

专家洞察

IBM 商业价值
研究院

将数据经纬融入 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

虽然数据经纬 (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

数据经纬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数据经纬，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

要通过实施数据经纬以获得商业价值，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。

数据经纬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数据经纬的概念。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，数据经纬可帮助他们“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。因为数据经纬提供了新的方法，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：数据经纬

数据经纬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：虽然数据经纬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。

数据经纬 — 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，数据经纬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：数据经纬”）。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，其他人无法掌握数据经纬的概念。这意味着我们大多数人可能会对数据经纬感到困惑不解。

不要害怕。对于本报告而言，只要理解数据经纬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。数据经纬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。数据经纬就是要解决这个问题。

数据经纬：超越数据库

数据经纬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。数据经纬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。

数据经纬和因陀罗网

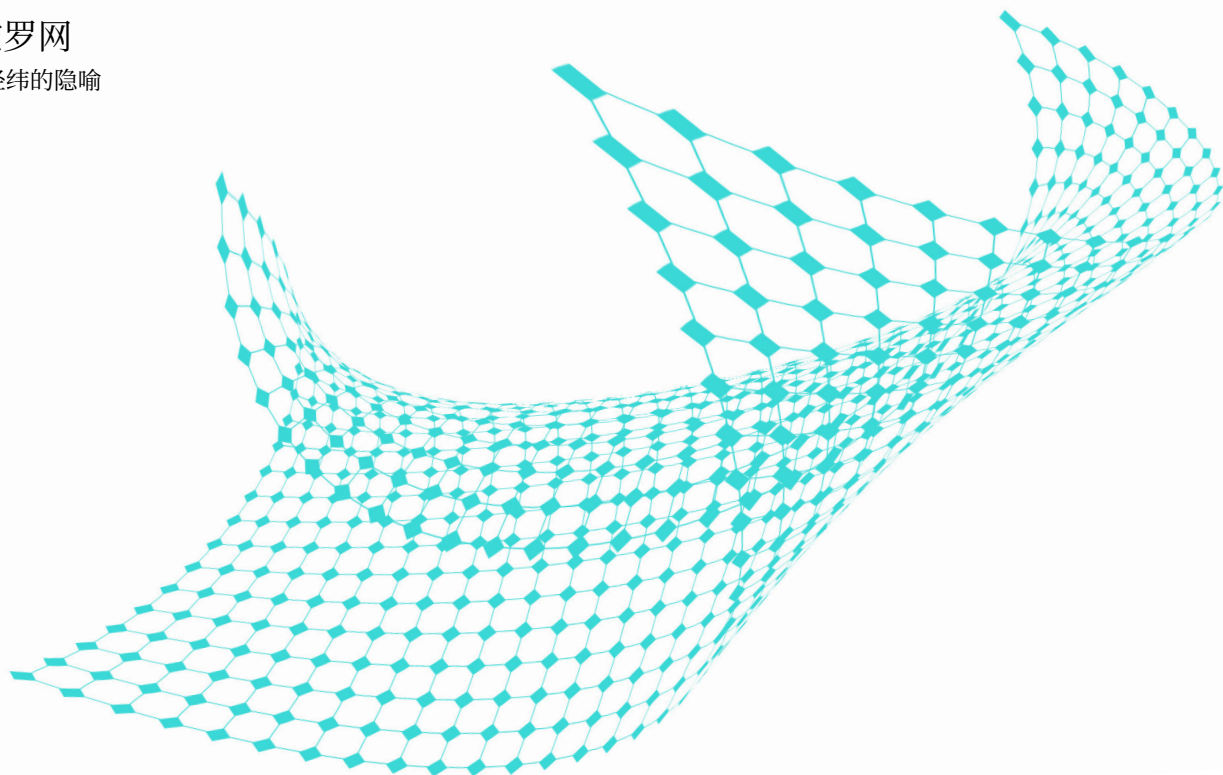
可以借用古老的东方教义因陀罗网来比喻数据经纬。² 因陀罗是众神之王，他有一张广阔无边的网格，在格子的线上四处都缀有珍贵的珠宝。这张网把无数的珠宝串连在一起，每一

