

# IBM Cloud Pak for Data 上的 Db2

基于 Red Hat OpenShift 构建的集成式多云数据平台

# 上下文中的数据

## 目录

- 2 上下文中的数据
- 4 将容器用于AI 数据库
- 5 Cloud Pak for Data 上的 Db2 的价值
- 5 在 Cloud Pak for Data 上部署 Db2 的三个主要好处
- 5 在单个统一平台上启用混合云和多云
- 6 集成和管理简便
- 8 无需移动数据的跨多个数据源查询
- 9 后续步骤

数据是企业最重要的资源之一。但其价值取决于能否以易于理解的方式全面呈现数据。人工智能 (AI) 技术承载着更快地挖掘更深洞察力的巨大希望 - 加速发现的步伐, 拓宽可利用的数据范围, 自动执行以前需要人类专业知识的任务。AI 仅在全部数据值得信赖、可以访问并且符合标准时才能发挥效用。同时, AI 的兴起也突显了数据系统中长期存在的弱点和局限性 - 呼唤着新的策略。

现在您可以挑战各种数据来源和类型: 传感器日志、文本文档、图像、录音、地理位置数据、情绪数据等。现代分析和 AI 驱动的工具可从企业数据中发现前所未有的洞察力, 但如果您限于一种数据甚至一种数据库系统, 就会一叶障目。

这意味着, 虽然企业仍然持续需要强大、安全、可扩展和高性能的数据库, 但这些数据库不能孤立存在。必须将它们集成到功能强大的数据平台。

考虑企业中不同数据角色之间的典型数据流 (图 1)。每个角色都有不同的重点:

1. 管理人员制定目标和战略
2. 企业架构师使技术与企业目标和战略一致
3. 数据工程师构建并维护数据管道
4. 数据科学家构建并改进分析和 AI 算法
5. 开发人员将这些洞察力嵌入到实用的应用程序中

但以上所有角色都需要丰富、优质、可信的数据。利用这些数据, 他们可以改进决策、加速创新、优化客户体验、推高运营效率以及获得其他优势。

应用程序开发人员	数据工程师	数据管理员	数据科学家	企业管理人员	企业架构师
<b>收集</b>		<b>组织</b>		<b>分析</b>	
<ul style="list-style-type: none"> <li>- 数据虚拟化</li> <li>- 数据仓储</li> <li>- 按需数据库</li> <li>- 数据源提取</li> <li>- 分布式处理</li> </ul> <b>采用:</b> Db2 和 Db2 Warehouse 技术		<ul style="list-style-type: none"> <li>- 发现和搜索</li> <li>- 数据转换</li> <li>- 数据编目</li> <li>- 业务术语</li> <li>- 策略、规则和隐私</li> </ul> <b>采用:</b> InfoSphere、DataStage 和 InfoSphere Information Governance Catalog 以及 Watson Knowledge Catalog		<ul style="list-style-type: none"> <li>- 数据可视化</li> <li>- 机器学习</li> <li>- 模型构建和部署</li> <li>- 模型管理</li> <li>- 仪表板</li> </ul> <b>采用:</b> Watson Studio 开源和 Cognos	
<b>多云服务</b>	<ul style="list-style-type: none"> <li>- 日志记录</li> <li>- 监控</li> </ul>	<ul style="list-style-type: none"> <li>- 计量</li> <li>- 持久存储</li> </ul>	<ul style="list-style-type: none"> <li>- Kubernetes</li> <li>- 安全性</li> </ul>	<ul style="list-style-type: none"> <li>- 身份访问管理</li> </ul>	<ul style="list-style-type: none"> <li>- Docker 注册表/Helm</li> </ul>

## IBM Cloud Pak for Data

图 1. 企业设置中数据角色之间的典型数据流动

对于想利用 AI 技术的企业，对数据系统的需求特别强烈。由于 AI 技术可以扩展和加速数据的传统商业应用，因此在使用 AI 驱动的智能时，纯粹的历史追溯系统与自动发现洞察力并做出预测的主动式系统会产生明显的差异。机会无限：2021 年，AI 增强预计会产生 2.9 万亿美元的商业价值，提高 62 亿小时的人工效率。<sup>1</sup>

