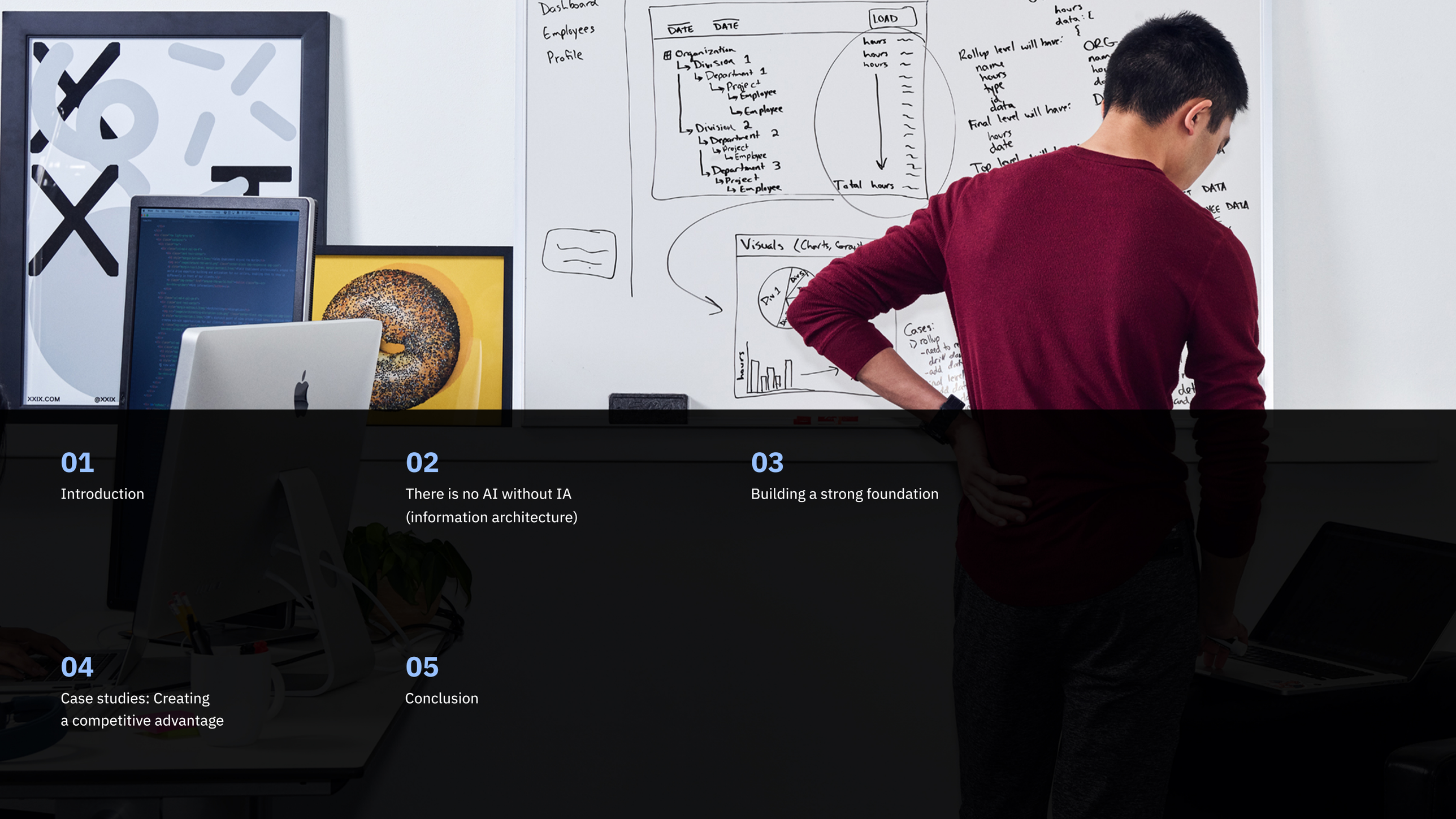
Storage for AI:

