *分析性数据*，它们又有所不同。这种类型的数据包括各种非常大型的数据集，用于揭示隐藏的模式，帮助做出预测和决策。大数据浪潮与分析性数据相关；目前，分析性数据已成为所谓的呈指数级发展的技术的基础，这些技术包括 AI、机器学习、物联网、边缘计算和 5G 等。

洞察：向客户进行交叉销售的价值主张

对于围绕不同客户产品发展数据平台的组织而言，用于在数据平台之间协商边界的“控制”模式是通用的。例如，一家大型银行可能有一个用于抵押贷款产品的数据平台，以及一个面向信用卡客户服务的数据平台。向信用卡客户交叉销售抵押贷款（反之亦然）的业务案例非常棒——这是利用大型企业的现有数据实现经济效益的经典案例。

在这个例子中，并不需要合并平台——业务部门也不会同意这样做。从长期来看，仅仅连接平台并不具有可扩展性。控制模式让每个平台都掌握在各自业务单位的手中，而且还支持较大型的企业在管理和资助集成平台的发展过程中发挥自己的作用。

洞察：超越数据进出问题

云之间的边界以一种非常有形的方式表现出来：较高的云服务提供商费用。云服务提供商收取费用的前提是，一旦客户的数据存储在他们的云中，就会一直保存在那里。向云中添加数据可能会增加费用，但从提供商的云中提取数据，并将其转移到另一个云中也非常昂贵。

以前面的银行交叉销售为例，我们假设银行已将其抵押贷款应用交由云服务提供商 A 托管，并将其信用卡应用交由云服务提供商 B 托管。在日常运营中，将数据直接从一个云实时转移到另一个云难度极大，成本极高。凭借 Data fabric，数据将从云 A 移动到作为数据服务的数据平台，不仅云 B 可以访问，而且任何云中的所有请求应用都可以访问。

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_46744

