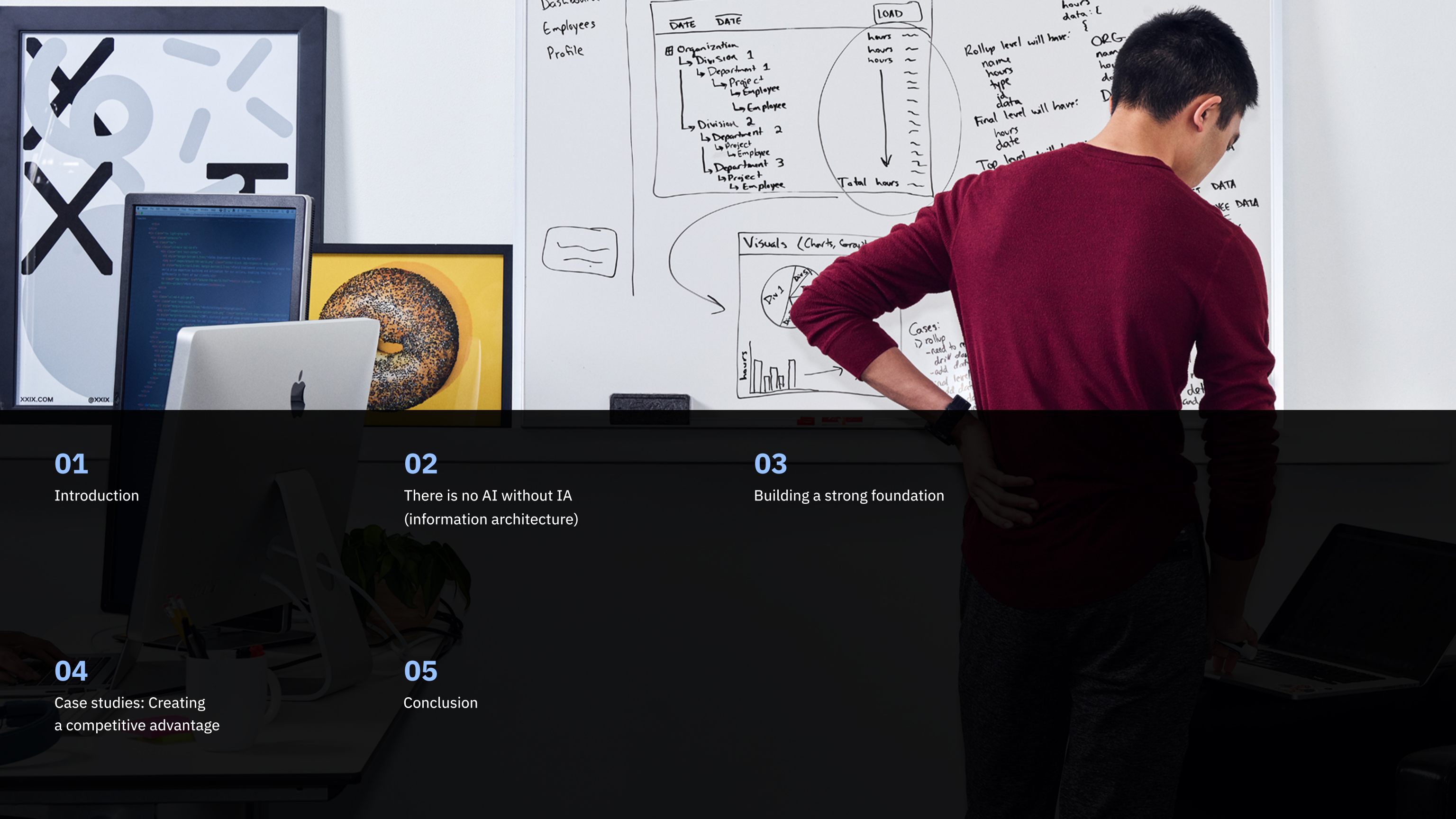




Storage for AI:

