

使用 IBM Spectrum Scale 优化 Cloudera 分析

提高生产效率、降低成本并提升 SLA

亮点

- Spectrum Scale 已经过认证，可与 CDP Private Cloud Base 一同使用
- 易于管理：单个全球数据湖
- 经济性：优化成本和性能
- 稳健：确保数据可用性、完整性和内置安全性
- 全局并行性能：消除瓶颈

通过企业级数据服务在单个集群中将边缘连接到核心再到公有云，进而实现高性能并行数据访问

AI 分析

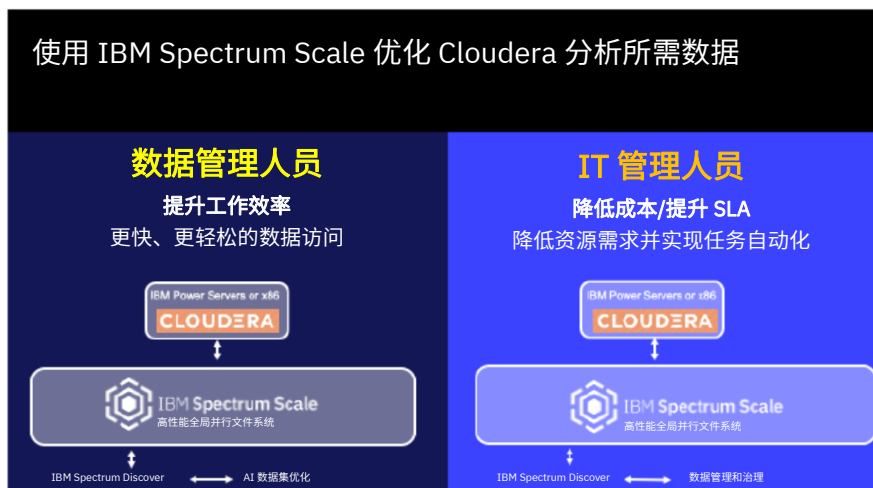
HPC

高性能工作负载



高性能存储

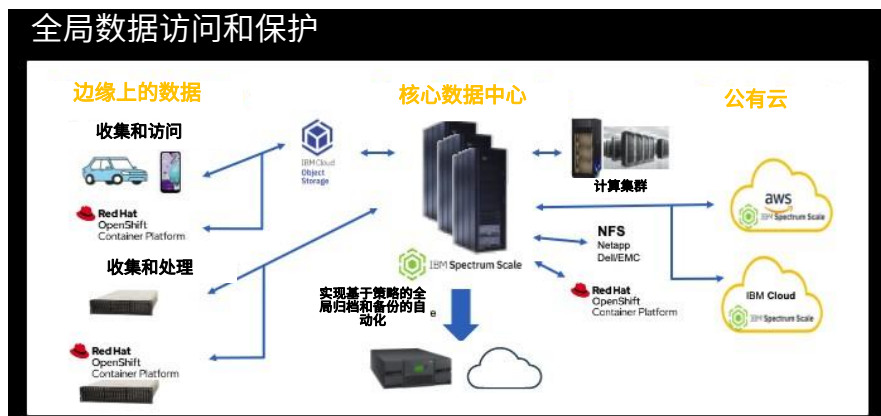
人工智能和分析平台需要软件平台能够与针对海量数据进行优化的基础架构平台和谐整合，这已成为解决方案不可或缺的一部分。很多时候的数据问题都是因为性能瓶颈的持续存在，以及数据未能在适当的时间放置在适当的位置。由于经常使用多个数据副本并且没有基于策略的归档功能，因此数据还会很快变得成本高昂。数据仍然是大数据工作负载的关键组件，但通常只是事后才想到的商品组件。存储基础架构必须提供一组灵活且可扩展的服务，提供最佳的高级分析和 AI 应用体验。基础架构平台必须确保对经济高效的数据和存储资源的动态、高性能访问。



IBM 和 Cloudera

Cloudera 与 IBM Storage 的合作促使这些功能被整合到单个解决方案之中，通过释放大数据中潜藏的洞察，为数字化企业赋能。2019 年，Cloudera 发布了 Cloudera Data Platform (CDP) Data Center（现为 CDP Private Cloud Base），这是从边缘到 AI 的集成分析方面最全面的内部平台，涵盖了摄入、处理、分析、实验和部署等多个环节。它结合了 Cloudera Enterprise Data Hub 和 HDP Enterprise Plus 的优势，融合了最新、最一流的开源数据管理和分析技术，这些技术被集成到一起并针对数据中心内的部署进行了优化。目前，Cloudera 已通过验证，确认 Spectrum Scale 可与 CDP Private Cloud Base 一同使用。

