

# IBM SpectrumAI with NVIDIA® DGX™

## AI 数据管道引擎

IBM SpectrumAI with NVIDIA DGX 是一款集成式的计算与存储解决方案，它采用最新的系统和软件创新，可为 AI 数据管道的整个生命周期（从数据准备到训练再到推理）提供支持。

### 亮点

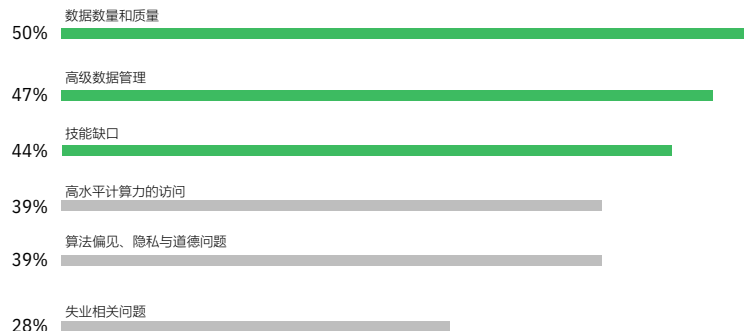
- 聚合型解决方案：可面向最苛刻的机器学习/深度学习项目进行部署
- IBM Spectrum Scale v5：采用软件定义技术，可通过 AI 数据管道简化数据移动
- NVIDIA DGX-1 服务器：专为 AI 和机器学习而构建的解决方案
- NVIDIA DGX 软件堆栈：经过优化可实现 GPU 训练性能的最大化
- 久经验证的数据性能：可实现超过 100GB/s 的吞吐量，单个堆栈可支持多达 9 个 DGX-1 服务器
- 针对您的容器化 AI 工作负载构建共享数据服务时可依赖的基础
- 通过业务合作伙伴简化了支持模式，同时确保了整个解决方案的功能性，由 IBM 和 NVIDIA 联合提供支持

### 挑战：

行业预测报告显示，在未来几年内，几乎所有的技术、服务和科学都将会借助 AI 而得到增强。凭借强大的 GPU 及经优化的 ML/DL 框架（以容器的形式交付）的支持，AI 正在快速为各个行业和领域所采用。计算机视觉与对象检测、人机交互、数据分类及高级范例检测等领域的创新层出不穷，这些技术的应用范围也在不断扩大。机器学习、深度学习和神经网络训练的性能与可用数据的质量和数量密切相关。如何实现数据基础架构与强大服务器的合理搭配是一个巨大挑战，因为确保对海量共享数据的高性能访问非常关键。

如果不能访问适当的数据，数据科学团队的生产效率将会受到严重影响。据广泛报道<sup>1</sup>，数据准备与摄入会占用数据科学家的大部分时间。每个项目都需要训练和测试数据集，还需要对这些数据集进行有序的组织 and 标记，使其可用于模型开发与持续验证。通过对 AI 数据管道进行适当的管理，包括数据治理、可扩展的元数据标记、归档及灵活的共享存储，组织就能够构建一个强大的数据库，用于构建和训练多个模型。

图 1: AI 工作负载部署方面的挑战



IDC 相关调研结果显示，在 AI 工作负载部署方面的三大挑战分别是：数据数量和质量、高级数据管理和技能缺口。<sup>2</sup>

### 解决方案描述：

数据科学生产效率的核心在于 ML/DL 工作负载构建与训练所用的基础架构和软件。一个必要的要求就是要使用高性能的共享存储，同时采用最新的 GPU 服务器来运行经优化的容器 AI 工作负载。随着 GPU 密度的增加，对存储密度和速度的需求也随之提升。全闪存存储技术领域的进步为构建此类存储系统奠定了基础，通过与最新的高速网络相连接，在单个机架配置中最多可驱动 72 个 GPU。此外，还需要专业级的设计和配置，以消除数据瓶颈并确保系统始终以最优性能运行。

不过，若要真正地实现 AI 项目的扩展，数据科学和 IT 团队需要相互协作，改善从数据摄入到推理再到归档的整体 AI 数据管道的效率。一般而言，数据来源会不断发生变化。若要实现优化，就需要一个通用的数据存储，而且这个存储要具有充足的容量和灵活性，可以从多个来源摄入数据，同时支持分析和脚本编辑工具，而且能够剔除使用频率非常少的数据。

