

# 面向对象存储的 AI 加速和数据编排

## 使用 IBM Spectrum Scale 和 IBM Cloud Object Storage 实现高性能对象存储访问

---

### 亮点

- 可通过 IBM Spectrum Scale 访问 IBM Cloud Object Storage
  - IBM Cloud Object Storage 并发访问和文件访问
  - 自动回收未使用的数据，为所需数据腾出空间
  - 消除不必要的副本及其相关管理
  - 提供缓存功能，让企业可以预取所需数据
  - 可选择将 IBM Spectrum Discover 用于策略引擎
- 

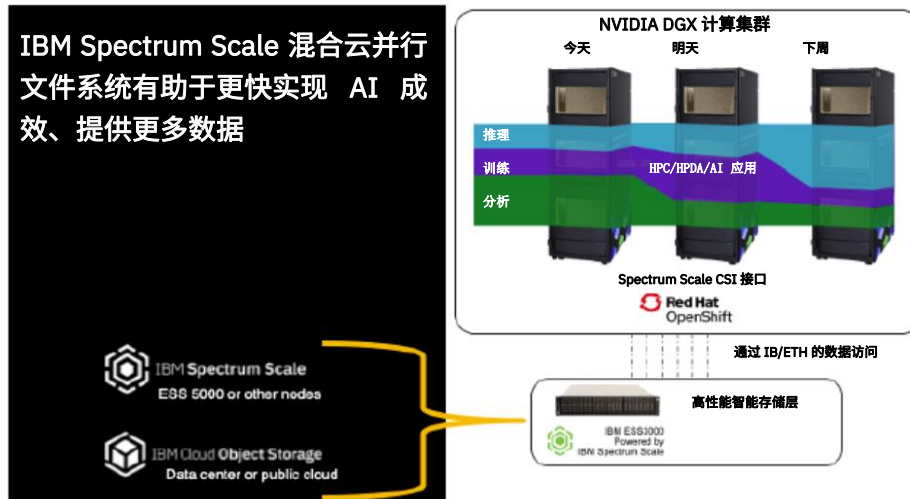
随着大多数组织开始采用 AI，支持 AI 模型所需海量数据集的存储基础架构变得至关重要。IBM Storage for Data and AI 最近推出了一项新功能，旨在实现 AI 加速以及从高性能文件系统自动访问对象存储数据。IBM Spectrum Scale 是业内面向 HPC、AI、分析及其他高性能工作负载的高性能文件访问的领导者。IBM Cloud Object Storage 是一款面向云原生 S3 应用的广泛可扩展且高效的存储解决方案。随着 IBM Cloud Object Storage 等对象存储系统成为企业存储海量数据的一种经济高效且高可靠的方法，客户正在寻找通过高性能和轻松访问将对象数据连接到 AI 和分析工作负载的方法。



### IBM Storage for AI

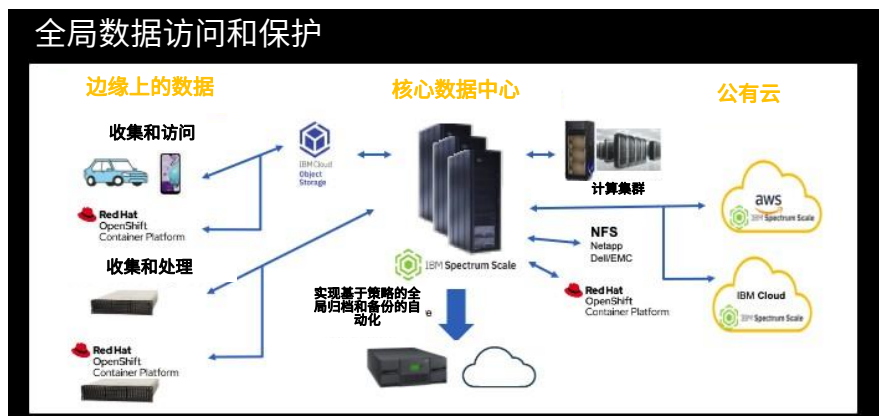
当前的大多数向外扩展 NAS 系统都支持仅将对象存储或云存储用作归档层。如果客户可以将其存储在内部对象存储系统中或直接存储在云端文件存储中的所有数据用于高性能应用，情况将会怎样？客户可以同时将数据摄入到对象存储系统中，并直接利用这些数据来满足 NVIDIA 工作负载或其他高性能需求。现在，通过基于 Spectrum Scale 的新的 AI 加速，客户可以做到这一点。这是一项关键技术，它可将文件存储和对象存储连接到一起，充分利用这两种存储解决方案的最佳功能并将其整合到单个可访问的名称空间之中。该项新功能成为可能主要是得益于 Spectrum Scale 和 AFM 的高级设计，它将有助于推动 IBM Storage for Data and AI 进入许多新的用例和客户环境之中。

