



专家洞察

多云世界 数据蓝图

IBM 商业价值研究院

IBM

主题专家



Tony Giordano

CBDS 数据平台服务
高级合伙人兼副总裁
IBM Services 全球主管
[linkedin.com/in/
tony-anthony-giordano-3111091](https://www.linkedin.com/in/tony-anthony-giordano-3111091)
tonygior@us.ibm.com

Tony 在绩效营销、大数据和分析、信息监管、客户关系管理和项目管理方面拥有超过 25 年的全球专业服务经验，并且在金融服务、生命科学、零售和汽车等领域拥有丰富的行业经验。他撰写过两本关于信息管理的书籍。



Mehdi Charafeddine

IBM Services
副合伙人，
解决方案主管架构师
[linkedin.com/in/mehdichara](https://www.linkedin.com/in/mehdichara)
mchara@ibm.com

Mehdi 是一位技术领导，在实施大型复杂项目方面声誉卓著，他善于从全新角度看待问题，运用新方法解决问题。Mehdi 还是开源技术的忠实拥护者，不断探索对客户具有潜在价值的新技术。这既需要对软件有深入的了解，还要准确把握 IBM 客户能够接受的风险 / 回报率。



Dan Sutherland

IBM Services 数据平台
杰出工程师，首席技术官
[linkedin.com/in/
dan-sutherland-4661a67](https://www.linkedin.com/in/dan-sutherland-4661a67)
dsutherl@us.ibm.com

Dan 擅长帮助客户将战略业务愿景转变为创新解决方案，同时，还不断探索边缘技术。他是解决复杂数据架构和设计问题的能手，定期提供咨询服务，对象主要是以海量数据为中心的大型项目，通过建议项目的关键组成部分，即坚实可靠的数据架构和设计方法，确保成功实施解决方案。



Becky Carroll

IBM Services
认知业务决策支持，
认知与分析
副合伙人
[linkedin.com/in/beckycarroll](https://www.linkedin.com/in/beckycarroll)
rlcarroll@us.ibm.com

Becky 在帮助大型企业的业务领导将传统业务流程转变为新业务模式方面，拥有超过 25 年的丰富经验。她曾领导过多个项目，涵盖大数据和分析、认知和人工智能、数字和社交业务战略与分析、CRM 以及客户服务等领域。Becky 的工作涉及众多行业，包括高科技、电信、娱乐与媒体、金融服务、航空公司、汽车、工业品、生命科学、游戏、教育等。Becky 拥有电子工程和计算机科学的学士学位以及营销管理的工商管理硕士学位。

扫码关注 IBM 商业价值研究院



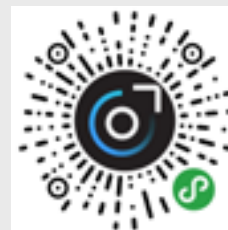
官网



微博



微信



微信小程序

计划周详、执行有力的数据战略有助于避免多云环境中出现令人措手不及的意外情况。

—

谈话要点

管理多个云环境中的应用和数据的数据战略和架构

对于当今在多云环境中运营的企业而言至关重要。

业务驱动的数据战略应当与企业不断发展的多云架构全面整合

防止出现数据孤岛，避免应用难以访问。

强大的数据战略应专注于业务转型机遇

避免被任何特定的云“锁定”，并使用 DataOps 方法自动为数据科学家执行准备工作，从可用数据中获得尽可能多的信息。

在多云世界中移动数据

随着越来越多的企业从初步尝试转为深入采用云技术，他们都体验到了多云环境的优势和挑战。多云环境的优点显而易见：有助于提升创新能力，改进产品和服务，并实现敏捷的业务流程。¹

事实上，IBM 商业价值研究院 (IBV) 最近的一项调研表明，全球 85% 的受访企业已经在多云环境中开展运营。² 到 2021 年，96% 的受访企业计划使用多个云环境。³ 众多企业正在将其自身重塑为“数字化”企业，将人工智能 (AI) 置于重要业务流程的核心。随着数字化的不断推进，业务重心会逐渐向多云环境偏移，因为通常需要从不同的云供应商那里获得基于云的服务。但是这种迁移绝不仅仅是将现有架构迁移到云端的新位置那样简单。

每个云供应商都有自己的架构方法。每个供应商组织数据的方法可能与其他云环境并不兼容。如果精心设计的战略没有考虑到解决这种不兼容问题，那么企业可能会遇到性能不理想以及成本高于预期的情况。