在不限制工具选择的情况下消除数据移动需求，将提高团队生产效率、降低成本并简化数据治理需求。软件定义存储可提供一流的可扩展性和管理，使您可以构架数据流的单一视图、向外进行扩展并对系统进行分层，以满足各种业务需求。IBM Spectrum Scale 能够帮助您消除使用传统存储阵列经常会出现的数据访问障碍及数据复制开销。

<sup>1</sup> <https://www.information-management.com/slideshow/how-data-professionals-spend-their-time>

<sup>2</sup> 来源：认知、ML 与 AI 工作负载基础架构市场调研，2018 年 1 月；受访对象数量=405, 1,000 多名员工（美国）；500 多名员工（加拿大）；IDC Technology Spotlight 报告（编号：US43977818）

## 生产效率、灵活性和可扩展性

IBM SpectrumAI with NVIDIA DGX 是一款一流的 AI 数据基础架构，企业可借助该架构高效构建数据科学服务。IBM SpectrumAI 平台是由 IBM 和 NVIDIA 联合推出的一款聚合型产品，专为数据科学的整个生命周期而构建，涵盖数据准备、训练和推理等等。IBM SpectrumAI 采用可组合的架构，该架构使用软件定义存储技术及 NVIDIA 在硬件和软件堆栈上的创新，可帮助您确保数据管道扩展所需的灵活性，而其一流的性能能够轻松应对多种用户和多个模型的需求。IBM SpectrumAI 可为数据科学家提供他们所需的功能、数据和灵活性。

经过测试和调化，IBM SpectrumAI 是一款聚合型、但可单独扩展的计算、存储与联网解决方案，它由 NVIDIA DGX-1、基于全闪存的 IBM Spectrum Scale 和 Mellanox 网络产品构成。由于这种设计，组织可以从小规模开始部署然后不断增加；IBM SpectrumAI with NVIDIA DGX 可以首先部署 3 台 DGX-1 服务器和单个 NVMe 存储系统。如果采用 3 个 NVMe 系统搭配 9 台 DGX-1 服务器的配置，整个堆栈可扩展到 72 个 Tesla V100 Tensor Core GPU，实现 120GB/s 的数据吞吐量。由于 IBM Spectrum Scale 属于软件定义存储，因此可跨越多个机架或其他类型的存储系统，还可实现数 EB 数据的自动分层、数据保护和归档。

NVIDIA DGX-1 是一款全堆栈的 AI 计算解决方案，由 8 个全球最高级的数据中心加速器 NVIDIA Tesla V100 Tensor Core GPU 提供支持。每台 DGX-1 服务器的混合精准训练性能可达到 1 petaFLOPS，而且易用性非常高；此外，它采用了一个经优化的软件堆栈，包括 NVIDIA CUDA、深度学习库和 Docker 容器管理产品。DGX 容器注册表为团队提供了一个扩展的目录，包括经 NVIDIA 优化的 AI 和数据科学容器。NVIDIA DGX 可帮助数据科学家快速开发机器学习和深度学习模型，轻松进行迭代和试验，还能进行规模化的训练和推理。

IBM Spectrum Scale v5 是一个软件定义的并行文件系统，旨在满足要求苛刻的最新工作负载。它曾帮助全球规模最大、最智能的超级计算机<sup>3</sup>实现了扩展，使其可运行超过 40,000 个 NVIDIA GPU；该产品已针对最新的全闪存存储进行了优化，可支持 AI 工作负载在吞吐量和混合大小文件方面的需求。它支持多个协议，包括 NFS、SMB、Object 及 HDFS，可确保用户在选择协议时的灵活性。它的高级数据管理功能可基于脚本、使用范例或策略在存储系统中或云端对数据进行分层，同时确保对用户的完全透明度。

IBM Spectrum Discover 可帮助用户更好地完成数据分类、跟踪和治理；作为一款现代化的元数据管理软件，它能够针对 IBM Spectrum Scale 和 IBM Cloud Object Storage 中的非结构化存储提供数据洞察力。它能够从数百万个文件和对象中摄入、合并元数据并对其进行索引处理，进而为数据科学家、存储管理员和数据管理员提供一个丰富的元数据层，使其能够高效地管理、分类海量的非结构化数据并从中发掘洞察力。

