



面向 AI 的存储：

**从数据摄取到洞察力实
现的快速通道**



01

引言

02

无 IA（信息架构），不 AI

03

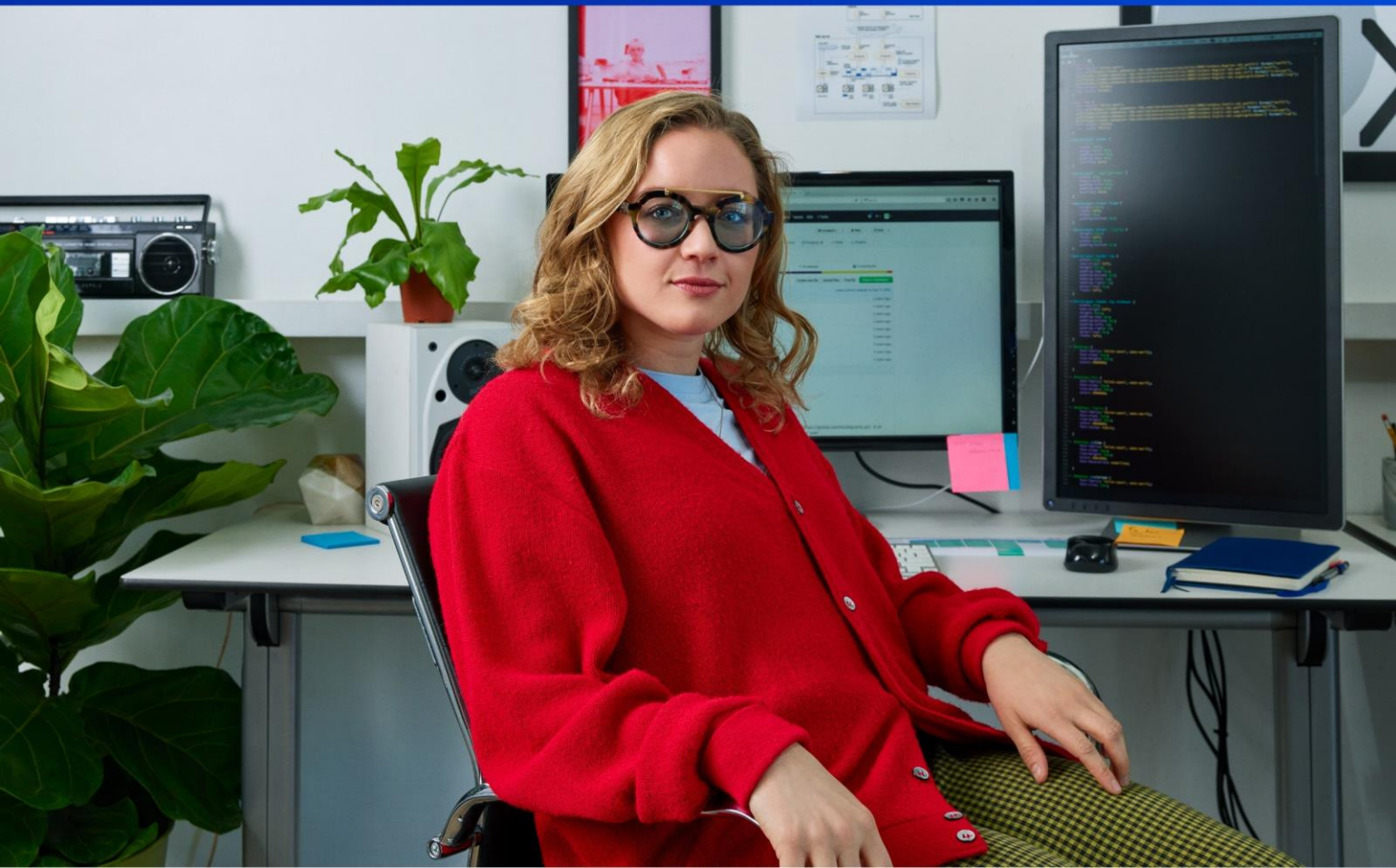
构建强大的基础

04

案例研究：建立竞争优势

05

结语



引言

如今，企业都认同一点：人工智能（AI）为他们提供了开展创新和提高生产效率的捷径。大多数企业已经在测试、采用和实施 AI，并释放 AI 的所有潜能。结果就是，企业的 AI 解决方案投资预计将在未来几年大幅增加。