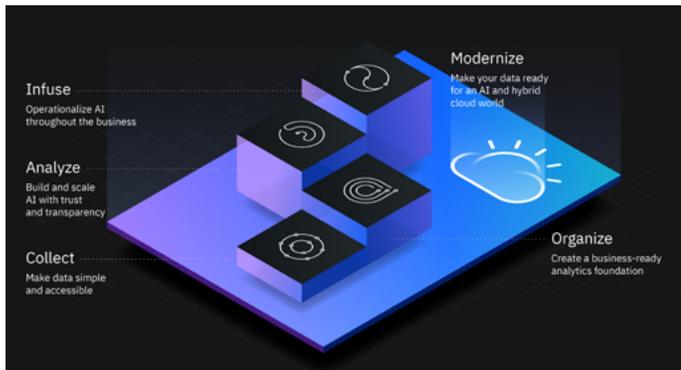


图 2. AI 阶梯：通往成功的 AI 计划的五级阶梯

但高效的 AI 需要最高质量的信息架构 (IA) 支持。这正是 IBM 开发 AI 阶梯（请参见图 2）的原因，AI 阶梯从独特的角度阐述企业如何准备和实施 AI 技术。企业必须实现其数据平台的现代化；收集、组织、分析其数据；将 AI 融入整个企业。

所有这些因素 - 从数据种类到数据用户需求再到 AI 的兴起 - 促成了数据库部署和使用方式的重大转变。

数据库是 AI 之旅的关键，因此 IBM® Db2®（结合 BLU Acceleration® 内存计算改进性能，结合 pureScale® 技术提供分布式和高度的扩展性）仍然与您的整个数据战略高度相关。

但通过将这些数据库技术与数据平台的其他部分无缝集成，企业可以全面而自动地配置数据，同时保持所需的性能、安全性和治理。容器化架构，尤其是部署在云平台（如 IBM Cloud Pak® for Data）上的容器化架构，是这种转变的关键。

# 将容器用于 AI 数据库

IBM 之所以开发 IBM Cloud Pak for Data，是因为市场需要快速、稳健、强大的数据平台来支持 AI 计划，这个数据和 AI 平台旨在管理整个企业对数据的自助使用。Cloud Pak for Data 是一个混合云和 AI 平台，为 AI 提供信息架构。通过 Cloud Pak for Data 可以在统一的云原生平台上解锁所有数据的价值，自动完成企业将数据转变成洞察力的过程。

通过 Cloud Pak for Data 部署 Db2 后，企业可以利用那些使 Db2 成为行业领先 2 数据平台的所有强大功能。作为 [AI 数据库](#)，Db2 针对 AI 构建，并且受 AI 支持。

## 针对 AI 构建

- 使用自然语言查询更快地探索数据
- 使用流行的语言和代码示例构建 AI 应用程序
- 使用 SQL 数据图表建模复杂的关系
- 本地分析区块链数据

## AI 驱动

- 数据虚拟化
- 自适应工作负载管理和资源优化
- 机器学习查询优化
- 基于置信度的查询

Cloud Pak for Data 上的 Db2 是 Db2 Advanced Edition 的容器化版本，基于微服务架构，已针对 Kubernetes 进行优化。通过 Cloud Pak for Data，Db2 可在世界领先的容器协调平台 Red Hat OpenShift 上本地运行。

将 Db2 数据库功能分解为微服务而非单体栈，您可以获得多种机会：

- 数分钟完成部署；支持标准部署和管理，同时保持按需修改参数的灵活性。
- 具有开箱即用的 enhanced Kubernetes 可用性，并且支持高可用性/灾难恢复(HA/DR)自动故障切换，因此可靠性高。
- 自动化更新减轻管理负担。一键部署服务包、版本和模式。
- 通过“应用程序组”自动化管理；管理员可以使用名称空间来管理访问控制和配置选项。
- 借助平台和服务级功能在应用程序级监控和管理。
- 独立扩展微服务，以应对不断变化的需求。