IBM SpectrumAI 的网络连接由 Mellanox 提供。Infiniband 选项的采用可帮助您实现可靠的 100Gbps 网络连接，充分利用 RDMA 等高级技术。由于规模和配置得当，因此可确保存储吞吐量，还可实现 NVIDIA DGX-1 服务器之间的节点间通信。



IBM SpectrumAI 与 DGX-1 POD 搭配组合 (9:3 配置)

<sup>3</sup> <https://www.ibm.com/blogs/systems/fastest-storage-fastest-system-summit/>

## 提升您的 AI 数据管道

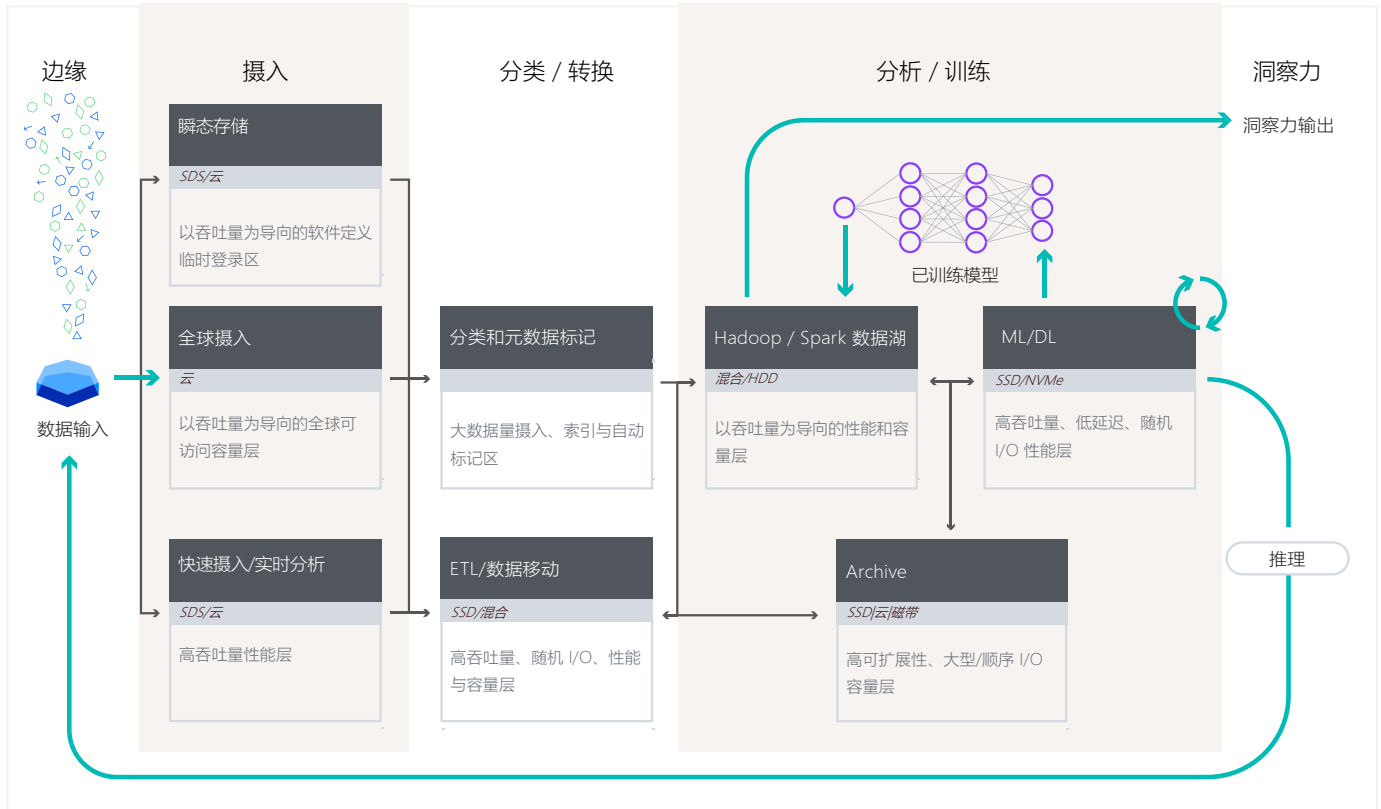
IBM SpectrumAI with NVIDIA DGX 可为组织的 AI 战略奠定一个坚实的基础。该产品由 IBM 及其专业渠道合作伙伴 NVIDIA 联合开发，可帮助数据科学家快速提升工作效率，为未来的工作做好充分准备。借助 NVIDIA 的专业知识和支持，数据科学家可以访问最新的经 NVIDIA 优化的 GPU 加速工具，同时还可确保存储性能。