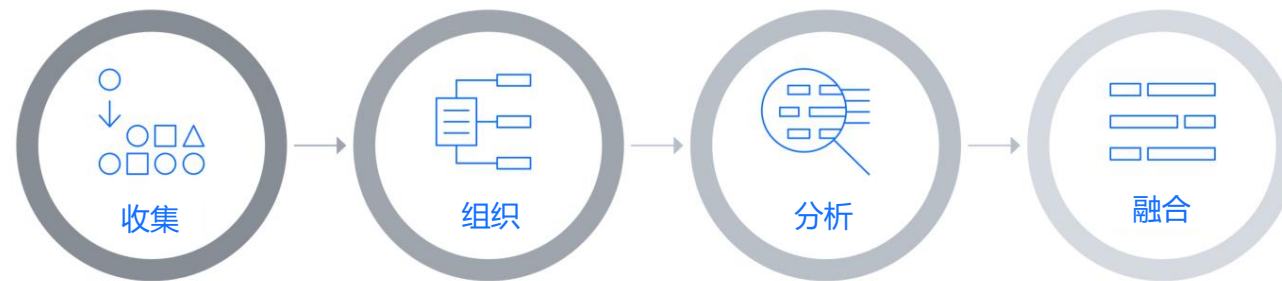
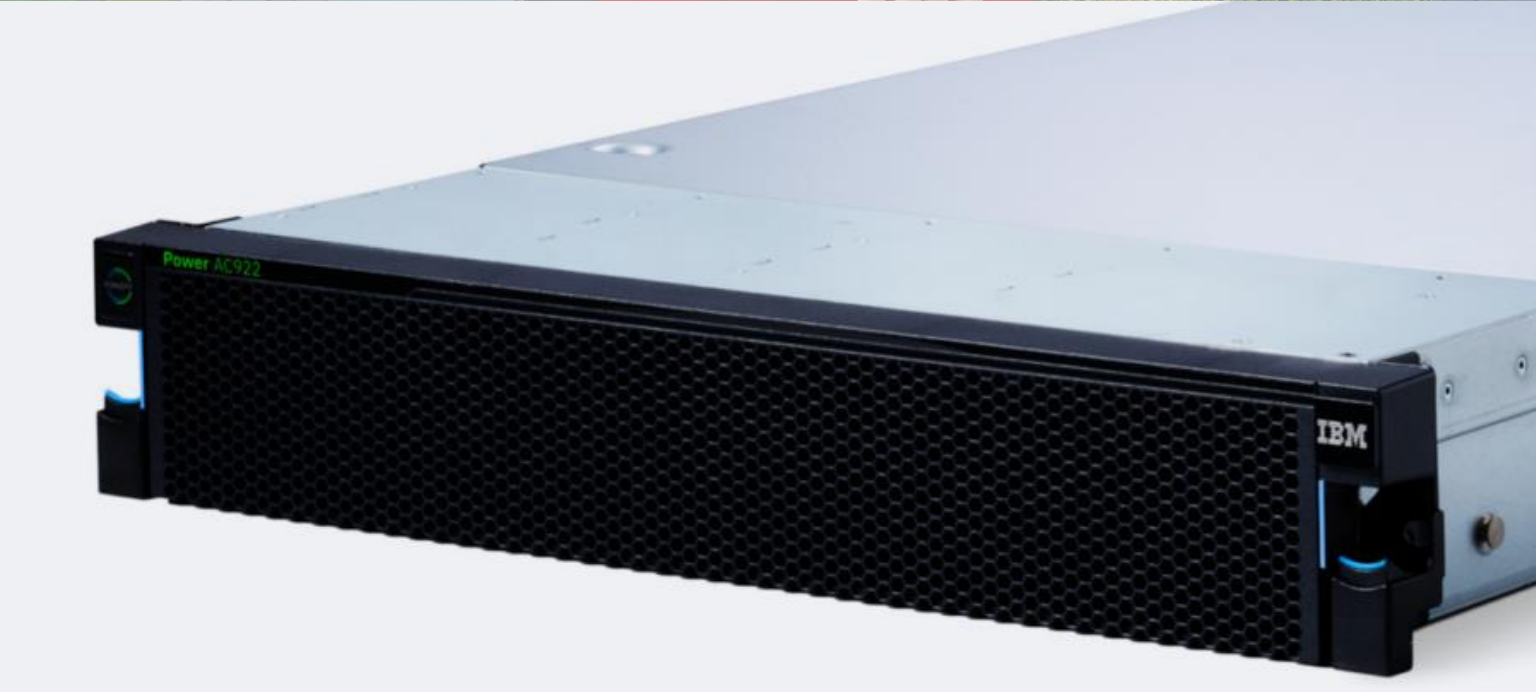
每个成功的 AI 项目都会经历一个多步骤流程，这个流程的第一步是拥有正确的数据，最后一步是广泛使用 AI。