The fast track to
an AI information
architecture



01

Introduction

02

There is no AI without IA
(information architecture)

03

Building a strong foundation

04

Case studies: Creating
a competitive advantage

05

Conclusion



Introduction

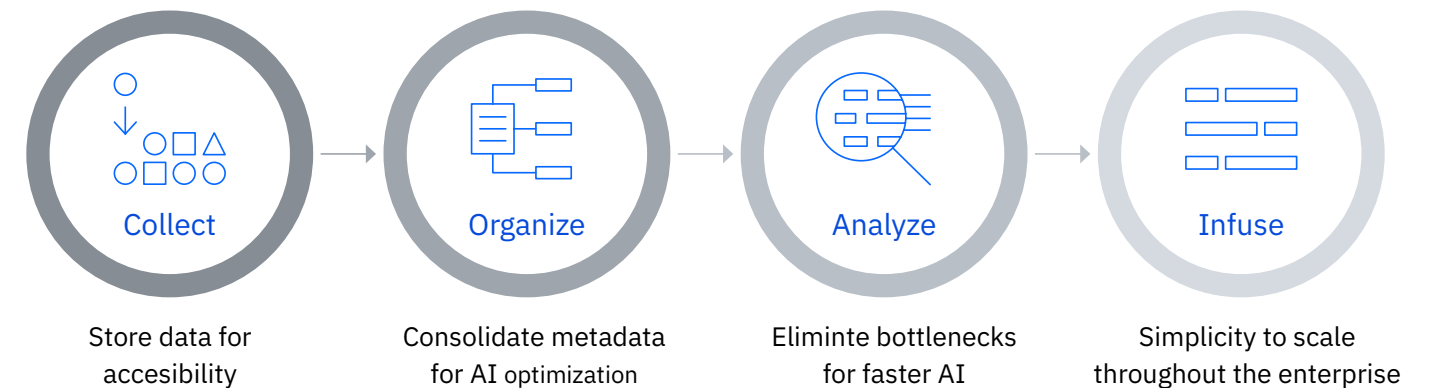
AI is a journey that begins with data. Data is the fuel for AI, and AI cannot exist without an IA (information architecture). The best AI is built on a foundation of data that is collected and organized as carefully as it is analyzed and then infused into the business. Organizations are challenged with gaining insights from their data for many reasons. Data silos make it difficult to get a holistic view of all your information, limiting the value of AI. Current infrastructure that was not built for AI is not flexible enough to respond to new demands without adding complexity.

Every successful AI project goes through a multi-step process that starts with having the right data and progresses to using AI broadly.

IDC predicted that storage spending on AI systems will reach

\$10.1 billion by 2020

¹



Adopting AI is not without its challenges. Open source and commercial developer tools and frameworks make it straightforward to deliver your first AI project or proof-of-concept. However, organizations face challenges¹ when supporting AI development teams or deploying and scaling production AI workloads:

- **Data volume and quality.** AI requires high-quality, diverse, and labeled data inputs. Identifying the right data sets across multiple data sources with dynamic data characteristics can be daunting.

- **Advanced data management.** Organizing and tracking data sets in AI projects is a challenge for developers who need to repeatedly test, re-use and expand data sets to improve AI model accuracy.
- **Skills gap.** The increasing demand for AI services means a corresponding increase in the need for skilled professionals. Since AI is still a relatively new field, it's difficult to find trained personnel and best practices for data science productivity.

It's no surprise that many organizations aren't sure how to proceed and don't have a clear understanding of how best to leverage AI/ML to their advantage. That's why IBM is here to help you at every step along the way.

