

IBM watsonx.data: 一个开放、混合且受 治理的湖仓一体方案

随时随地为所有数据扩展 AI 工作负载

■

要点

选择适用的查询引擎

应用内置数据治理、
安全性和自动化

共享单个数据副本

数分钟内即可连接至数据

扩展 AI 需要可信数据,然而大多数组织都难以应对基本的数据挑战。根据 IDC 的数据,到 2025 年,存储数据量将增长 250%¹,数据会在本地、云端、应用和多个位置之间快速传输,且可能存在质量问题。这种情况会导致数据孤岛增多,成本增加,复杂程度加深,组织更难以管理 AI 和数据工作负载。

如何将数据仓库的高性能和可用性与数据湖的灵活性和可扩展性相结合,从而应对当今复杂数据环境蕴含的挑战并扩展 AI? 通过优化您的 AI 和分析工作负载,并以适当的成本为相应的工作负载选择正确的引擎来实现目标——无论您的数据驻留在何处。

全新 IBM® watsonx.data™ 平台之所以能做到这一点,是因为它是一个开放、混合且受治理的数据湖仓,且已完成优化,适用于所有类型的数据和 AI 工作负载。该产品可帮助组织从分析生态系统中获得最大价值,以下三大核心优势提供了支持。



跨混合云访问所有数据

通过跨所有云和本地环境的共享元数据层的单一入口点访问所有数据。



数分钟内即可开始使用

数分钟内即可连接到存储和分析环境,并通过内置的治理、安全性和自动化功能增强对数据的信任。



将数据仓库成本降低多达 50%²

使用适合的方案(如 Presto、Spark、IBM Db2®、Netezza® 等),优化数据仓库并实现数据湖现代化。

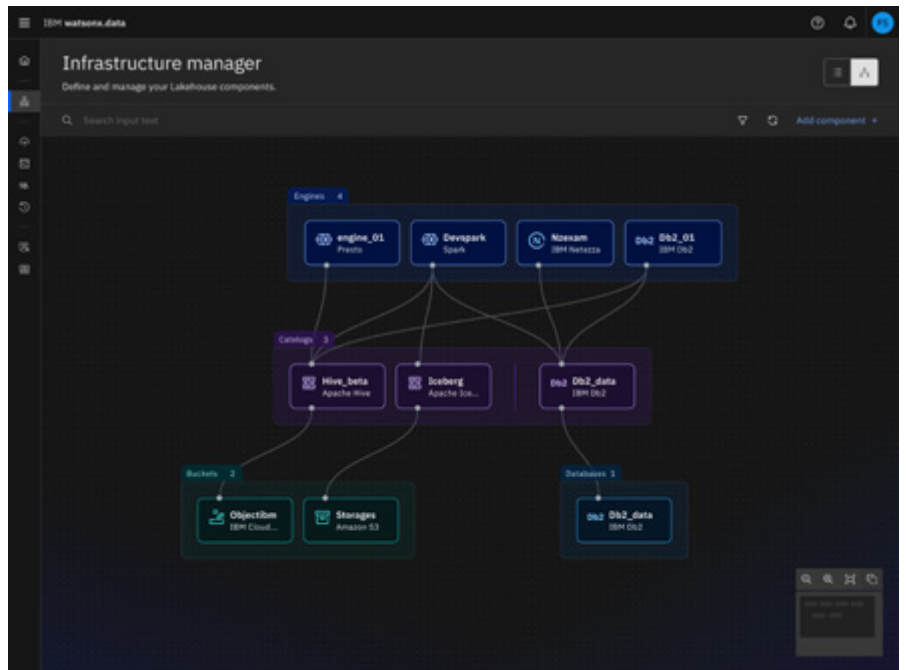


图 1. 确保采用的查询引擎符合所需目的, 进而通过低成本的计算和存储以及适用的分析引擎 (包括 Presto 和 Spark) 寻求助力, 动态调整规模的大小, 帮助降低分析成本。

选择适用的查询引擎

没有哪一个分析引擎可以做到满足所有分析需求。要满足如此广泛的分析需求, 就需要多个分析引擎。

您可以使用经济高效的对象存储和适用的查询引擎来优化工作负载, 从而优化成本高昂的仓库工作负载, 并帮助将数据仓库成本降低多达 50%。² 这些包括针对 BI 工作负载而优化的 Presto; 以及针对机器学习和数据科学 (ML/DS) 工作负载而优化的 Spark; 它们都会随着需求的变化自动调整规模的大小。只需单击几下, 您即可快速添加所优选的新查询引擎, 且符合企业的性价比要求。

应用内置数据治理、安全性和自动化

借助 watsonx.data 内置的治理、访问控制和企业安全性, 帮助保护数据、管理合规性, 并针对各类应用于 AI 的数据, 维护所需的信任度。与 IBM 的集中式治理功能相集成, 自动实施策略, 在整个企业内构建负责、透明且可解释的数据和 AI 工作流程。最后, 您可以通过 watsonx.ai 模型和对话式用户体验 所具备的强大功能来发现、增强、优化和可视化 watsonx.data 数据和元数据。