IBM Spectrum Scale 和 Elastic Storage Systems (ESS) 构件块提供了一个运行 IBM Spectrum Scale 软件的集成存储系统（该系统可确保高性能访问），还提供了一个可确保企业级数据服务（包括优化的数据放置）的全局并行文件系统。在 IBM Storage 和 ESS 基础上整合使用 Cloudera 的客户通过利用 IBM Spectrum Scale 的数据生命周期管理和数据策略，以及将所有数据保存在单个全球混合云存储系统上的能力，可以实现高达 60% 的存储基础架构空间节省。



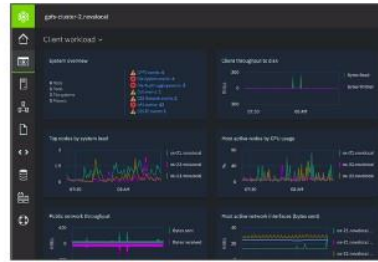
AI 信息架构

IBM Spectrum Scale 是一个基于全局并行文件系统架构的软件定义存储系统，该架构提供了一个高性能的文件、对象和经优化 HDFS 接口，该接口支持 HDFS API。IBM Spectrum Scale 可以用作 CDP 环境的存储层，替代原生 Hadoop 分布式文件系统 (HDFS)。对 HDFS API 的支持可实现对企业存储的原位分析，而不是将数据从企业存储复制到分析孤岛。原位分析不仅消除了数据重复，还避免了针对陈旧数据运行分析所带来的问题。Spectrum Scale 提供了一个从边缘到核心数据中心再到公有云的全球联合存储。这意味着数据用户和业务线可以在单个位置充分利用海量数据执行分析和 AI 工作流。对 POSIX 访问的支持可实现超快速数据摄入。此外，Spectrum Scale 还提供了一个容量可扩展到数 YB 的环境及企业级数据服务，以优化数据放置并降低数据容量成本。数据管理员对所有的自动数据优化称赞有加，这些优化能够提升基于策略的归档和安全功能的效率，进而让数据管理员的工作变得更轻松。到 CDP 的共享存储允许计算和存储解耦，以实现经优化的低成本配置。

IBM Spectrum Scale

- 快速实现成效：
全球混合云并行文件系统
- 降低运营开支：
简化管理、确保在线可扩展性
- 降低资本开支：
基于策略的数据优化及透明的数据生命周期管理
- 保护资产、降低风险：
数据可用性、完整性和安全性
- 实现数据访问的现代化：
经优化的容器原生存储

面向 AI 和混合云数据中心的全面企业级存储



IBM Spectrum Scale

面向 Cloudera 的 Spectrum Scale，其背后的价值理念很简单：通过企业级数据服务在单个集群中将边缘连接到核心再到公有云，进而实现高性能并行数据访问。这使得 Spectrum Scale 在 AI/分析、HPC 和高性能工作负载方面首屈一指。

安全访问，随处可行。
企业级数据服务，全覆盖。
混合云，全员支持。

用例	AI/ML	IBM Cloud Pak	Cloudera/Hadoop	HPC	视频和图像	大数据分析	NVIDIA	数据协作
集成	IBM FileNet	IBM Watson	Splunk	Red Hat OpenShift	Radars	Spark	NVIDIA	CLUSTERFLUX
组织	IBM Spectrum Discover							
混合云文件系统								
边缘 数据中心 私有云 公有云								

全球数据访问

为什么选择 IBM?

IBM Storage for Data and AI 不仅仅是存储产品或存储解决方案。它蕴含着一种可帮助客户完成 AI 与混合云数据中心之旅的存储战略。IBM 将会继续强化其在可扩展、高性能工作负载，以及面向基于文件和对象的解决方案的高效、安全、可扩展的容量存储方面的领导地位。我们的产品为 AI 和混合云提供了一种经增强的战略。我们为边缘、核心数据中心以及 Kubernetes 容器和 Red Hat OpenShift 平台等公有云提供了一个面向未来的基础。IBM Storage for Data and AI 面向可融入整个组织的混合云，帮助企业降低复杂性和成本，实现与 AI 信息架构的增强集成。我们的宣传口号很简单：“面向数据和 AI 的存储化繁为简”。

下一步行动

- [IBM Spectrum Scale 技术资料](#)
- [IBM ESS 技术资料](#)

有关更多信息

访问我们的解决方案 Web 页面：
<https://www.ibm.com/cn-zh/it-infrastructure/storage/ai-infrastructure>

© 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® Spectrum Scale



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。