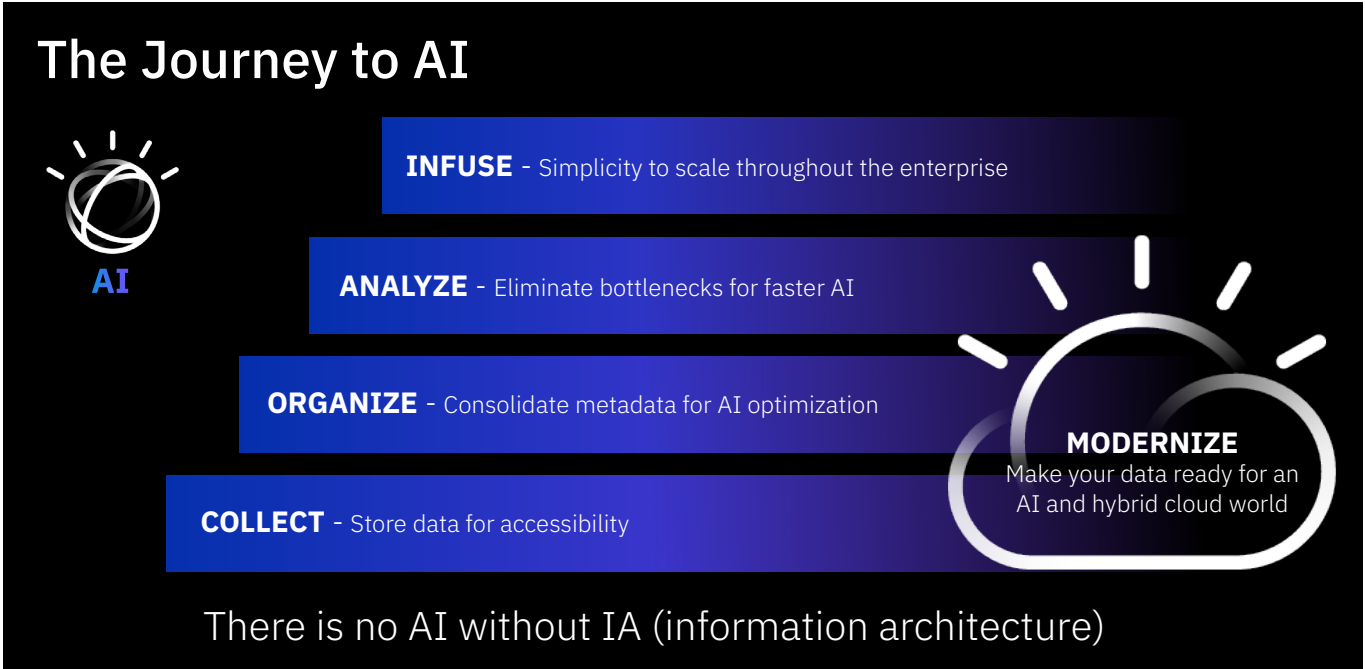
Data volume and quality, advanced data management, and a skills gap are among the core challenges organizations face when supporting AI development teams or deploying AI workloads.





There is no AI without IA (information architecture)

The AI Ladder is a framework that accelerates your ability to collect and organize data, gain deeper insights by leveraging AI-driven data analysis, and infuse your entire enterprise with these capabilities and insights.



Collect data

Data is the fuel that powers AI, but it can become trapped or stored in a way that makes it difficult or cost prohibitive to maintain or expand. Customers need to unleash that data so it can expand from edge to inference in a simple and cost-effective infrastructure. IBM Storage for data and AI makes data simple and accessible for a hybrid multicloud infrastructure with AI storage solutions that fit your business model.

Organize data

AI can only be as good as the data it relies on. Businesses must fully understand what data they have so they can leverage it for AI and other organizational needs, including compliance, data optimization, data cataloging and data governance. IBM Storage for data and AI delivers data in real-time which means as new data is ingested it automatically is updated to the storage catalog and can leverage the policy engine for advance search or even integration to IBM Cloud Pak® for Data.

Analyze data

Analysis is critical to the AI journey and must provide high performance for fast analysis and easily connect to both the data lake and the storage catalog. Organizations must plan for issues beyond the deployment of AI; they need to build AI infrastructures with confidence in scalability. IBM Storage for data and AI provides a simple and integrated AI infrastructure for analysis and the ability to infuse throughout the organization.