容器化数据和 AI 平台可让您将 Db2 运行于混合云环境（云与非云平台的组合）或多云环境（来自不同提供商的云）的组合，以对每种数据类型使用适当的基础架构。

最近容器之所以流行，正是得益于这些优势。根据“Red Hat 2019 全球客户技术概况”，57% 的企业已经在使用容器，预计容器的使用率在接下来 2 年还会增长 89%。<sup>3</sup>

但选择 Cloud Pak for Data 作为容器化架构特定版本的优势远不止于此。利用 Cloud Pak for Data 与 IBM 产品组合中其他元素的全面集成，您还可以轻松地利用其他数据功能，例如使用 IBM Watson® 的机器学习；使用 Watson Knowledge Catalog 和 IBM InfoSphere® DataStage® 的数据运营；以及使用 Watson Studio 和 Watson OpenScale™ 的集成数据科学。借助 Cloud Pak for Data 上的 Db2，您可以更轻松地全面使用 IBM 服务的专业知识，以设计、部署和管理有助于您实现商业价值的高级分析。

# Cloud Pak for Data 上的 Db2 的价值

如图 3 所示，Cloud Pak for Data 上的 Db2 结合了 Db2 容器化微服务架构以及多个额外的功能层：

## 1. 服务层

借助 Db2 的容器化微服务版本，访问 IBM 和第三方的逾 45 项分析服务和模板。

## 2. 数据虚拟化功能

快速而轻松地查询多个数据源，无需移动数据。

## 3. 平台接口层

通过集成数据管理、数据治理和分析的单一用户体验快速实现价值，提高效率并改进资源的使用。

## 4. Kubernetes 层

利用领先的混合云、企业容器平台 Red Hat OpenShift 实现创新、快速的部署策略。

## 5. 基础架构层

通过多云方法避免锁定，利用任何云基础架构。

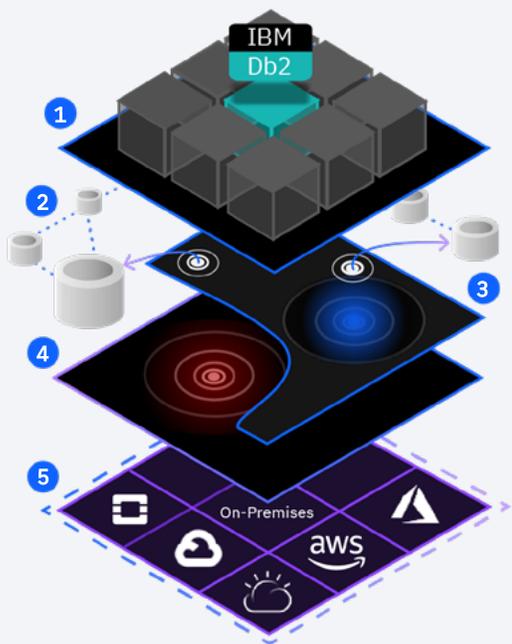


Figure 3. Db2 on Cloud Pak for Data

# 在 Cloud Pak for Data 上部署 Db2 的三个主要好处

## 01

### 易于在单个统一平台上启用混合云和多云

Cloud Pak for Data 上的 Db2 可让企业更轻松部署和管理混合云及多云数据平台模型。对于复杂的企业数据平台，混合云模型在控制及效率两个方面都堪称最佳。企业可以在私有云基础架构上运行敏感、监管严格以及任务关键的应用程序和工作负载，或者具有合理稳定性和容量要求的工作负载。它们可以在公有云上运行不太敏感、更加动态甚至是临时的工作负载，例如新应用程序的开发和测试环境。

通过私有云与公有云之间的适当集成和协调，企业在需要时甚至可以同时利用两者运行同一工作负载。例如，通过“云爆发”，企业可以利用额外的公有云容量来应对私有云应用的需求高峰。多云固有的灵活性具有诸多优点，包括缓解风险、优化以及直接访问所需的服务。