随着数据需求和业务需求的不断增加，IBM Storage 会提供丰富的选项和创新，以满足各种 AI 数据管道在性能、经济性及数据治理方面的需求。IBM Spectrum Scale 可提供跨多种媒介的存储服务，包括 AWS 公有云。它可以与 IBM Cloud Object Storage 和磁带共享数据和元数据，进而实现地理分散存储或本地存储方面的灵活性。

对于具有前瞻性思维的企业而言，通过合并端到端的 AI 数据管道，他们可以持续提升生产效率、降低成本并简化数据治理；数据科学家可以缩短获取洞察力所需的时间，同时最大程度的减少数据副本的数量；而对于基础架构团队而言，则可以简化管理、确保可扩展性并改善 TCO。

IBM SpectrumAI with NVIDIA DGX 是一款一流的数据科学平台，您可以借助该平台开发并交付新的 AI/ML/DL 应用 - 从数据准备到训练再到推理和可视化。该款久经验证的企业级解决方案可实现快速的插件式/提升式部署，同时还可确保无与伦比的性能。IBM SpectrumAI with NVIDIA DGX 集成了 AI 计算、高性能存储及网络连接等领域的最新创新，可为您提供获得洞察力的最快路径，而且能够在开发生命周期内支持多个项目，同时由经简化的支持模式提供支持，可确保您的 AI 工作负载持续正常运行。

AI 数据管道及其存储需求



### IBM SpectrumAI 的技术功能

- DGX-1 服务器 - 专为 AI 和机器学习构建的解决方案
- 经优化的 NVIDIA DGX 软件堆栈，可最大程度地提升 GPU 加速的训练性能
- IBM Spectrum Scale v5 是一款专为 AI 工作负载而设计的领先软件定义文件存储产品，有助于提升小文件、元数据和随机 IO 存取的性能。
- NVMe 全闪存存储可实现超低延迟，还有助于提升能耗效率和数据密度。借助 IBM Spectrum Scale 的分布式数据保护功能，它能够在每个 2U 构建块中交付超过 300TB 的数据，单个机架的数据吞吐量最高可达到 120GB/s。
- 该产品可实现多个机架、其他 IBM SpectrumAI 配置及工作站之间的无缝数据管道连接，因此能够为数据科学家提供有关其 IBM SpectrumAI 数据环境的统一视图。
- 通过与 IBM Spectrum Discover 相集成，该产品可在 IBM Spectrum Scale 和 IBM Cloud Object Storage 中实现可扩展的数据治理与元数据标记。



拨打免费咨询热线

手机请拨打：400-810-1818 转 5122

座机请拨打：800-810-1818 转 5122

更多问题，请联系我们

[https://www.ibm.com/account/reg/cn-zh/signup?  
formid=MAIL-storage](https://www.ibm.com/account/reg/cn-zh/signup?formid=MAIL-storage)

© Copyright IBM Corporation 2018

IBM Systems  
3039 Cornwallis Road  
RTP, NC 27709

美国印刷

IBM、IBM 徽标及 ibm.com 是 International Business Machines Corporation 在美国和/或其他国家/地区的商标。如果这些和其他 IBM 商标术语在此类信息中首次出现时使用商标符号 (® 或 ™) 做了标记，则表明在此类信息发布时，这些术语已在美国进行了注册或者已为 IBM 所拥有的普通法商标。这些商标也可能是在其他国家或地区的注册商标或普通法商标。Web 站点 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) 上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表。

NVIDIA、NVIDIA 徽标及 DGX-1 是 NVIDIA, Inc. 或其分公司在美国和/或其他司法辖区的商标或注册商标。

Linux 是 Linus Torvalds 在美国和/或其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 及 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

其他产品、公司或服务名称可能是其他公司的商标或服务标记。

在本出版物中，但凡提及 IBM 产品或服务时，并不表示 IBM 可以在 IBM 业务所涉及的所有国家或地区予以提供。