

# IBM watsonx.data: 受治理的开放式混合 数据存储

随时随地扩展 AI 工作负载, 让所有数据  
充分发挥作用

## ■ 要点

选择专用的查询引擎

应用内置的数据治理、  
安全性和自动化

共享单一数据副本

数分钟内即可连接至数据

要扩大 AI 的应用规模, 离不开可信的数据, 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, 存储的数据量将增长 250%<sup>1</sup>, 数据将在多个本地环境、云端、应用和地点之间快速传播, 而且可能存在质量问题。这种情况会产生更多数据孤岛, 导致组织治理 AI 和数据工作负载的成本和复杂度显著提升。

如何将数据仓库的高性能和易用性与数据湖的灵活性和可扩展性有机结合, 应对当今复杂的数据环境所带来的挑战并轻松扩展 AI? 要做到这一点, 必须优化 AI 和分析工作负载, 并以适当的成本为适当的工作负载选择适当的引擎, 而无论数据位于何处。

全新的 IBM® watsonx.data™ 平台能够帮助您实现这一目标, 因为它是受治理的开放式混合数据湖仓, 针对所有类型的数据和 AI 工作负载进行了优化。该产品借助以下三大核心优势, 帮助组织从分析生态系统中获得最大价值。



### 访问混合云中的所有数据

通过单一入口点, 借助覆盖所有云和本地环境的共享元数据层访问所有数据。



### 只需几分钟即可轻松入门

数分钟内即可连接到存储和分析环境, 借助内置的治理、安全性和自动化功能加深对数据的信任。



### 最高可将数据仓库的成本降低 50%<sup>2</sup>

使用适合作业的工具 (如 Presto、Spark、IBM Db2®、Netezza® 等), 优化数据仓库, 并使数据湖实现现代化。

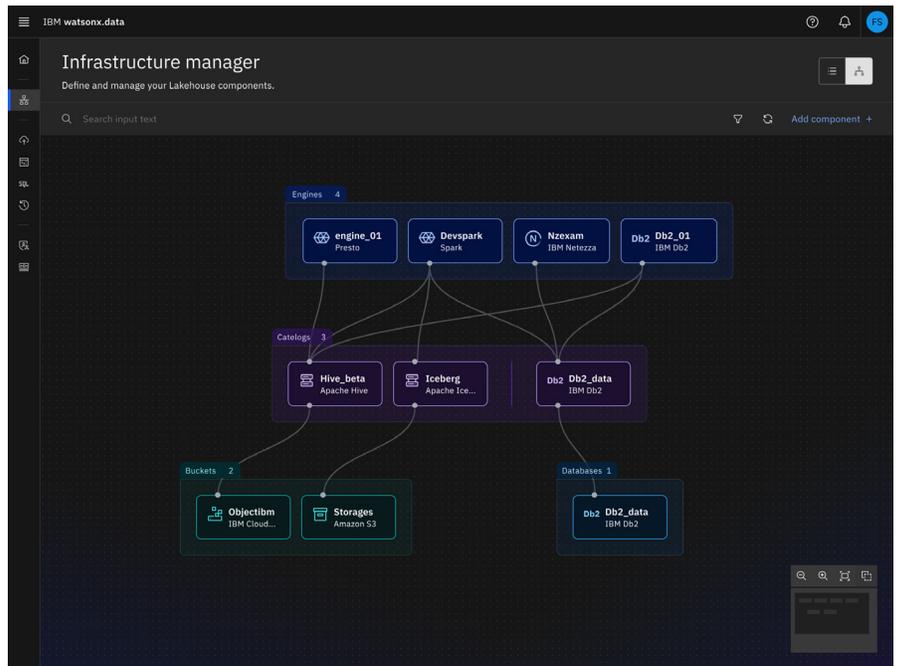


图 1. 专用的查询引擎通过低成本的计算和存储以及专用的分析引擎 (包括 Presto 和 Spark) 动态调整规模, 帮助降低分析成本。

### 选择专用的查询引擎

不存在一体适用的分析引擎。要满足如此广泛的分析需求, 需要采用多个分析引擎。

您可以使用经济实惠的对象存储和专用的查询引擎, 优化成本高昂的数据仓库工作负载, 最高可使数据仓库的成本降低 50%。<sup>2</sup> 这包括针对 BI 工作负载进行优化的 Presto, 以及针对机器学习和数据科学 (ML/DS) 工作负载进行优化的 Spark; 它们都可根据需求的变化自动调整规模。只需几下点击操作, 即可快速添加所选的新的查询引擎, 满足企业的性价比需求。