多云可从两个方面帮助缓解风险：限制单一供应商方法的风险，以及防止供应商锁定。在多云环境中，如果一个供应商的云服务中断，只会影响一个供应商的服务。例如，如果您托管的电子邮件服务关闭了几个小时，但其他提供商（如网站或软件开发平台）的服务仍可运行。如果某供应商无法再满足您的需求，不会妨碍您选择另一个供应商的解决方案。

多云还可让用户选择最符合其需求的服务，以优化性能。一项服务可能提供额外的功能，或者采用更易于满足用户合规要求的安全协议。

多云的另一个重要优点是对技术的使用。例如，分析解决方案可以轻松部署为云服务，没有前期资本开支。这也意味着用户可以更快地运行服务，加快实现价值。

类似地，当用户能够自由选择任何解决方案的任何提供商时，相比于单一供应商的目录，他们可以更快地使用新的创新技术，并结合多个供应商的服务创建具有独特竞争优势的应用程序。

## Cloud Pak for Data 中 Red Hat 的价值

Cloud Pak for Data 基于的 Red Hat OpenShift 的优势，增强了混合云和多云选项。Red Hat 栈与 IaaS、OpenShift 及 Kubernetes 一起运行的优势尤为明显。它使您能够开发安全且可扩展的 Kubernetes 应用程序，而不会被大规模手动 Kubernetes 管理的复杂性所困扰。OpenShift 提供一个全面平台，该平台支持自动化操作，并提供对 Java、Node.js、Ruby 和 Python 等语言的开箱即用支持。OpenShift 还提供监控、身份验证和授权以及网络管理等支持服务。Red Hat 通过 OpenShift 增加的这些功能不在 Kubernetes 中。

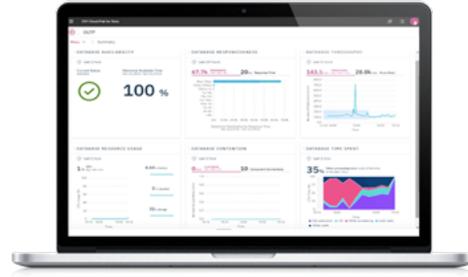
此外，随附的 Kubernetes 发行版为企业级，并且可从每个版本的数百个安全性、缺陷和性能修复程序中受益。同时提供经验证的适用于 Kubernetes 的流行存储和网络插件。最后，开源 Red Hat 工具提供了其他功能选项，例如用于流数据的 Apache Spark 或用于机器学习应用程序的流行 Python 和 R 语言。附加功能可确保企业利用通过 OpenShift 平台开发、部署和运行应用程序所需的必要开源工具。

当这些不同的资源全部成为 Cloud Pak for Data 中单个统一平台的一部分后，集成和管理就变得比以往更加容易。

## 02

### 数据集成和管理简便

集成数据管理、数据治理和分析的单一平台可以提高效率，改进资源的利用。这可以在不同数据源之间提供无与伦比的集成，从而改进数据访问，帮助维护数据安全性，并支持团队之间的自助服务协作。



### 统一的控制台

统一的数据平台需要统一的访问点，因此 Db2 控制台基于 Open Restful API 构建。这意味着所有仪表板功能都可通过开放的 API 访问。

这个单一门户可让您：

- 通过简便易用的交互式界面 (UI) 配置、执行、监控、管理和优化 100 多个任何规模的 Db2 数据库的性能。
- 在所有数据服务中为所有数据用户提供一致、高效、可扩展的体验。
- 通过可组合的用户界面进行协作 - Db2 控制台允许用户选择要显示的内容来组合自己的界面；这种选择可与队友共享。

### 预先集成的数据体验

通过 Cloud Pak for Data 上的微服务模型，可以使用针对特定行业预建的能力配置标准资源部署，例如欺诈检测或资料匹配。这些预建体验的优势包括：

- 通过基于角色的可定制工作流程提供公共的外观。
- 通过数据库中存在并且直接运行的模型提供 API 式的功能，改进性能并减少数据移动。
- 通过生态系统适应明确定义的 API，而不是依赖搭配和一套文件系统。
- 自动执行常见的服务，包括用户管理、身份验证模型、安全性配置、预配、协作等。