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：数据经纬的隐喻”）。

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、行业生态系统参与者和 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。

在介绍了数据经纬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
数据经纬的隐喻



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。

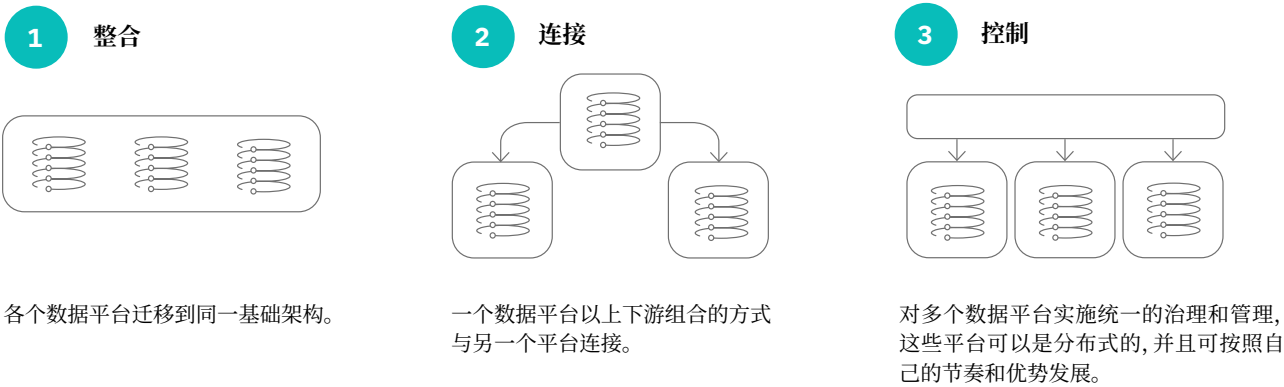
这些边界是从可用数据中获得更多价值的一种权宜之计。但现在，有些变化让事情有所不同。首先，当企业部署新的业务模式并构建客户生命周期互动的单一视图时，按照业务单位划分的客户数据“孤岛”就勉为其难了。其次，随着供应链日益数字化，构建和运行客户价值链的端到端视图所需的数据不太可能位于任何单一数据平台上。第三，随着更出色数据分析的出现，我们可以发现数据平台之间存在连设计师都未曾想到的联系和洞察。

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”）。这些方法都可以是数据经纬的一部分，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数据经纬架构中都很重要（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。

数据经纬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，通过数据经纬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。应用数据经纬还可能揭示出，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。凭借数据经纬，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。

数据经纬可使用 API, 提供直接植入事务性应用或业务工作流程中的洞察即服务。

分析性数据的两个特征对数据经纬非常重要。首先，分析性数据的许多使用场景都需要高速采集大量数据。其次，它可能包含 *非结构化数据*：即无法以数据单元形式存储的数据。视频、图像、音频、文本、社交媒体、传感器信号都是数据，但无法以电子表格的形式捕获。

现代分析需要 *实时或接近实时* 的数据。传统的数据仓库或数据湖无法提供分析所需的这种速度，因为它们需要通过批处理和成本不菲的数据移动才能支持分析。等到洞察产生时，可能已经过时了。