### 应用内置的数据治理、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, 帮助保护数据安全, 管理合规性, 并保持对 AI 所用数据的信任度。与 IBM 的集中式治理功能集成, 自动实施策略, 在整个企业范围内支持负责任、透明而且可解释的数据和 AI 工作流程。最后, 可通过 watsonx.ai 模型和对话式用户体验的强大功能, 发现、扩充、优化和直观呈现 watsonx.data 数据和元数据。

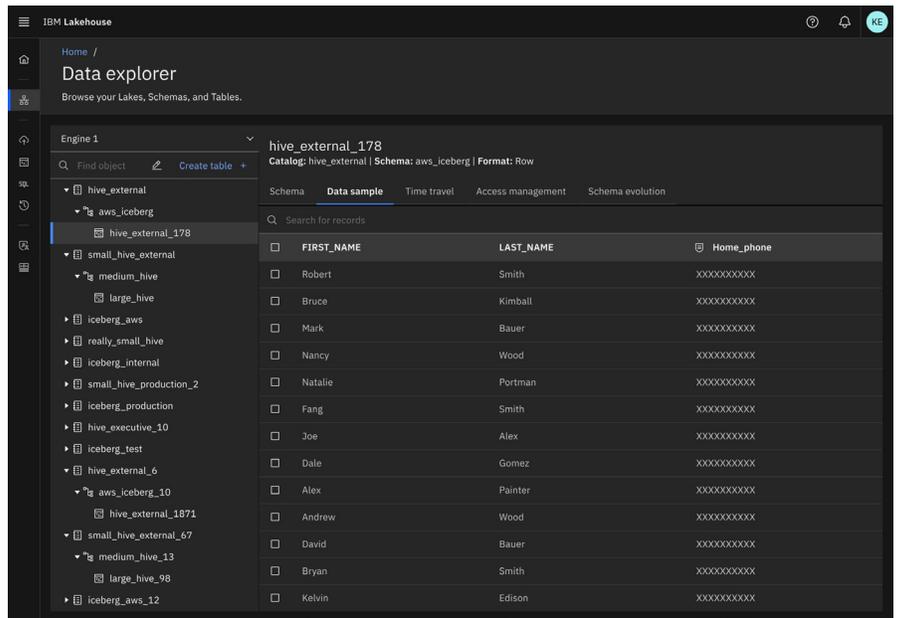


图 2. 内置的治理和安全功能在整个数据生态系统中实施统一治理，帮助解决企业的合规与安全问题。

### 共享单一数据副本

开放促进协作。这还有助于减少支持不同用户和工具所需的数据副本数量，从而提高数据完整性，降低安全风险。减少副本意味着所需的软件更少，硬件需求和存储成本更低。

借助 watsonx.data，可以轻松访问多个数据库和数据湖中的所有数据。通过开放表格式（例如 Apache Iceberg），可以轻松共享大量数据，从而有力推动高性能分析和大规模数据处理。为分析数据集支持多种供应商开放格式，允许使用 Parquet、Avro 和 Apache Orc 等工具的不同引擎同时访问和共享相同的数据。依靠 watsonx.data，针对所有分析和 AI 工作负载使用单一数据副本，在多个查询引擎之间共享元数据。

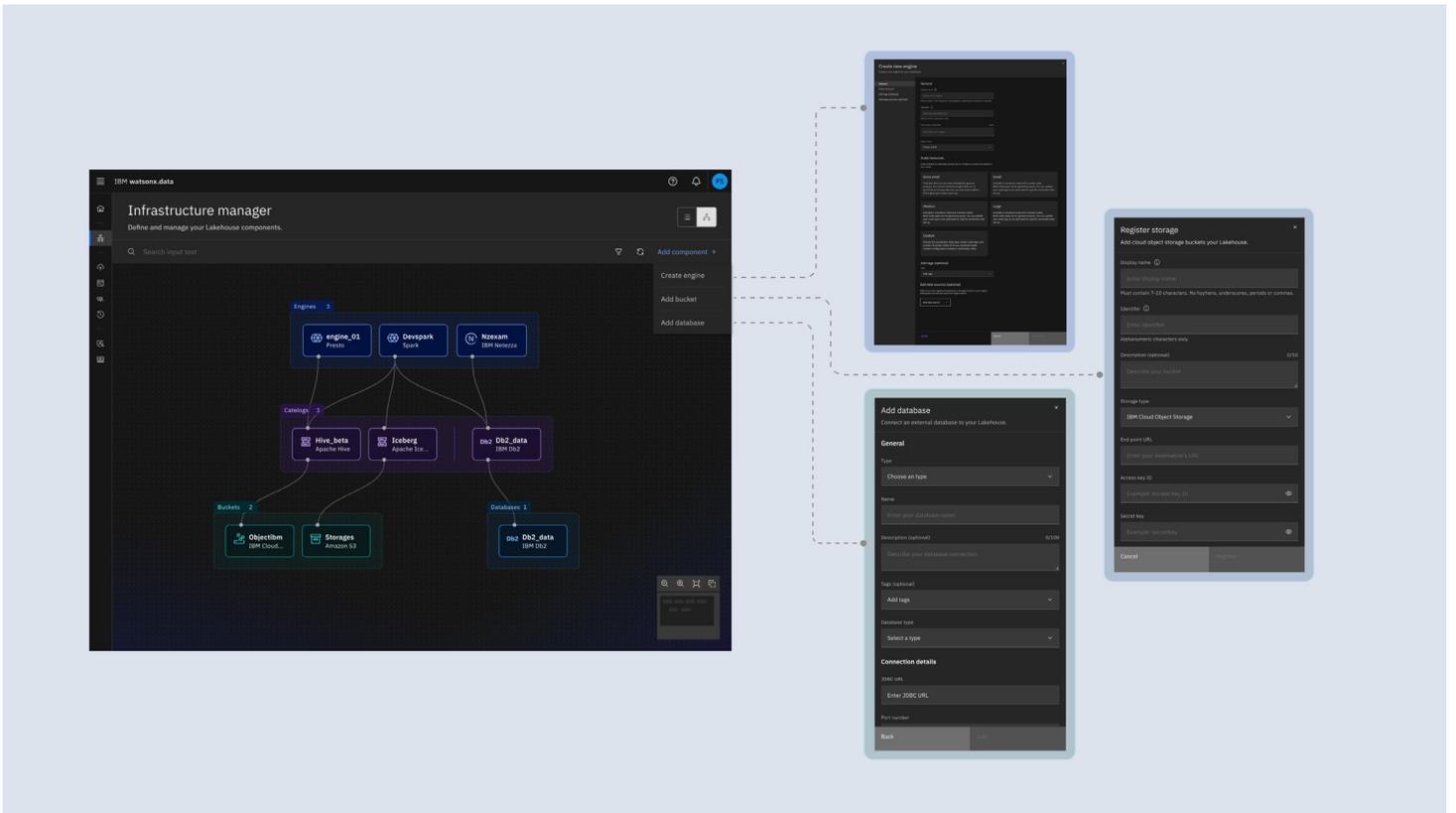


图 3. 借助易于使用的集成控制台，只需数分钟即可连接至现有分析数据并部署查询引擎。

**数分钟内即可连接至数据**

只需数分钟，即可将现有数据与新数据连接起来，挖掘全新而且可信的洞察，而不会带来管理、复制和移动数据方面的成本和复杂性。用户可使用通用 SQL 来探索和转换数据。watsonx.data 还支持与强大的 IBM 和第三方技术生态系统集成，帮助简化分析工作负载的开发和部署。

无论您决定如何部署 watsonx.data，都可以在几分钟内开始使用。它可以通过 IBM Cloud® 和 AWS 上的 SaaS 访问，也可以作为容器化软件使用，非常方便。通过简单的用户体验和控制台，团队可以更快地采集、访问和转换数据以及运行工作负载。

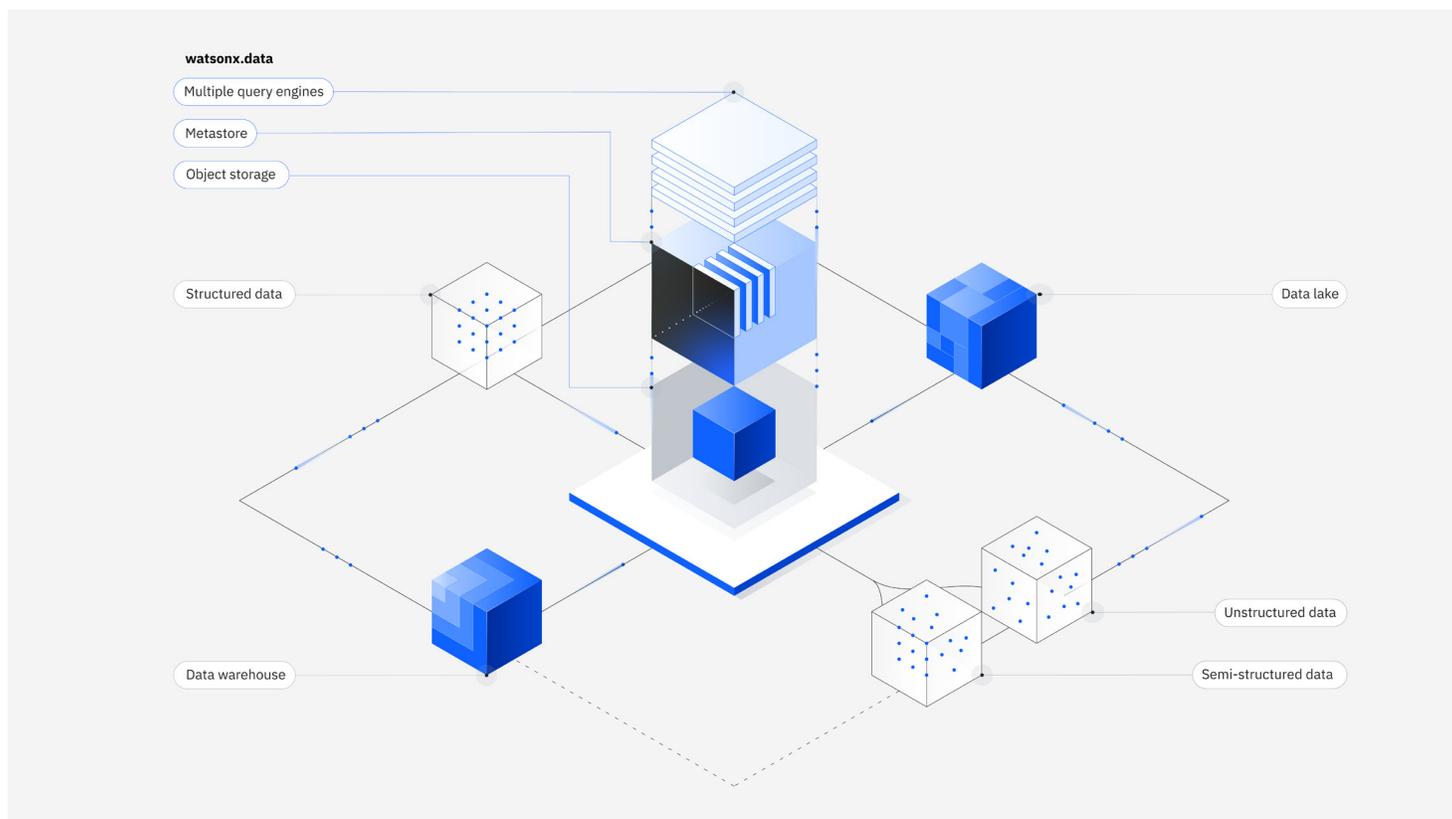


图 4. IBM watsonx.data 是唯一针对所有数据和 AI 工作负载进行优化的数据湖仓。

### 总结

Watsonx.data 支持您访问各种云和本地环境中的所有数据。它内置了治理、安全性和自动化功能，只需数分钟即可建立数据连接并开始使用。可利用多个查询引擎运行分析和 AI 工作负载，最高可将数据仓库的成本降低 50%。<sup>2</sup> watsonx.data 是一种受治理的开放式混合数据存储；针对所有数据和 AI 工作负载进行了优化；旨在帮助您从分析生态系统中获得更大价值，让 AI 发挥更大作用。

### 为何选择 IBM?

IBM 深受客户信赖，能够帮助他们管理最重要的任务关键型数据和应用。我们在企业数据解决方案领域拥有丰富的创新经验，包括市场创新型数据库解决方案以及面向企业的 AI 等。我们可帮助客户在几乎所有云或本地环境中运行解决方案，并始终秉持“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](https://ibm.com/cn-zh/products/watsonx-data)。

免费试用 watsonx.data →

申请实况演示 →

1. IDC Global DataSphere Forecast 2022-2026 <https://www.idc.com/getdoc.jsp?containerId=US49018922>
2. 将 2023 年公布的 watsonx.data 的 VPC 小时数标准化标价与几家主要云数据仓库供应商进行比较。具体节省金额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 6 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。Web 地址 [ibm.com/trademark](https://www.ibm.com/trademark) 上提供了 IBM 商标的最新列表。

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。

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。

本文档内的信息“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。