## IBM 生态系统

借助 Cloud Pak for Data 上的 Db2，可以轻松利用来自更广泛 IBM 和开源生态系统的各项功能。例如，在设计时集成全套 AI 功能的 IBM 解决方案，经认证可以协同作用，并且由 IBM 支持。它们包括：



### Watson Studio

AI 的构建阶段：使用开放、可扩展的数据科学工具来探索数据、准备数据和开发模型。



### Watson Machine Learning

AI 的运行阶段：部署、保留和管理模型。



### Watson OpenScale

AI 的融入阶段：持续评估结果；查找 KPI 和生产指标。



### Watson Knowledge Catalog

在主动元数据和策略管理的支持下提供智能编目。



### InfoSphere DataStage

利用混合云和多云环境的 ETL 功能，包括元数据管理。



### Cognos® Analytics

对整个分析生命周期（从发现到运营管理）使用单一商业情报平台。

开源生态系统的更多优点包括：

- 创作工具 - Jupyter 笔记本、R Studio、H2OAI
- 机器学习、优化 - Spark、Anaconda
- 深度学习运行时 - Torch、Caffe、Keras、TensorFlow
- 编码和视觉建模选项 - Graph、Python、R 等
- 模型生命周期管理 - 自动版本控制、发行更新和 SLA
- 简化的生命周期管理 - 通过 Helm 滚动更新

### 安全性和可靠性

安全性和可靠性是 **Cloud Pak for Data** 的基础。容器化的架构有助于隔离任何失败，例如，使控制台级别的失败不影响整个系统。集中数据治理策略并应用到企业中的所有数据，使用数据时无需移动数据或授予更广泛的访问权限，因此数据安全性更高。

Cloud Pak for Data 基于安全性而设计，具有许多安全特点，包括访问控制、身份验证和授权；数据保护；以及安全日志记录。Db2 支持安全工程开发实践：威胁建模、风险评估、静态和动态代码分析、渗透测试、容器扫描等。它构建为支持安全运营最佳实践，例如审核日志合并与分析、用户访问管理以及安全事件管理。

所有这些功能大大简化了数据治理和合规措施。例如，合规控制措施可由外部机构制定，Db2 包含用于维护合规安全态势的系统安全计划。Cloud Pak for Data 上的 Db2 包含符合《联邦信息安全管理法案》(Federal Information Security Management Act, FISMA) 的系统安全计划，涵盖 350 项控制措施。

由于具有部署模型灵活性、易于集成和统一管理等诸多优势，Cloud Pak for Data 上的 Db2 可以支持企业数据的最终目标：快速而轻松地获取任意位置的数据中的洞察力，无需先移动数据。

### 03 无需移动数据的跨多个数据源查询

利用 Db2 on Cloud Pak for Data，您可以获得所有数据的可信视图和全面的洞察力，无需移动、复制数据或使其暴露给安全风险。

这两套功能可帮助您获得数据的单一视图：

#### 数据治理和自助服务分析

Db2 中的数据自动集成治理功能，提供自动发现、编目和搜索。这些自动流程需遵守数据安全策略和访问规则，支持最广泛数据用户的自助服务发现。由于数据经过预先清理和预先验证，而且访问控制可以集中管理，因此用户可以快速使用数据，因为他们知道这些数量的质量很高，而且访问的都是针对性的数据。

#### 生态系统和虚拟化

Db2 系列包括针对 OLTP (Db2 Database)、OLAP (Db2 Warehouse) Hadoop/大数据 (BigSQL) 和串流“快数据”(Event Store) 优化的功能。Cloud Pak for Data 中还可以利用 Db2 与服务提供商（如 MongoDB、Postgres 等）的其他正式合作。