Infuse data

Business challenges can become an opportunity to explore, understand, predict and bring an AI infrastructure to your entire organization. IBM Storage is empowering customers to use data and AI storage in order to leverage that infrastructure in more ways that bring value to the organization.

Building a strong foundation

Growing an AI practice seems complicated, but it doesn't have to be. AI projects are easier and more likely to succeed if they're built on a solid foundation. IBM Storage for AI provides that foundation, with a collection of offerings that put you on the fast track to AI productivity by addressing the top business challenges associated with deploying AI workloads.

IBM Spectrum® Scale

IBM Spectrum Scale is a high-performance file system solution that automatically grows with and unifies your storage infrastructure. It is software-defined to balance performance and costs by moving file data to the optimal storage tier quickly and efficiently. IBM Spectrum Scale is available both as a software only solution or as an integrated appliance. IBM Spectrum Scale provides continuous real-time updates to IBM Spectrum Discover for organizing data for AI.

IBM Elastic Storage® System

The new ESS is designed for data lakes with increased performance, density and scalability. With a new ESS generation, consolidate massive data volumes and increase simplicity, speed and exabyte scalability.

ESS 3000 combines the speed of NVMe flash in a small 2U building block that can scale both capacity and performance using one to thousands of nodes.

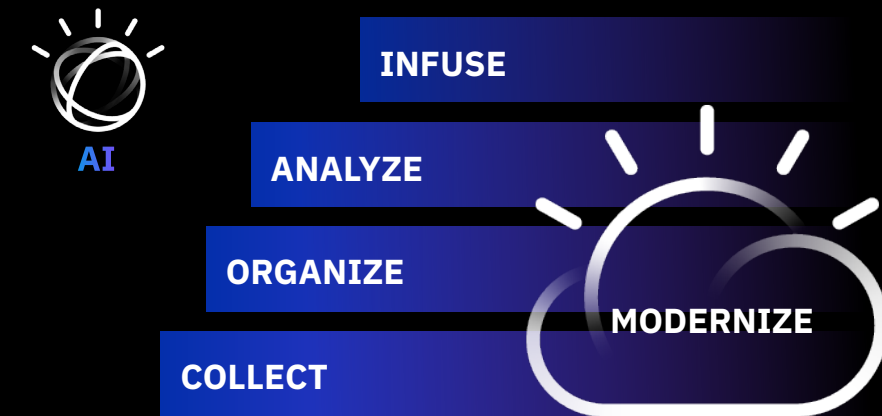
IBM Cloud™ Object Storage

IBM Cloud Object Storage is a software-defined storage platform that easily scales capacity and throughput from terabytes to exabytes. With space saving geo-dispersed data protection and cloud native API access it is an excellent choice for building EB scale cloud storage data lakes and optimized AI storage. IBM Cloud Object Storage provides continuous real-time updates to IBM Spectrum Discover for organizing data for AI and can seamlessly integrate with IBM Spectrum Scale and the Elastic Storage Systems to provide high performance access to data.

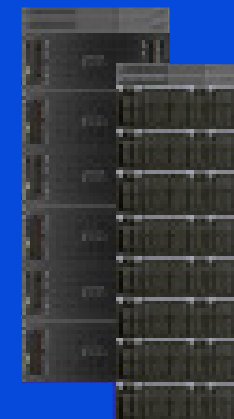
IBM Spectrum® Discover

IBM Spectrum Discover is a storage data catalog with modern metadata management software that can rapidly ingest, consolidate, and index metadata across multiple storage platforms, including public cloud. It increases productivity by enabling data users to efficiently unify, index, and enrich the data catalog to increase insights and validate governance in seconds instead of hours or days from their growing, diverse stores of unstructured data. With one click integration data can be seamlessly integrated to IBM Cloud Pak for Data.