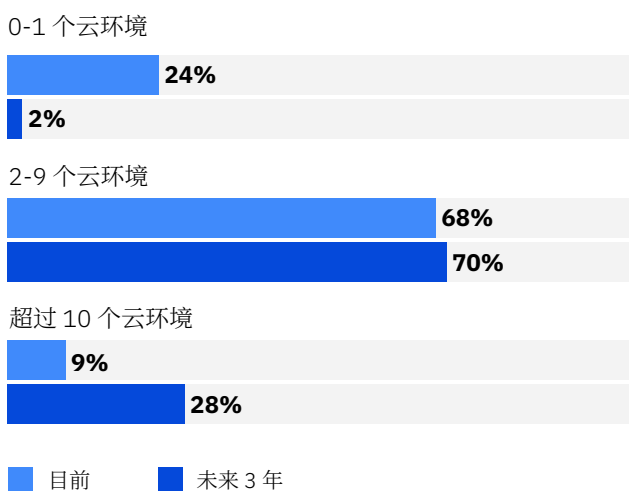
挑战的根源在于数据的性质及其快速增长。当企业增加所使用的服务和应用的数量时，数据会呈指数级增长。但真正的问题并不在于数据的数量，而在于数据所在的位置。数据所在位置通常取决于收集和创建数据的位置。

Data Fabric 有助于在整个企业范围实现更理想的数据可视性和可用性。

这种由不同的业务单位或职能部门创建的分散数据被称为“数据孤岛”。孤岛的存在严重影响了数据的易用性。数据可能重复，但以不同的方式组织，因此很难相互关联。由于收集数据的方式不同，可能导致数据出现不一致。企业中其他部门的应用可能不知道这些数据的存在，或者无法访问这些数据。

由于几乎所有企业很快都将开始多云环境运营，因此随着企业将流程、应用和数据包分散到多个云环境中，内部数据孤岛造成的负面影响将会不断加剧（见图 1）。

图 1
98% 的受访组织表示，他们计划在未来三年内使用混合云环境



来源: IBV 2018 年多云管理调研, Q2 和 Q3。

Q2 你们整个企业目前使用多少个云服务和平台?

Q3 您计划未来三年内在整个企业中使用多少个云服务和平台?

这些新情况引出了一个重要问题: 企业如何才能成功地应对多云环境运营所带来的复杂性、高成本和延迟问题?

答案是什么呢? 那就是在多云世界中定义“Data fabric”, 也成为“数据蓝图”。

Data Fabric 帮助企业轻松驾驭多个云环境

Data fabric 是数据资产组织方式的概念性表示或架构。这种蓝图是正式的结构, 用于在整个企业中定义和查看数据, 并且独立于任何特定的基础架构或云计算需求。

Data fabric 有助于在整个企业范围实现更理想的数据可视性和可用性。它用于指导如何维护和监管数据。它基于这样一个概念: 业务流程可能随时发生变化, 但支撑它们的数据一直保持稳定。

数据被分组到称为“中心”或“湖”的集合中, 以实现可视性和可访问性。Data fabric 有助于避免或缓解数据孤岛、低可靠性和低可扩展性、依赖于原有系统以及成本效率低下等问题。但是要在多云环境中有效开展工作, Data fabric 必须能够适应不同的使用模式。

下面我们举三个例子, 展示企业如何在不同的云环境中分配工作负载、应用和数据。

示例 1

在第一个例子中，一家公司希望使用 AWS 中已有的数据来运行 IBM Cloudpak 中提供的分析包。它的数据采集和管理服务

位于一个云环境中，只有一个特定的分析服务从另一个云环境调用这些数据（见图 2）。

图 2

通过架构组件在整个多云环境中管理数据



要在多云环境中有效开展工作，Data Fabric 必须能够适应不同的使用模式。

示例 2

在第二个例子中，企业既被 Azure 上低廉的应用价格所吸引，同时又希望其他现有流程保留在 AWS 上。因此，该公司的原始数据采集功能以及数据统一和使用功能被分散