特别值得一提的是数据虚拟化，它虽然是一项普遍的技术，但要在不将所有数据移至单一存储库的情况下整理不同数据中的洞察力，数据虚拟化是不可或缺的，没有它不可能实现（由于不同类型的数据）或很难实现（由于成本、延时和安全风险）这一总体目标。虚拟化可帮助克服数据孤岛固有的障碍。

如果没有虚拟化，繁琐的数据资源搜索、移动和治理任务将要耗费大量的成本，羁绊技能熟练的数据团队。仅仅几年前，数据科学家有 80% 的时间用在这方面，<sup>4</sup> 当然，进展依然不尽如人意。<sup>5</sup> 数据传输也耗费时间，并且可能费用惊人。虽然 IBM Cloud® 不对数据传输收费，但其他很多提供商会收。

对比之下，经过治理的数据虚拟化（图 4）让您不需要知道或了解数据的实体格式或位置，也不需要移动或复制数据，便可以查看、访问、操作和分析数据。数据虚拟化将数据孤岛连成一片，就好像它们是单一数据集一样。虚拟化过程对数据所在的服务器执行分析查询，然后将结果返回到原始应用程序。

使用 Cloud Pak for Data 上的 Db2，您可以避免与移动数据相关的孤岛（数据子集）、治理、安全性和重复问题。您还可以解释数据科学家，让他们专注于战略性的计划和模型构建，而不是手动数据清理和操作。

#### 了解更多

[下载白皮书](#)或[观看视频](#)，以了解有关虚拟化和使用 Cloud Pak for Data 消除数据孤岛的详细信息。

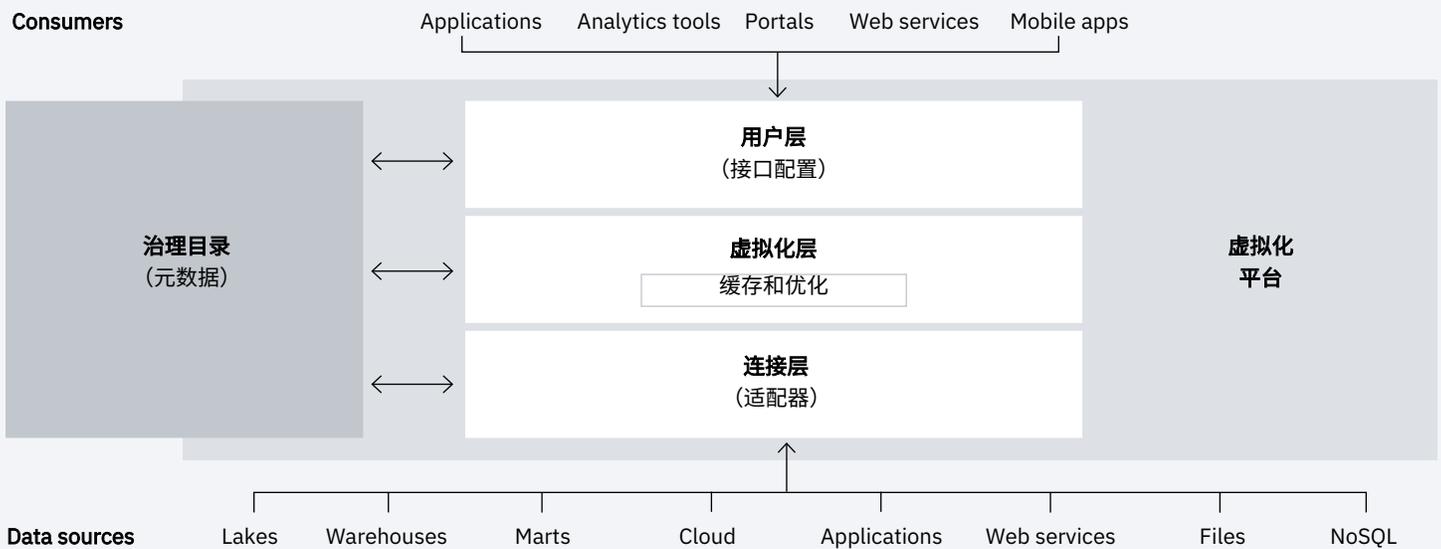


图 4. 虚拟化层如何连接任何数据与任何用户的示例

通过 IBM Cloud Pak for Data 部署时，Db2 不只是可用于收集数据的强大的数据存储库。它已成为基于微服务的数据平台的一部分，还可以帮助您组织和分析数据，并将 AI 功能融入您的企业。