IBM 一直都在与 NVIDIA 合作，使用 NVIDIA DGX 系统和 IBM Spectrum Scale 存储构建经验证的参考架构。与 NVIDIA 的合作就是充分利用 IBM Storage for Data and AI 全面产品组合的示例之一。NVIDIA 和 IBM 的合作能够简化企业 AI 环境，并提供所需的存储和 GPU 性能，帮助客户更快实现业务成效。IBM 以往已经证明：当将更多的 ESS 3000 节点连接到 DGX 集群时，IBM Spectrum Scale Storage 和 ESS 3000 2U NVMe 构建块可提供高达 40GB/s 的吞吐量和线性可扩展性。如果将数据摄入或归档到对象存储系统之中，IBM Spectrum Scale 可以透明地将此类数据纳入到 IBM Spectrum Scale 全局并行文件系统之中，实现对 NVIDIA AI 分析的高性能访问。



扩展访问更多数据的示例

IBM Storage for Data and AI 目前提供了优于其他存储选项的独特数据编排功能。这种数据编排功能将 ESS 3000 从“高性能存储”转变为“高性能智能存储层”。它能够将 ESS 3000 连接到组织中的数据湖（文件或对象存储），还能以智能的方式将 AI 建模所需的数据缓存到 DGX Pod 之中。



AI 信息架构

只有 IBM 的产品可以将所需数据从文件和对象存储缓存到容量最高可达 8YB 的全球联合名称空间之中。这种方法不仅有助于我们的客户优化 TCO、提升生产效率，还可以为其 AI 工作负载提供卓越的高性能。

为了提供这种差异化价值，我们增强了 IBM Spectrum Scale 的活动文件管理 (AFM) 功能，使其成为 IBM 的原生数据移动程序。现在，该数据移动程序可以连接到任何基于 NAS 或 S3 的数据存储，例如 IBM Cloud Object Storage 及 IBM Spectrum Scale 全球联合名称空间。

**IBM Spectrum Discover**

- 数据洞察力：
  - 多供应商连接（包括 Kubernetes）
- 更高质量的 AI：
  - 在一秒内完成数十亿条记录的搜索
- 优化数据 workflow：
  - 基于策略的自动化和自动标记数据
- 数据安全和合规
  - 提前发现安全、合规和治理问题
- 业务价值
  - 将数据连接到 IBM Watson 解决方案和 IBM Cloud Pak for Data

借助数据目录和策略引擎实现全面的搜索和 AI 分析

IBM Spectrum Discover

### IBM Spectrum Discover

IBM Spectrum Discover 是一个多源数据目录，每当实时使用元数据进行更改时，该目录便会自动、连续地为对象和文件建立索引。如此一来，便可构建一个功能强大且可自定义的数据库，它带有一个易于使用的界面，允许用户查找和识别最相关的数据，不管数据的类型或位置为何。用户可以使用简单的 SQL 查询命令或可执行的 API 脚本或命令，快速高效地获得数据的综合洞察力。Spectrum Discover 还可用于创建自定义标签和基于策略的工作流，用以编排内容检查并激活人工智能 (AI)、机器学习 (ML) 和分析 workflow 中的数据。Spectrum Discover 可用于更快的 AI 分析、合规分类、图像和视频索引编制、个人数据识别、AI 数据管道集成、实时数据发现，而且可提供新洞察力，优化数据并找出不良或重复数据。数据源包括 IBM Spectrum Scale、IBM COS、AWS S3、NFS 或 SMB 数据源，包括 Netapp、Isilon 和 Red Hat (Ceph 和 OCS)。

## 为什么选择 IBM?

IBM Spectrum Discover 能够基于元数据索引选择正确的数据集进行移动，可谓是数据编排背后的“大脑”。IBM Spectrum Scale 数据移动程序与 Spectrum Discover 结合使用，旨在提供一个全面的数据编排解决方案，为主动归档、数据迁移、缓存或 AI 加速等用例提供服务。如此一来，IBM Spectrum Scale 和 ESS 不仅可为传统的内部 AI/HPC/HPDA 工作负载提供价值，还可为混合云和边缘计算工作负载提供价值。

## 有关更多信息

<https://www.ibm.com/cn-zh/it-infrastructure/storage/ai-infrastructure>

## 下一步行动

- [IBM Spectrum Scale Web 页面](#)
- [IBM Spectrum Discover Web 页面](#)
- [IBM Cloud Object Storage Web 页面](#)

---

© 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](https://www.ibm.com/legal/us/en/copytrade.shtml#section_4) 包含了可能在本文档中提及的所选第三方商标列表。

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

IBM® Spectrum Scale、IBM Spectrum® Discover



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。