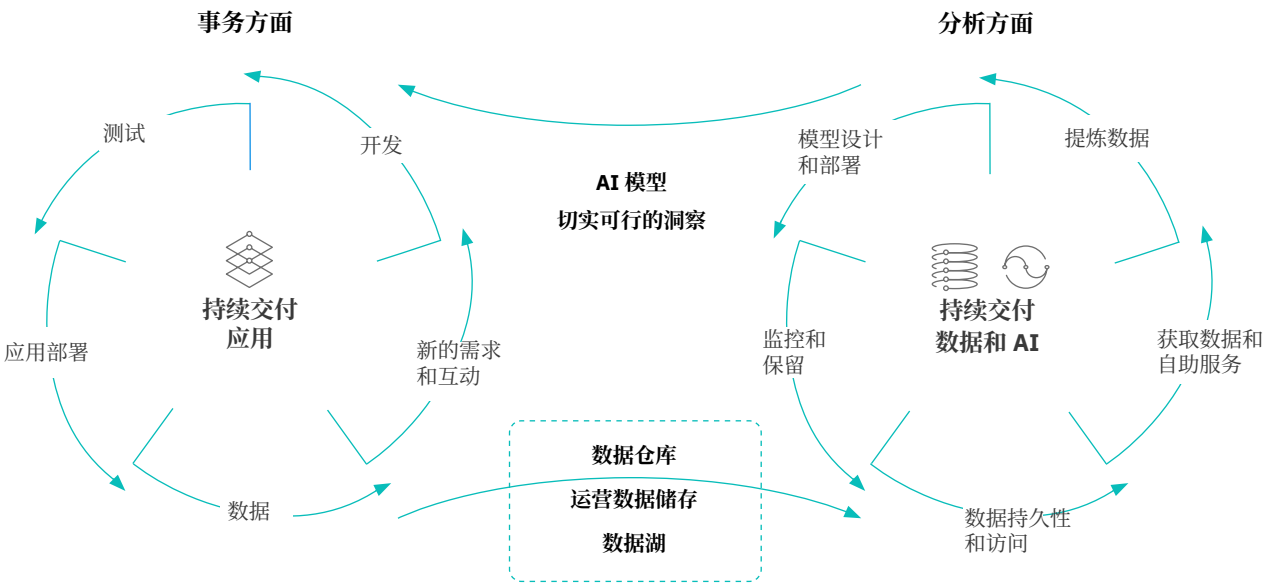
而数据经纬可提供事件流、应用集成和实时分析（甚至可分析移动中的数据），从而解决这一难题。此外，数据经纬提供数据虚拟化能力，支持基于分布式数据源构建分析。当数据源分布在多个云中时，这一点尤为重要，因为在这种情况下，将数据移入移出云和云提供商环境的成本非常高。

数据经纬可使用 API，提供直接植入事务性应用或业务工作流程中的洞察即服务。这意味着数据的用户无需等待对数据仪表板进行成本不菲的人工解读，即可获得切实可行的洞察。在某些情况下，这些洞察可在系统中直接触发自动行动，而无需人工干预。

管理社区边界创造了运营模式机遇

将事务和分析方面结合在一起，可对企业的运营模式产生深远的影响。⁴ 数据经纬通过技术手段拉近事务和分析方面的距离，带来了新的机遇——让应用开发团队以及数据和分析团队开展更紧密的合作。现在，数字产品的开发和管理周期可与生成数据洞察、为 AI 应用模型以及创造新的增值工作流程的周期结合在一起。这两个社区中的团队可针对共同的目标开展合作——向应用用户和客户提供丰富的数据价值（见图 3：“成功的联合：管理事务性和分析性数据社区之间的边界”）。

图 3
成功的联合
管理事务性和分析性数据社区之间的边界



促进应用交付和数据分析社区之间的客户/提供商关系

这就是奇迹发生之处——数据经纬将来自整个混合多云 IT 环境的数据连接起来，使其可用于数字产品。它还使数据人才中心可使用由数字产品生成的数据。由于数据经纬不需要集中式架构，而且它可与分布式异构混合多云环境完美结合，因此有助于实现原本不可能的创新。当数据经纬打破大型企业常见的技术和组织障碍时，就会产生效益。

同样的优点有助于实现覆盖所有平台的数据和分析的统一治理，从而打破更多类型的孤岛（请参阅边栏“洞察：发展提供商和客户关系”）。借助统一的治理，业务部门可使用各个平台，同时还可通过公共的数据治理流程，将各种不同的分析计划联系在一起。例如，通过数据经纬，企业级数据目录和模型注册表可以让企业中的可信数据和分析资产可见并且可用。

数据经纬可能是一种很难理解的技术，但它的优点显而易见：可连接整个企业和企业使用的多个云环境中的不同数据。也许最重要的是，它可以连接企业的应用交付团队与数据人才中心。我们鼓励企业认真研究数据经纬、混合云、应用开发和数据分析相结合的领域。这是一个充满潜在业务价值的领域。