到不同的云环境中。该公司还订购了 IBM Cloudpak 上的分析服务（见图 3）。

图 3

通过数据中心层在整个多云环境中管理数据



示例 3

在第三个例子中，该公司出于独特的需求，要求各个业务部门在不同的云环境中运营，但同时选择将涵盖

所有其他云环境的关键跨业务职能（例如，安全）保留在 Azure 中（见图 4）。

图 4

管理多个云环境中的数据



合适的数据战略能够预测变化，帮助企业保持灵活性。

正如上面三个例子所示，这种新的多云现实创造了更多新的选择和机遇。但同时也给 IT 领导带来了新的挑战。近期市场情报显示，82% 的 IT 领导担忧如何将所有这些云环境与传统 IT 环境衔接。⁴ 73% 的受访者表示，他们需要采用更好的方式，更有效地在多个云环境中移动应用、工作负载和数据。67% 的受访者担心如何以一致的方式管理这种新的多厂商云与环境组合，而不会带来新的安全与合规风险。⁵

为了应对这些挑战，我们推出了三项指导原则，帮助企业制定和执行数据战略，以便顺利迁移到多云环境并在其中运营。

原则一：业务转型之时，向云迁移之日

我们的数据显示，降低成本仍然是对多云环境投资的重要战略考量因素。然而，一些企业在从多云投资中实现成本节省方面遇到了困难，迟迟无法收获成果。根据我们与客户合作的经验，许多企业都面临着成本上升的问题，在某些情况下，甚至高达 300%。为什么会这样？

迁移到云端会放大原有 IT 格局的现有缺陷。云供应商对数据传输和网络使用的收费比大多数内部 IT 组织要高得多。因此，孤岛型数据（大多数原有系统无法避免的现实）在从一个云环境迁移到另一个云环境以供各种应用使用时，会产生巨大的成本和性能问题。

因此，向云端迁移之时，需要审视当前的业务流程和数据管理，评估哪种改变方式有助于控制成本以及最大程度发挥新多云环境的潜力。这不是简单的 IT 讨论或者重组当前数据的问题。业务部门和 IT 部门必须密切协作，一切以业务需求和价值为导向，并由业务部门高管负责牵头。

但也存在一个挑战，那就是快速变化的技术格局：今天适用的工具可能明天就毫无用处。此外，转型机遇还意味着，迁移原有系统可能需要采用新技术，而不是继续重复使用现有技术，比如 Hadoop。

原则二：预测并规划未来发展、变化和创新

在迁移到多云环境时，企业应该避免“供应商锁定”。云提供商和产品在不断快速变化，他们所支持和提供的技术也是如此。因此，最适合企业的产品或服务也会随着时间的推移而不断改变。

合适的数据战略能够预测变化，帮助企业保持灵活性。在某种情况下，企业可能希望将数据和应用从一个云环境迁移到另一个，扩展存储和处理能力而不产生负担，并将数据存储在最适合数据科学和工作负载的位置。

为了预测变化，企业的数据战略必须包含三个关键组成部分：容器化、无服务器能力和实用的数据设计。

“容器化”是指企业将数据应用打包成不依赖于单个云实施的“容器”。容器化的应用可在多个云环境和操作环境中运行。这样，企业就能够随着生态系统的发展而迁移应用和数据，特别是在各种流程越来越多地跨云互动的情况下。企业也不会完全依赖于特定的云来运行业务的某些部分。

变化的另一个方面是新应用和服务生成的海量数据。企业基本上不可能扩大数据管理团队的规模以及增加与存储和处理数据相关的成本。而无服务器战略可以提供帮助。

企业不需要保留一定规模的资源（例如，1000 台服务器），而是由云供应商负责根据数据和使用需求的逐渐增加、使用高峰或减少而缩放资源。企业为所使用的资源付费，可以将主要精力放在支持运营所需的业务案例和代码上，而无需关注运行所需的 IT 资源。无服务器战略与容器化一起，有助于降低新部署在复杂性、技能和成本方面的风险。这种方法帮助企业既不受云供应商的限制，也不受技术的限制。

预测变化的第三个要素是适当的数据结构设计。在为传统报告、数据科学、数字化和运营用例等设计数据结构时，务必将工作负载部署在靠近相应数据的位置，以便减少网络流量并降低成本。数据设计还应包括涵盖多个云提供商的数据策略，以确保安全性、合规性以及数据生命周期管理。这些数据设计要素可整合为一系列统筹管理的数据工作流程，覆盖多个云提供商。

Yara：数据和多云数字农业平台

总部位于挪威的 Yara 是全球最大的化肥生产商之一，该公司的愿景是帮助全世界以可持续的方式消除饥饿。为了通过数字化引领核心业务模式创新，Yara 着手构建世界领先的数字农业平台。