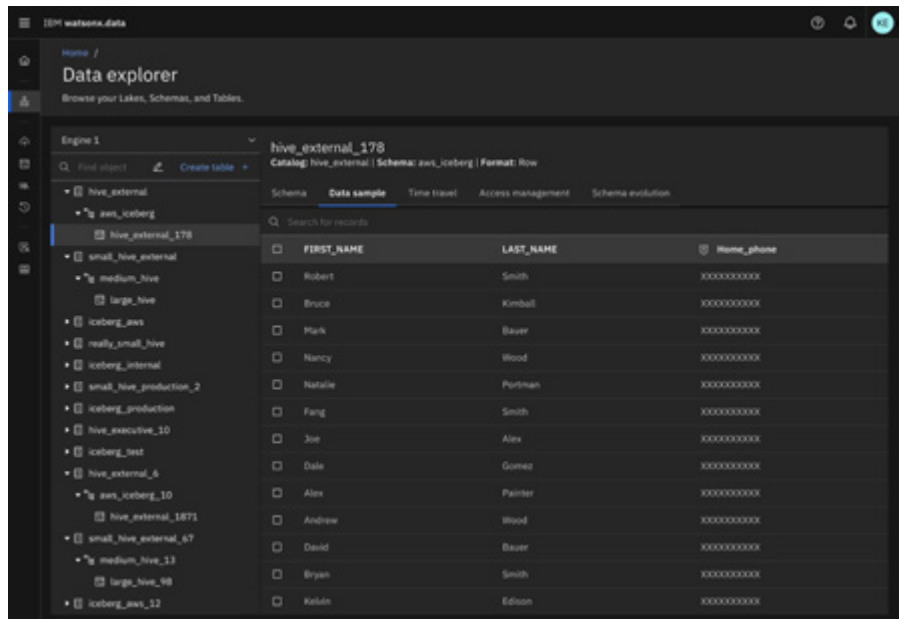


图 2. 内置的治理和安全性通过跨数据生态系统完成统一治理, 帮助解决企业合规性和安全性问题。

共享单个数据副本

开放促进协作。它还可以减少支持不同用户和工具所需的数据副本数量, 提高数据完整性并帮助解决安全风险。减少副本意味着更少的软件、更低的硬件要求和更低的存储成本。

借助 watsonx.data, 您可以跨数据库和数据湖访问所有数据。通过开放表格式 (例如 Apache Iceberg) 共享大量数据, 这些格式专为高性能分析和大规模数据处理而构建。支持针对分析数据集的多种供应商开放格式, 允许不同的引擎访问和共享相同的数据, 且同时还可使用 Parquet、Avro 和 Apache Orc 等工具。依靠 watsonx.data, 在采用单个数据副本的多个查询引擎之间共享元数据, 以满足所有分析及 AI 工作负载的运行需求。