洞察：发展提供商和客户关系

第三次使用这个银行示例，该银行的数字抵押产品和数字信用卡产品都生成事务性数据。在客户群之间开展交叉销售的业务案例基于这样一种想法：抵押贷款应用生成的一些（但不是全部）数据对信用卡业务有用，反之亦然。这两类应用的产品所有者都需要对另一方的数据进行一些分析，甚至可能需要一些像预测性分析或工作流程自动化这样更高阶的能力。为此，需要数据分析社区。

相反，数据分析社区则需要稳定、海量的数据流，在此基础上测试和构建先进的数据模型，并训练机器学习和 AI 产品。从这个角度来看，这两个社区彼此之间自然而然地形成提供商和客户的关系。但实际上，他们之间的互动很少，结果就是两个社区的绩效都不甚理想。

打破事务社区和分析社区之间的障碍是个运营模式问题。这需要我们转变思维：不再将“运行事务性应用”和“从事数据科学”视为不同的领域，而是将二者视为向客户交付价值所需的同一水平工作流的贡献者。

行动指南

扫清业务价值实现途径中的障碍

关于云采用、数据经纬和智能数据的讨论越来越交织在一起，也理应如此。关于管理实践和思维模式变化的类似综合交流却经常被遗漏——也就是不仅仅是实施数据经纬技术，而是要通过实施来获得业务价值。

要超越实施，实现价值，有三种方法。

1. 在投资时，将数据经纬计划视为产品，而不是项目

业务技术计划往往以项目的形式获得资金支持，这些项目建立、按照预先设定的要求交付，然后结束，所有权转移给 IT 部门以进行维护和支持。对于数据经纬这样的计划，这种方法行不通。

将数据经纬计划作为产品进行投资和管理更为合适，这样的投资反映了产品管理的生命周期。我们不应该只是实施技术，在取得任何业务成果之前就宣布大功告成；根据产品方法，形成数据经纬的工作版本只是开始。⁵提供资金和配备人员是一项需要持续多年的工作。

这种方法可能很困难，因为大多数大型企业仍在实施数字和云出现之前的那种投资流程和政策。它们通常弹性极大，防御性极强。一些企业并没有对传统的投资方案直接进行现代化，而是直接放弃当前的政策，取得了更理想的结果。

2. 采用敏捷实施原则

没有放之四海而皆准的数据经纬方法。每个企业都需要去探索如何让数据经纬为我所用。当数据经纬工作与云采用工作或大规模 AI 实施工作同时开展时，尤其如此。系统某一部分的改变可能会导致其他部分的改变，而且通常是以无法预料的方式发生。

通过采用敏捷原则，例如构建一点、测试一点、学习一点，有助于应对构建数据经纬时无法避免的复杂性。使用小型跨学科团队一同构建、集成和持续运行更完整、可扩展的数据编制版本，实现最理想的成果。在与云、AI 等计划同时运行的情况下，最好建立一个覆盖所有计划的产品办公室来支持集体团队，帮助整合团队的工作。

采用敏捷实施原则可能很困难，因为许多大型企业仍习惯于自上而下的命令与控制管理原则。在所有人都不太清楚数据经纬的工作方式时，在计划的一开始就制定固定的计划、时间表和里程碑不是什么好主意。即使拥有适当的人才、适当的技术以及像参考架构这样的有用工件，但这个计划以前从未实施过仍是不争的事实——从未在这个业务环境中与这个企业中的这些人一起完成过。

3. 将数据构造作为服务运行

理想情况下，数据经纬在后台运行，而用户（例如开发团队）并不需要知晓它的工作方式和内容。所以开发团队不必聘用全职的数据经纬专家。数据经纬最好作为服务提供，这样用户在响应客户请求的过程中使用服务，有时甚至不知道自己正在使用数据经纬。在服务驱动的业务技术模式中，数据经纬可完美地融入向运营模式中其他领域的各种“客户”提供的一系列基础的日常服务中。