Comprehensive Storage Solutions for the AI Journey

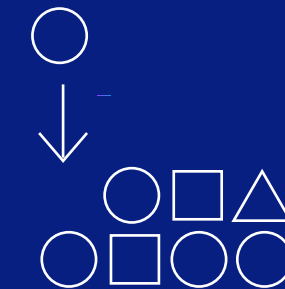


COLLECT



ESS 5000, IBM Spectrum Scale, Cloud Object Storage

ORGANIZE



IBM Spectrum Discover

ANALYZE



ESS 3000, IBM Spectrum Scale

Case studies: Creating a competitive advantage

Harnessing the power of your data provides a significant competitive advantage. AI is one key to unlocking the value of that data and transforming your business in innovative new ways, including:

- Predicting and shaping future outcomes
- Optimizing your workforce to engage in higher-value work
- Automating decisions, processes, and experiences
- Reimagining business models

Here's how some of our clients have used IBM Storage for data and AI to improve management of the entire data life cycle, accelerate their journey to AI, and transform their organizations:

Results:

96% reduction in runtime

for a standard genome analysis pipeline

1/3 the price

of using commodity solutions to perform the same work at scale

2 weeks from conceptual design to fully-functional IBM HPC environment that leverages the cloud

[Read the case study](#)

L7 Informatics

High-performance Genomic Cloud for ground-breaking research

Genomics — the study of an organism's complete set of DNA — requires scientists to process vast amounts of data. As a result, many organizations struggle to cope with the huge volume of data they generate.

L7 Informatics teamed up with IBM to build a high performance computing (HPC) environment that leverages IBM Spectrum Storage for Data and AI technology to:

- Unify data
- Work with high volumes of unstructured data
- Provide parallel access to data with no bottlenecks
- Provide built-in tiering for flexible data movement
- Allow seamless migration from labs to cloud for additional analysis and long-term archival storage.



University of Birmingham

Driving innovative research forward by taking control of data

Today's research simulations generate more data than ever before. To meet this ever-increasing demand, the University of Birmingham deployed IBM Spectrum Scale and IBM Spectrum Protect to:

- Provide a single data management plane across multiple storage systems
- Enable price-performance decisions when matching workloads to platforms, without causing complexity to spiral out of control
- Allow researchers to deploy applications where it makes sense with immediate data availability

Results:

Supports compliance with data protection regulations at low cost and without disruption

Up to 2 FTEs estimated saving due to enhanced operational efficiency

5,000 researchers supported by infrastructure that helps them find solutions to key issues faster

We support research in a wide range of areas, including applying and developing techniques to use AI and deep learning. For example, we're collaborating with the University of Nottingham on the Centre of Membrane Proteins and Receptors project. By analyzing the super high-resolution images produced by the latest generations of microscopes, the project will shed light on how cardiovascular disease, respiratory disorders and cancer can be better prevented and treated.

Simon Thompson, Research Computing Infrastructure Architect, University of Birmingham

[Read the case study](#)





Conclusion

The journey to AI starts with a single successful proof-of-concept, and creating a simple and comprehensive AI infrastructure that can be infused throughout the organization. Navigating that journey successfully starts with creating a robust, agile IT infrastructure foundation optimized for the unique data requirements that drive productivity and adoption. The right storage platform must deliver simplicity, performance, scalability, and flexibility, which AI projects demand. The decisions you make as you build that foundation have far-reaching implications that will impact you at every step along the way and, ultimately, determine your success. That's why having the right partner from the outset is critical.

IBM Storage for AI provides end-to-end optimization of AI Journey to improve data governance and accelerate time to insights. By combining industry-leading offerings, innovation and proven leadership, IBM enables you to build the infrastructure you need to manage your data, handle AI workloads, leverage the power of AI, and ultimately drive deeper and faster insights that create better business outcomes.

[Learn how IBM Storage for Data and AI can put your organization on the fast track to an AI information architecture.](#)

Resources

1. IDC Worldwide Artificial Intelligence Market Shares, 2018, Doc #US45334719, July 2019.

© Copyright IBM Corporation 2020. U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp. NOTE: IBM web pages might contain other proprietary notices and copyright information that should be observed.

IBM, the IBM logo and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