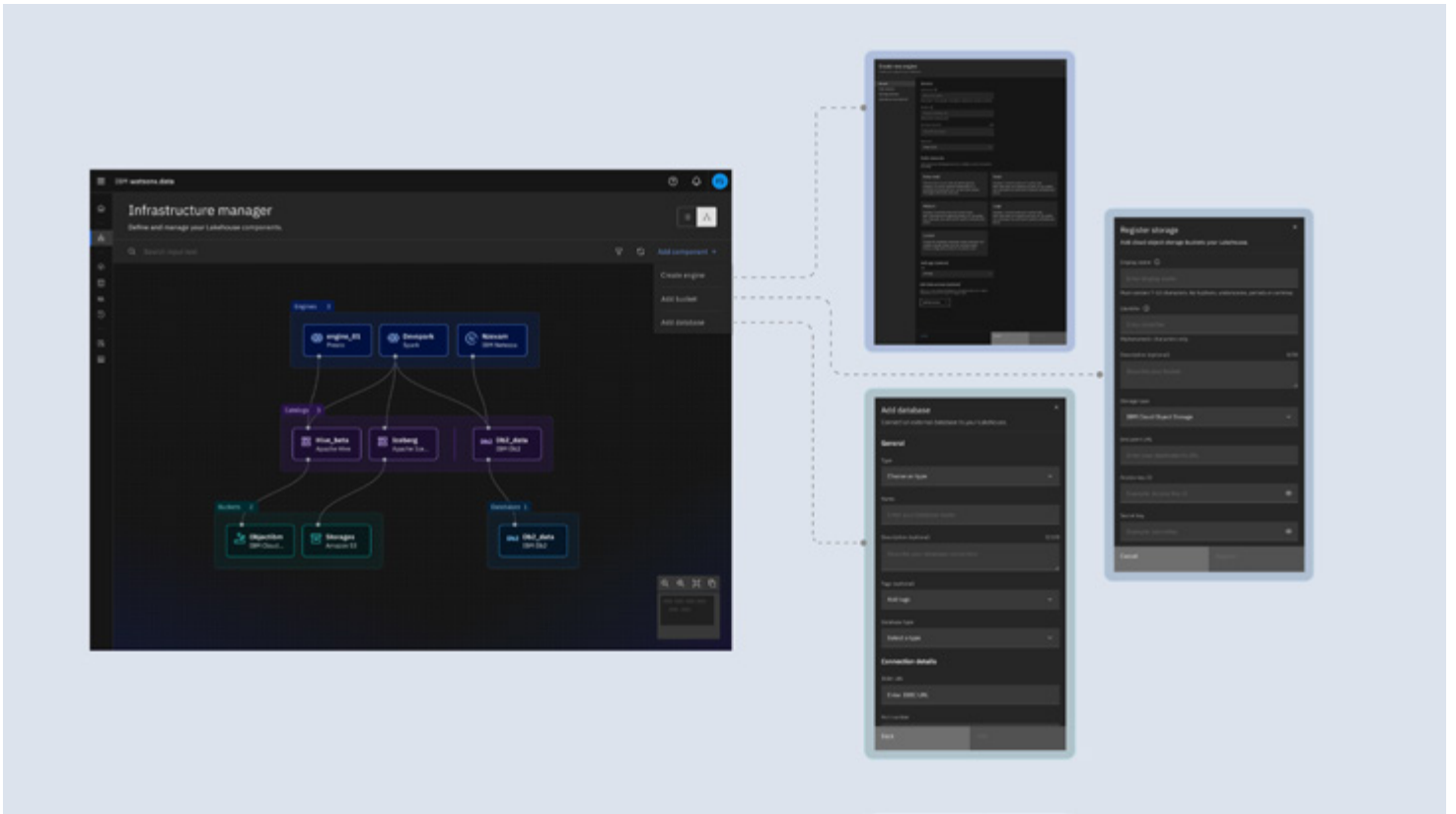


图 3. 易于使用的集成控制台可确保您数分钟内即可连接至现有分析数据并部署查询引擎。

数分钟内即可连接至数据

数分钟内即可将现有数据与新数据连接起来，解锁新的可信洞察成果，而无需承担数据治理、复制和迁移成本以及各种相关联的复杂事宜。用户可以使用通用 SQL 来浏览和转换数据。Watsonx.data 还支持与强大的 IBM 生态系统和第三方技术相集成，助力简化分析工作负载的开发和部署过程。

无论您决定如何部署 watsonx.data，都可以在数分钟内投入使用。它可以通过 IBM Cloud® 和 AWS 上的 SaaS 或作为容器化软件轻松访问。通过简单的 UX 和控制台加快团队采集、访问和转换数据以及运行工作负载的速度。



图 4: IBM watsonx.data 是唯一面向所有数据和 AI 工作负载而优化的数据湖仓。

结论

Watsonx.data 支持您跨云和本地环境访问所有数据。数分钟内即可建立数据连接并开始使用，且内置治理、安全性和自动化功能。充分利用多个查询引擎运行分析和 AI 工作负载，从而将数据仓库成本降低多达 50%。²

watsonx.data 属于一种面向所有数据和 AI 工作负载而优化的开放、混合且受治理的湖仓一体方案，可帮助您从分析生态系统中获取更大价值，同时还可以与 AI 有效结合使用。

为什么选择 IBM?

IBM 深受各方信赖，可帮助客户管理各种对任务最为关键的数据和应用。我们在企业数据解决方案领域拥有丰富的创新经验，包括市场开创型的数据库解决方案和企业就绪型人工智能方案等。我们几乎可帮助客户在所有云或本地环境中运行解决方案，并秉持“客户的数据完全属于客户”的理念。

更多信息

如需了解有关 watsonx.data 的更多信息，请联系您的 IBM 代表或 IBM 业务合作伙伴，或者访问 ibm.com/cn-zh/products/watsonx-data。

详情请拨打热线：400 810 1818 转6216

1. IDC 全球 DataSphere 预测 2022-2026 年 <https://www.idc.com/getdoc.jsp?containerId=US49018922>。
2. 针对 watsonx.data 所对应的 VPC 小时数而发布的 2023 年标准化定价单, 将其与多家主要云数据仓库供应商产品价格相比较。具体节省数量可能因配置和工作负载而异。

© Copyright IBM Corporation 2023

国际商业机器 (中国) 有限公司
了解更多信息, 欢迎访问我们的中文官网:
<https://www.ibm.com/cn-zh>

美国出品 2023 年 5 月

IBM、IBM 徽标、IBM Cloud、Db2、Netezza 和 watsonx.data 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。关于 IBM 商标的最新列表可参见以下 Web 网址: ibm.com/cn-zh/trademark。

本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。

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。

本档内的信息“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, 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。