数据经纬作为客户使用的服务的优点之一，就是能够自动保持目标上数据经纬的范围和规模。数据经纬服务随着需求的扩大而扩大；随着客户驱动的需求的缩减或变化而缩减或变化。

将数据经纬作为服务来运行难度较大，因为传统的 IT 组织是围绕垂直的职能领域而构建的，而服务驱动的运营模式最终需要改变为与以客户为中心的igital 产品交付流一致的水平结构。

关于专家洞察

专家洞察代表了思想领袖对具有新闻价值的业务和相关技术主题的观点和看法。这些洞察是根据与全球主要的主题专家的对话总结得出。要了解更多信息，请联系 IBM 商业价值研究院：ibvchina@cn.ibm.com

选对合作伙伴，驾驭多变的世界

在 IBM，我们积极与客户协作，运用业务洞察和先进的研究方法与技术，帮助他们在瞬息万变的商业环境中保持独特的竞争优势。

IBM 商业价值研究院

IBM 商业价值研究院 (IBV) 站在技术与商业的交汇点，将行业智库、主要学者和主题专家的专业知识与全球研究和绩效数据相结合，提供可信的业务洞察。IBV 思想领导力组合包括深度研究、专家洞察、对标分析、绩效比较以及数据可视化，支持各地区、各行业以及采用各种技术的企业做出明智的业务决策。访问 IBM 商业价值研究院中国网站，免费下载研究报告：<https://www.ibm.com/ibv/cn>

© Copyright IBM Corporation 2021

IBM Corporation
New Orchard Road
Armonk, NY 10504
美国出品
2021 年 4 月

IBM、IBM 徽标及 ibm.com 是 International Business Machines Corporation 在世界各地司法辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的注册商标。以下 Web 站点上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表：ibm.com/legal/copytrade.shtml。

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。

本文档内的信息“按现状”提供，不附有任何种类（无论明示还是默示）的保证，包括不附有关于适销性、适用于某种特定用途的任何保证以及非侵权的任何保证或条件。IBM 产品根据其提供时所依据的协议条款和条件获得保证。

本报告的目的仅为提供通用指南。它并不旨在代替详尽的研究或专业判断依据。由于使用本出版物对任何企业或个人所造成的损失，IBM 概不负责。

本报告中使用的数据可能源自第三方，IBM 并未对其进行独立核实、验证或审查。此类数据使用的结果均为“按现状”提供，IBM 不作出任何明示或默示的声明或保证。

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

备注和参考资料

- 1 “From Data Science to Data Diplomacy.” IBM, 2019, www.ibm.com/thought-leadership/institute-business-value/c-suite-study/cio?mhsrc=ibmsearch_a&mhq=data+diplomacy.
https://www.ibm.com/thought-leadership/institute-business-value/c-suite-study/cio?mhsrc=ibmsearch_a&mhq=data%20diplomacy; Giordano, Tony. “A Blueprint for Data in a Multicloud World.” IBM, 2019, www.ibm.com/thought-leadership/institute-business-value/report/multicloud-data-strategy
- 2 O'Brien, Barbara. “Indra's Net: Buddhist Metaphor for the Interpenetration of All Things.” Learn Religions, www.learnreligions.com/indras-jewel-net-449827
- 3 Settle, Mark. “CIOs Can Weaponize Their API Architectures Through Better Governance.” Forbes, Forbes Magazine, 1 June 2020, www.forbes.com/sites/marksettle/2020/05/04/cios-weaponize-apis-through-better-governance/?sh=27972aee6076
- 4 “How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh.” Martinowler.com, martinowler.com/articles/data-monolith-to-mesh.html; “Data Mesh Principles and Logical Architecture.” Martinowler.com, martinowler.com/articles/data-mesh-principles.html
- 5 Kersten, Mik. “Project to Product: How Value Stream Networks Will Transform It and Business.” IT Revolution PR, 2018. <https://itrevolution.com/book/project-to-product>

扫码关注 IBM 商业价值研究院



官网



微博



微信公众号



微信小程序