40%

到 2019 年，40% 的数字化转型项目将使用 AI 服务¹

47.9 亿美元

2019 年用于 AI 的 IT 存储开支达到了 47.9 亿美元¹



让数据变得简单、数据访问变得更便捷

构建值得信赖的分析基础

利用机器学习，扩展洞察力

跨流程进行预测，并实现自动化和扩展

企业在采用 AI 时肯定会碰到挑战。开源和商用开发人员工具与框架能够让首个 AI 项目或概念验证的交付变得很容易。但是，企业在支持 AI 开发团队或者部署和扩展生产 AI 工作负载时，却面临挑战¹。

- **数据量和质量。**AI 离不开高质量、多样化且带标签的数据输入。在具有动态数据特征的多个数据源中识别正确的数据集是一项棘手的工作。

- **高级数据管理。**对于需要不断测试、复用和扩展数据集以提高 AI 模型准确性的开发人员而言，整理和跟踪 AI 项目中的数据集已成为一项挑战。
- **技能缺口。**企业对 AI 服务的需求与日俱增，这意味着，他们对高技能专家的需求也将随之增加。由于 AI 依然是一个相对新的领域，企业很难找到经过培训的人才和最佳实践，来推动数据科学生产力。

这也就难怪很多企业不知道如何推进下一步工作，也不清楚如何以最佳方式让 AI/机器学习为己所用。这也是 IBM 全程助您一臂之力的原因。

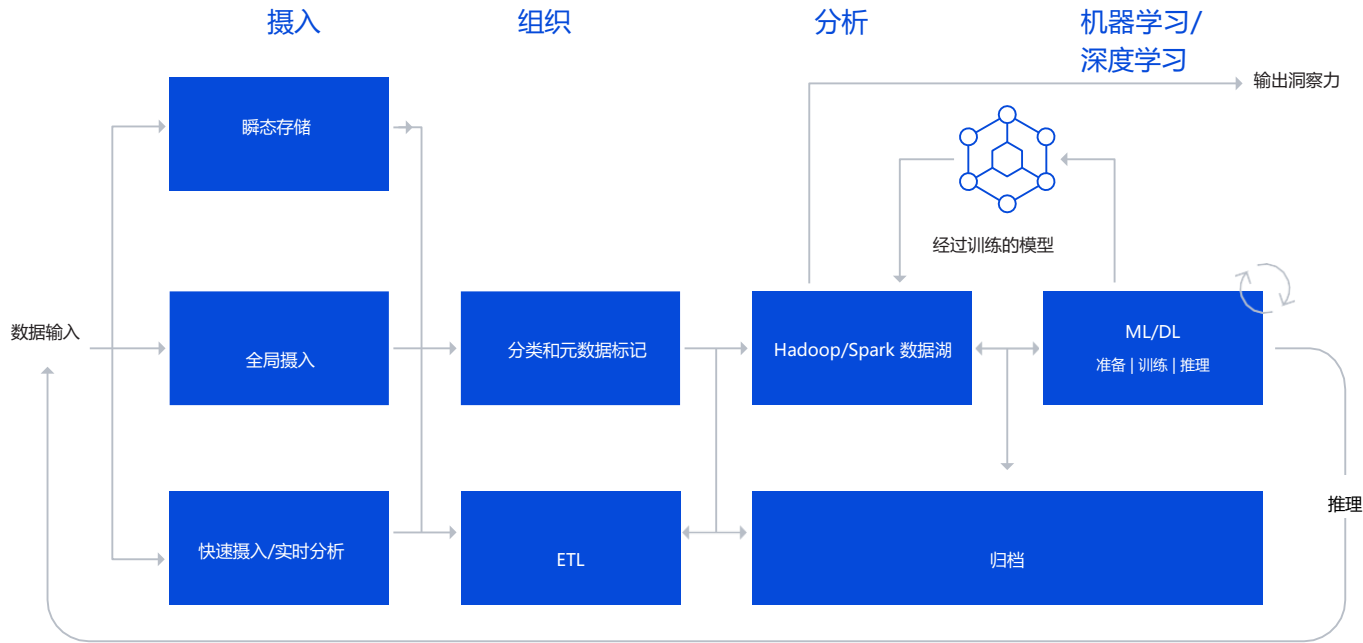
企业在支持 AI 开发团队或部署 AI 工作负载时，数据量和质量、高级数据管理以及技能缺口是他们面临的核心挑战。





无 IA（信息架构），不 AI

AI 管道指的是您摄取、整合和分析数据乃至最终训练模型以便从数据中生成 AI 驱动型洞察力的方式。AI 管道对于高效的数据科学至关重要。AI 管道的效率与您利用正确的 IT 基础架构解决上述挑战直接相关。



统一的数据访问

数据孤岛是阻碍企业高效利用数据，尤其是在 AI 领域高效利用数据的一大障碍。实施 AI 项目时，收集数据是最耗时的阶段。您投资的数据集整理和分类技能应该应用于所有 AI 项目。为了实现这一目标，您需要一个满足以下要求的数据和存储基础架构：最小化延迟；提高效率；在多个项目中启用通用、共享的数据，并支持各种数据分析工具。

数据吞吐量性能

AI 模型的准确性取决于良好的数据输入以及用于对此进行分析的充足计算资源。图形处理单元 (GPU) 常用于 AI，因为 GPU 能够快速分析大型数据集。IT 基础架构必须配合相应的存储性能，才能匹配计算资源对数据的使用能力。同样的，可以用于获取实时洞察力的流式数据，需要具有多种能够正确分配数据工作负载的属性。

通过支持容器，提高敏捷性

通常企业在容器内管理 AI 项目，因为容器具有轻量级、可快速部署的特性，可与多个程序和脚本一起使用。要想从初期试验迅速扩展至生产级别的 AI，您需要采用能与 Kubernetes 和 Red Hat OpenShift 协同运行的永久性存储。容器不仅能简化开发，还能提高 IT 基础架构的敏捷性，以应对企业 AI 服务需求的增长。

构建强大的基础

开发 AI 实践看似复杂，但是其实大可不必如此复杂。如果您基于稳固的基础实施 AI 项目，AI 项目将变得更加简单，成功几率也会更高。面向 AI 的 IBM 存储提供了这样的基础，同时还提供了一系列产品，旨在帮助您解决与部署 AI 工作负载有关的重大业务挑战，进而将您推上实现 AI 生产力的快速通道。

IBM Spectrum® Scale

IBM Spectrum Scale 是一款高性能文件系统解决方案，它能够自动随着存储基础架构的扩展而扩展，并确保存储基础架构的统一。它是一款软件定义解决方案，通过快速、高效地将文件数据移动至最佳存储层，实现性能与成本之间的平衡。借助 IBM Spectrum Scale，您可以安全地收集和整理数据，利用统一的数据基础提供对任意位置数据的访问，这让 AI 采用变得简单。

[了解 IBM Spectrum® Scale](#)

IBM Cloud™ Object Storage

IBM Cloud Object Storage 赋予了云原生应用和 AI 框架性能与可扩展性。这是一个安全的软件定义存储平台，它能够轻松将容器和吞吐量从 TB 级扩展至 EB 级。假如团队使用的是需要数据安全性或高性能本地数据的最新云开发环境，那么 IBM Cloud Object Storage 是他们的理想选择。

[了解 IBM Cloud Object Storage](#)

IBM Spectrum Discover

IBM Spectrum Discover 是一个现代化的元数据管理软件，它能够跨越多个存储平台（包括公有云）快速摄取、整合和索引元数据。借助 IBM Spectrum Discover，数据科学家和存储团队能够对元数据执行高效的统一、编目及扩充操作，从不断增长的非结构化数据多样存储中挖掘更多洞察力，最终提高生产效率。

[了解 IBM Spectrum Discover](#)



案例研究：建立竞争优势

发挥数据的力量，您能够提供巨大的竞争优势。AI 是解锁数据价值、以创新方式变革企业的钥匙，具体方式如下：

- 预测和重塑未来结果
- 优化员工队伍，让员工从事更高价值的工作
- 实现决策、流程和体验的自动化
- 重构业务模式

下面我们将介绍 IBM 客户如何使用 IBM 存储解决方案，改进整个数据生命周期管理，加速 AI 之旅，并实现企业转型：

结果：

运行时缩短 96%

基因组分析管道的运行时缩短 96%

1/3 的成本 大规模执行同样的工作，其成本只相当于商用解决方案的 1/3

2 周 从概念设计到功能全面的云端 HPC 环境只需 2 周

[阅读案例研究](#)

L7 Informatics

用于突破性研究的高性能 Genomic Cloud

基因组学指的是有机体的整个 DNA 组研究，这门学科需要科学家处理海量数据。因此，很多组织在处理自己生成的大量数据时都面临重重困难。

L7 Informatics 与 IBM 携手，合作构建高性能计算 (HPC) 环境，进而利用 IBM Spectrum Storage 技术，以便：

- 统一数据
- 处理大量非结构化数据
- 提供对数据的并行访问，同时消除瓶颈
- 提供内置分层，支持数据的灵活移动
- 允许从实验室无缝迁移至云端，用于分析和长期存储



伯明翰大学

通过管控数据，推动创新型研究的发展

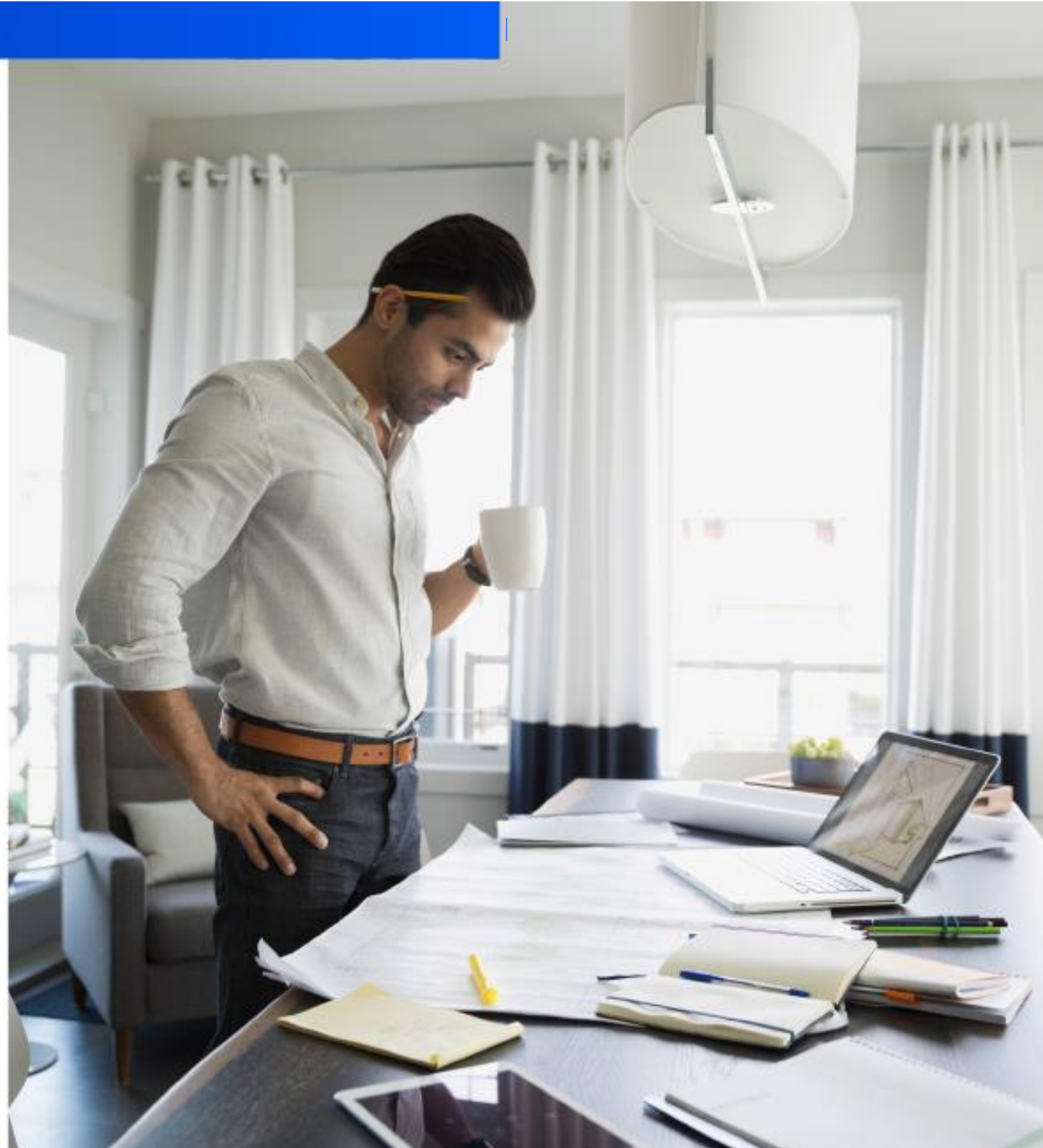
当今研究模拟所生成的数量远远超过以往任何时候。为了满足这一日益增长的需求，伯明翰大学部署了 IBM Spectrum Scale 和 IBM Spectrum Protect，旨在：

- 提供跨越多个存储系统的单一数据管理平台
- 在匹配工作负载与平台时，做出性价比更高的决策，避免因复杂性的增加而失去可控性
- 允许研究人员在适当的位置部署应用，即时获取数据

我们支持多个领域的研究，包括利用和开发各种技术来使用 AI 和深度学习。比如，我们与诺丁汉大学合作开展 Centre of Membrane Proteins and Receptors [COMPARE] 项目。通过分析最新一代显微镜生成的超高清图像，该项目能够为更好地预防和治疗心血管疾病、呼吸系统疾病与癌症提供启示。

Simon Thompson，伯明翰大学研究计算基础结构架构师

[阅读案例研究](#)





结语

AI 之旅起步于一个成功的概念验证，然后迅速扩展至整个企业。为了成功驾驭 AI 之旅，您首先应该建立一个强大、敏捷的 IT 基础，该基础可优化以满足独特的数据要求，推动数据生产力和采集。理想的存储平台必须交付 AI 项目要求的性能、可扩展性和灵活性。您在建立基础时所制定的决策将产生深远的影响，它们会影响您在 AI 之旅的每一步，甚至最终决定您的成败。正因为此，从一开始就拥有理想的合作伙伴这一点至关重要。

面向 AI 的 IBM 存储能够端到端地优化数据管道，进而改进数据治理，加速获取洞察力。通过结合利用行业领先的产品、创新成果和经过验证的领导力，IBM 能帮助您构建所需的基础架构，用于管理数据，处理 AI 工作负载，发挥 AI 的潜能，最终更深入地挖掘洞察力，进而获得更卓越的业务成果。

[了解面向 AI 的 IBM 存储如何将贵企业推上从数据摄取到洞察力实现的快速通道。](#)

[探索面向 AI 的 IBM 存储](#)

资源

1. IDC Worldwide Storage for Cognitive/AI Workloads Forecast, 2018-2022

© Copyright IBM Corporation 2019. U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp. 注：IBM 的 Web 页面中可能包含有应遵守的其他所有权声明和版权信息。

IBM、IBM 徽标及 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 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。