## IBM Cloud Pak for Data 上的 Db2

1. 使用 AI 功能构建，专为 AI 项目打造
2. 数据虚拟化无需移动数据即可释放洞察力
3. 通过单一接触点访问预先治理的自助服务

Cloud Pak for Data 上的 Db2 提供容器化架构、Red Hat 基础结构，数据连接性和更广泛 IBM 功能生态系统的独特组合，对于希望为未来机遇准备数据基础的企业而言，是极具吸引力的选择。

要开始使用，请[免费试用 Cloud Pak for Data](#)。安排与混合数据管理专家的[免费一对一咨询](#)。

©IBM Corporation 版权所有，2020年

IBM Corporation  
New Orchard Road, Armonk, NY 10504  
美国印制，2020 年 6 月

IBM、IBM 徽标、ibm.com、Db2、BLU Acceleration、pureScale、IBM Cloud Pak、IBM Watson、InfoSphere、DataStage、OpenScale、Cognos 和 IBM Cloud 是国际商业机器公司的商标，已在全世界许多司法辖区注册。其他产品和服务名称可能是 IBM 或其他公司的商标。当前的 IBM 商标列表请见网站的“版权和商标信息”版块：[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)。http://www.ibm.com/legal/copytrade.shtml Red Hat® 和 OpenShift® 是 Red Hat, Inc. 或其下属公司在美国和其他国家/地区的商标或注册商标。本文档包含截至发布之日的最新信息，IBM 可能随时更改。并非所有产品或服务在 IBM 开展业务的所有国家/地区均有提供。用户应负责对 IBM 产品和程序的任何其他产品或程序运行进行评估和确认。本文所载信息按“原样”提供，不做任何明示或暗示的担保，包括对适销性、特定目的的适用性的任何担保，以及针对非侵权的任何担保或条件。IBM 根据产品交付协议中规定的条款和条件为产品提供担保。

客户应遵守适用的法律与法规。IBM 不提供法律建议或声明或保证其服务或产品能够确保客户遵循所有法律或法规。良好安全实践声明：IT 系统安全性涉及通过预防、检测和应对来自企业内外的不当访问以保护系统和信息。不当访问可能导致信息被篡改、销毁、盗用或不当使用，也可能导致系统受损或被不当使用，包括被用于攻击他人。不应认为任何 IT 系统或产品是绝对安全的，任何一种产品、服务或安全措施都不能完全有效地防止不当使用或访问。IBM 系统、产品和服务被设计为合法的综合安全性方法的一部分，必然涉及其他操作过程，可能需要其他系统、产品或服务配合才能发挥最大效用。IBM 不保证任何系统、产品或服务不受任何一方的恶意或非法行为影响，也不保证您的企业不受任何一方的恶意或非法行为影响。

- 1 “Reshaping Business with Artificial Intelligence” by Sam Ransbotham, David Kiron, Philipp Gerbert, and Martin Reeves. MIT Sloan Management Review, September 2017. [www.sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/](http://www.sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/)
- 2 “The Forrester Wave™: Data Management For Analytics, Q1 2020” by Noel Yuhanna. Forrester, February 2020
- 3 [www.redhat.com/en/blog/red-hat-global-customer-tech-outlook-2019-automation-cloud-security-lead-funding-priorities](http://www.redhat.com/en/blog/red-hat-global-customer-tech-outlook-2019-automation-cloud-security-lead-funding-priorities)
- 4 “Cleaning Big Data” by Gil Press, Forbes, March 2016. <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says>
- 5 “What do Data Scientists Really Do?” by Hugo Bowne-Anderson, Harvard Business Review, August 2018 <https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists>