

借助 IBM Cloud Pak for Data 和 IBM Spectrum Discover 让 数据变得智能

从 Spectrum Discover 到 IBM Watson Knowledge Catalog 的一键式集成

亮点

- 充分利用 IBM Watson Knowledge Catalog 中的文件和对象数据
 - 可通过一键式集成功能充分利用数百万个或数十亿个文件和对象
 - 充分利用 IBM Cloud Pak for Data 中的文件和对象数据
 - 使用 Watson Studio 分析文件和对象数据
 - 充分利用面向数据和 AI 的集成存储解决方案组合
-

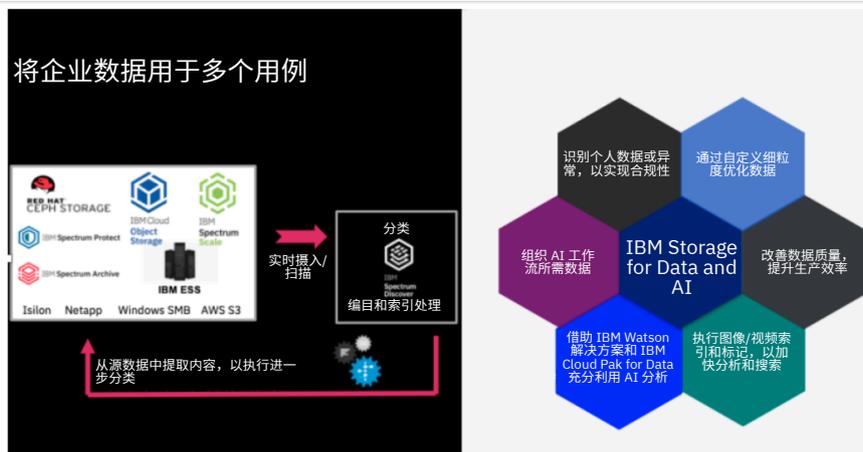
高质量的海量数据能够使人工智能 (AI) 驱动型应用变得更智能。据相关报告显示, 有 81% 的商业领导者不了解 AI 所需的数据和基础架构。

现在, IBM Storage 已将文件和对象数据集成到 IBM 的 AI 解决方案之中, 旨在让数据变得更智能。

希望充分利用 AI 价值的企业已经认识到, 馈送到 AI 应用中的数据准确性、一致性及其他质量特性, 会直接影响 AI 所产生结果的质量。更高质量的数据输入, 意味着更高质量的结果输出。

但这并不是当今组织所面临的唯一一个数据相关挑战。许多公司都因业务产生的数据太多而备受困扰, 导致他们无法跟踪所有数据, 或不了解自己的数据, 也就无法充分利用这些非常宝贵的资产来获利。特定的数据资产位于何处? 数千或数百万个文件中的每个文件具体包含哪些数据? 数据的“年龄”如何? 谁可以访问这些数据? 他们应如何访问? 是否符合最新的政府和国际法规?

IBM Spectrum Discover



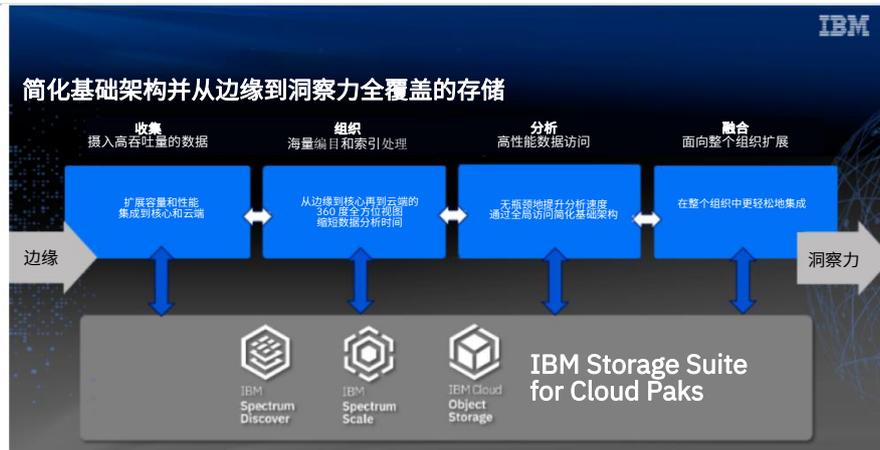
将 IBM Spectrum Discover 用于多个用例

IBM Spectrum Discover 可通过大容量数据编目和数据管理解决方案简化数据和 AI 组织，它能够提供 360 度全方位数据视图，进而提高生产效率，而且有助于优化数据，以实现更快的 AI 分析。

用于非结构化数据的传统元数据解决方案无法就存储使用情况、数据质量等方面提供如今所需级别的详情，也无法与其他 AI 解决方案很好地集成。IBM Spectrum Discover 在数据分类和标记方面起着重要的作用，它使用自定义元数据，不仅能让您更轻松地查找和调用数据用于分析，还通过赋予数据额外的语义和意义，真正增加数据的价值。

IBM Spectrum Discover 是一个内部交互式数据管理目录，它能够提供详细的实时数据视图。它不会破坏现有存储和应用，而且能够让用户创建自定义索引和报告，以识别和优化企业数据。

让面向 AI 的存储变得简单



面向数据和 AI 的集成式存储解决方案组合

作为 IBM 产品组合的一部分，IBM Spectrum Discover 可帮助企业简化 AI 之旅；具体来说，它提供了一个效率得到优化的信息基础架构，可帮助企业更快地获得结果，而此类结果可进行大规模扩展，而且从边缘到洞察力，可供全球使用。

借助 IBM Storage for data and AI，客户可以：

- 通过综合性的数据和 AI 产品组合简化 AI 之旅
 - 在摄入数据的同时，能够随着增长而进行扩展，而且可提供现代数据访问
 - 能够对来自多个来源的数据进行编目和索引处理，而且可提供实时数据视图
 - 通过对 GPU/CPU 的快速访问以及不断演变的数据科学工具构建更高性能的分析
 - 在整个组织中使用简单且全面的基础架构来扩展价值和效率
- 从边缘到核心再到云端，简化现代解决方案的 AI 基础架构
 - 可确保基于文件或 POSIX 的应用工作负载，或数百种经验证的基于 S3 的对象工作负载的灵活性

提供适于 Splunk SmartStore、SAS 网格、面向 AI 的 NVIDA、面向 AI 的 IBM Power Systems 的参考架构和蓝图。

通过与 IBM Cloud Pak for Data 相集成来简化多云和容器连接

- 通过从 Spectrum Discover 到 IBM Watson Knowledge Catalog 的一键式集成实现企业数据集成
- 使用 IBM Storage Suite for IBM Cloud Paks 的单个产品解决方案，可以灵活地根据文件、块或对象的需求选择存储选项（包括 Spectrum Discover）

让数据变得更智能



IBM Cloud Pak for Data 及 IBM Spectrum Discover

现在，数据科学家和数据用户只需简单点击一下，便可将来自 Spectrum Discover 的任何信息直接集成到 IBM Watson Knowledge Catalog 之中。这意味着 IBM Watson Knowledge Catalog 以及与该目录连接的所有工具均可充分利用元数据形式的信息，以及来自 Spectrum Discover 的定制化标签和自动标签。

基于 IBM Cloud Pak for Data 的 IBM Watson Knowledge Catalog 是一个与企业数据治理平台紧密集成的数据目录。IBM Watson Knowledge Catalog 可帮助数据公民轻松查找、准备、理解和使用他们所需的数据。该解决方案提供了基于元数据和主动策略管理的端到端体验，可帮助企业成功实现各种高级用例，比如面向《加利福尼亚消费者隐私法案》(CCPA) 和《通用数据保护条例》(GDPR) 的监管合规用例、数据湖治理用例，以及高质量数据自助使用用例。

IBM Watson Knowledge Catalog 提供了以下独特功能，可帮助企业作为 DataOps 实践的一部分加速实现业务价值并实施 AI 项目：

- 实时数据虚拟化支持
- 自动数据发现和元数据生成
- 借助 ML 从最常见的监管术语中提取业务术语表
- 通过动态数据屏蔽保护敏感数据
- 通过 Watson Knowledge Catalog InstaScan 实现非结构化数据的自动扫描和风险评估

IBM 致力于通过基于 IBM Cloud Pak for Data 的 IBM Watson Knowledge Catalog 帮助客户交付业务就绪数据，进而为 AI 和分析项目提供高质量数据。Spectrum Discover 提供的企业文件和对象数据信息集成功能，有助于为 AI 和分析项目交付更多数据和更多价值。

高效使用项目中的资产

将信息从 IBM Spectrum Discover 移至 IBM Watson Knowledge Catalog 并将数据移至数据中心之后，数据用户便可高效使用数据或分析资产来发现洞察力。数据中心是一个支持 S3 的存储系统，而且无论部署在何处，IBM Cloud Object Storage 均可被用作存储系统。您可以将数据与其他元数据一起从 Spectrum Discover 移至项目中。您还可以将项目用作登台区域，用以管理数据资产或构建分析资产，然后再将它们发回到 IBM Watson Knowledge Catalog。项目通常包含有目录合作者的所选子集。

您可以通过 Watson Studio 应用将目录中的资产添加到项目中，以供使用，而且可以添加其他工具来分析数据或构建人工智能模型。Watson Studio 和 Watson Knowledge Catalog 已完全集成，而且已融合到了 IBM Cloud Pak for Data 之中。

利用 IBM Storage for Data and AI 的集成示例

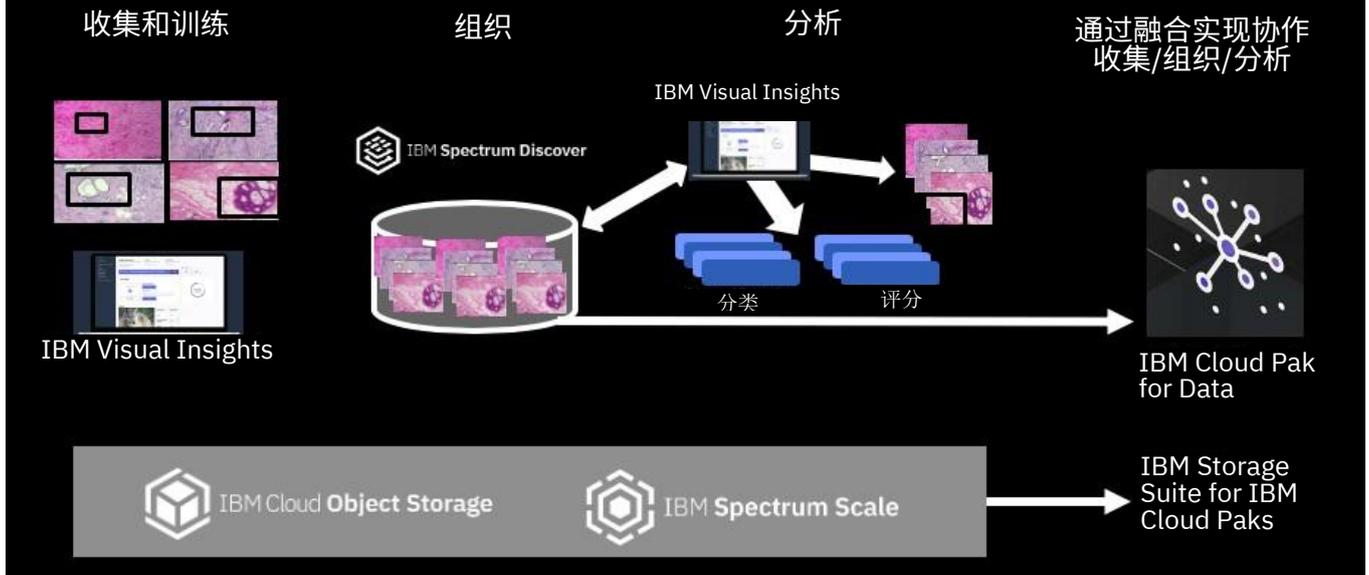
在该示例中，我们首先从癌症图像，甚至是 COVID-19 等病毒图像中收集并训练多个数据点。

这些数据是在 IBM Spectrum Scale 或 IBM Cloud Object Storage 中收集的。使用 IBM Visual Insights 开发一个推理模型，该模型能够基于一组现有的数据图像来识别缺陷、癌症或病毒。然后将数据发送到 Spectrum Discover，以执行 AI 流程中的组织阶段。Spectrum Discover 针对新数据调用“DeepInspect”策略（Spectrum Discover 的一个内置功能），以激活推理。现在，我们进入第一个分析阶段，在该阶段，我们将新图像与模型进行比较，以确定与我们的基本数据集匹配的可能性。触发 IBM Visual Insights，以分析新图像，并将它们重新编入到 Spectrum Discover 的索引。然后，Spectrum Discover 会显示结果，并将结果注册到 IBM Watson Knowledge Catalog，以执行数据治理、数据协作和进一步分析。来自 IBM Storage for Data and AI 的信息和数据也可以用于执行基于其他相关性的额外 AI 分析，例如地理数据、天气数据、物理数据或其他数据等，这些数据可以通过 IBM Cloud Pak for Data 中的工具（IBM Watson Studio 或 IBM Watson for AI and ML）进行其他分析。这不仅能够让数据变得更智能、更强大，也展现了集成式 AI 解决方案的优势。

IBM 面向数据和 AI 的解决方案可在整个企业范围内提供一种集成方法，使企业能够：

- 对整个企业中的非结构化数据进行基于事件的索引处理和标记
- 对 1000 多种非结构化数据进行自动检查和分类，包括基因组学和成像特定的文件格式
- 基于 Spectrum Discover 的搜索和过滤条件，通过 Watson Knowledge Catalog 自动注册资产，同时高效利用 Cloud Pak for Data 中的资产
- 基于来自 Spectrum Discover 的洞察力在 Cloud Pak for Data 中的 Watson Knowledge Catalog 中实施数据治理策略，同时高效利用 Cloud Pak for Data 中的资产

IBM Storage for Data and AI 已经集成到了 IBM Cloud Pak for Data 和 IBM Visual Insights



集成式 AI 方法示意图

为什么选择 IBM?

作为业内领先的数据存储产品提供商，IBM 对数据管理解决方案进行了大量投资，旨在改善存储经济效益、数据质量、数据治理水平及数据识别效率，进而为大规模分析和 AI 提供支持。IBM Spectrum Discover 是 IBM 在面向数据和 AI 的存储解决方案方面所具整体优势中的关键一环，可为客户提供强大的元数据管理功能，帮助他们实现数据可视性与分类，进而改善存储优化水平并提升数据科学技能。

有关更多信息

如欲了解有关以下所列任何一款产品的更多信息，请联系您的 IBM 代表或 IBM 业务合作伙伴，或访问以下 Spectrum Discover 网站：

<https://www.ibm.com/cn-zh/marketplace/spectrum-discover>

IBM Watson Knowledge Catalog:

<https://www.ibm.com/cn-zh/cloud/watson-knowledge-catalog>

IBM Cloud Pak for Data:

<https://www.ibm.com/cn-zh/products/cloud-pak-for-data>

IBM Storage Suite for IBM Cloud Paks:

<https://www.ibm.com/cn-zh/products/storage-suite-for-ibm-cloud-paks>

IBM Visual Insights:

<https://www.ibm.com/cn-zh/products/ibm-visual-insights>

© Copyright IBM Corporation 2020.

IBM、IBM 徽标及 [ibm.com](https://www.ibm.com) 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。Web 站点 <https://www.ibm.com/legal/us/en/copytrade.shtml> 包含了 IBM 商标的最新列表；Web 站点 https://www.ibm.com/legal/us/en/copytrade.shtml#section_4 包含了可能在本文中提及的所选第三方商标列表。

本文中包含了与以下 IBM 产品（IBM Corporation 的商标和/或注册商标）相关的信息：

IBM Cloud Pak[®]、IBM Spectrum[®] Discover、Watson[™]



有关 IBM 未来发展方向及意图的声明如有变更或撤销，恕不另行通知，且仅用于说明目标之用。