在构建该平台的过程中，Yara 集中精力制定并实现独立于云平台的战略，确保能够对所有环境实施一致的数据监管和数据安全策略。该公司还专注于 DataOps，自动执行数据功能，让数据科学家能够专心从事数据模型和创新工作。

该平台为全球农户提供整体数字服务和即时农事建议，通过增加现有农田的粮食产量，最终避免毁林开荒。Yara 数字平台的目标是覆盖全球 7% 的耕地。

独立于云平台的数字农业数据平台采用按使用量付费的商业模式，并且为 Yara 提供两项数据服务：Weather Data 和 Crop Yield as a Service。这些加速器是众多创新中的第一批；开放创新层使 Yara 能够创建突破性的新算法，为农户提供知识和决策洞察。

原则三：将 DataOps 添加到 DevOps

企业应当借鉴 DevOps 如何在过去十年左右给应用开发带来了革命性的变化。部署自动化可以帮助开发人员加速测试和证明自己的工作。数据科学家现在需要一场类似的革命，他们发现 DataOps 就是自己所要的：让拖慢进度的工作实现自动化。

目前，数据科学家需要花费大量时间准备、验证和清理数据源，然后使用这些数据源训练数据模型。他们花在数据模型设计上的时间少得惊人，而这是数据科学家工作中价值最高的部分。如果能够自动执行数据准备和训练工作，那么就可以将数据科学家解放出来，从而为企业带来更多洞察，最终实现新的价值。

数据战略、文化与人员

如果没有阅读文化，即使汗牛充栋，也毫无用处；同样道理，有了海量的数据和最具创新性的数据工具，仍需具备适当技能的人来使用它们，而企业文化必须提供这方面的支持。人们需要数据以及从中获得的洞察，也需要在工作的特定环境中使用数据。相同的数据可能对不同的用户有着不同的意义。科学家认为有用的洞察，可能与开发人员、产品经理、营销人员或流程专家需要的洞察完全不同。

特别是当人工智能应用于超大数据集时，意义和上下文就变得至关重要。应该提出的正确问题是什么？这些是用于提供答案的正确数据吗？如果这是正确的数据类型，我们能否信任它？我们能否相信针对该数据运行的人工智能算法，以及人工智能正在接受的训练（包括它正在接受的实际训练数据）？

成功的数据战略应包含所需技能和培训的清单，以及旨在培养充满活力的数据文化的长期计划：激励人们从数据中获得最大价值，并且有能力这样做，企业也支持这样做。

如何开始

制定与多云计划整合的数据战略并应用以上列出的原则，这看似颇具挑战性，但可采取一些步骤加以应对。

1. 制定业务转型战略，明确考虑并利用多云能力，有效管理数据以充分发挥这些能力的作用。
2. 考虑多种技术，确保让数据存储位置靠近使用位置，并且在多云环境中妥善监管数据。
3. 确定耗费数据科学家大部分工作时间的低价值任务，并使这些任务实现自动化。这样就可以将数据科学家解放出来，有更多时间从事更有价值、更具战略意义的工作。

— 准备好了吗？

- 贵企业对哪些核心业务流程进行了优先级排序并研究了转型可行性？
- 贵企业采取了哪些措施以营造适合多云环境的数据文化？
- 贵企业将如何应对与位于不同云环境中的数据相关的法规？

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：iibv@us.ibm.com。

备注和参考资料

- 1 IBV survey Assembling your cloud orchestra: A field guide to multicloud management (2018). Steve Cowley, Lynn Kesterson-Townes, Arvind Krishna, Sangita Singh. <https://www.ibm.com/downloads/cas/EXLAL23W>
- 2 Ibid.
- 3 Ibid.
- 4 IDC Cloud Forecast 2018-2020 – Market sizing + CAGR; BCG & McKinsey Study conducted for IBM – Multicloud + Priority concerns.
- 5 Ibid.
- 6 Source: IBM Institute for Business Value hybrid cloud survey (2016).
- 7 Internal IBM measurement, IBM Global Business Services.
- 8 IBM client experience

© Copyright IBM Corporation 2019

IBM Corporation
New Orchard Road
Armonk, NY 10504
美国出品
2019 年 10 月

IBM、IBM 徽标、ibm.com 及 Watson 是 International Business Machines Corp. 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 可随时对其进行更改。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（无论是明示的还是默示的）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组织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 并不独立核实、验证或审计此类数据。此类数据的使用结果均“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