The fast track from ingest to insights



Introduction

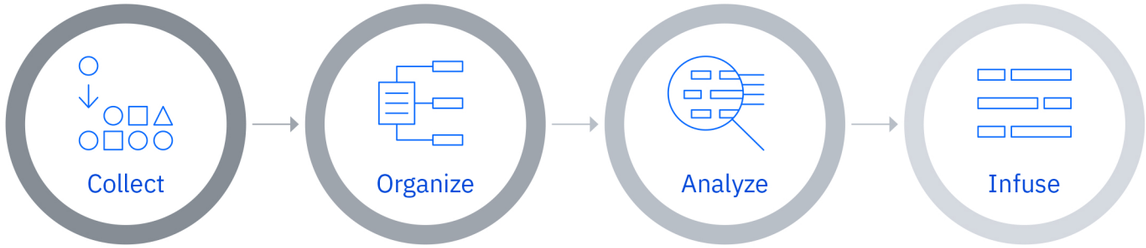
Organizations today agree that artificial intelligence (AI) is the fast track to innovation and productivity. Most organizations are already on their way to testing, adopting, implementing, and realizing the full potential of AI and, as a result, corporate investment in AI solutions is expected to increase significantly over the next several years.

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

40% of digital transformations will use AI services by 2019[1]

$4.79 billion

in IT storage spend for AI in 2019[1]



Make data simple and accessible

Create a trusted analytics foundation

Scale insights with Machine Learning

Predict, automate and scale across processes

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 challenges[1] when supporting AI development teams or deploying and scaling production AI workloads:

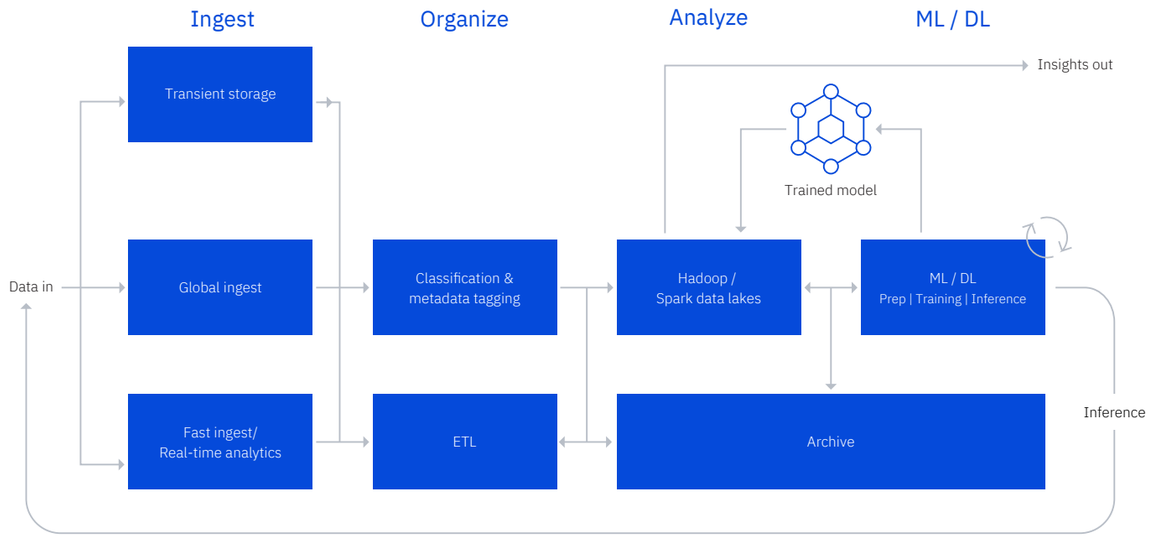
- **Data volume and quality.** AI requires high-quality, diverse, and labeled data inputs. Identifying the right data sets across multiple data sources with dynamic data characteristics can be daunting.
- **Advanced data management.** Organizing and tracking data sets in AI projects is a challenge for 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 pipeline — how you ingest, organize and analyze data and, ultimately, train models to create AI-driven insights from that data — is essential to efficient data science. Efficiency of your AI pipeline is directly tied to addressing the challenges above with the right IT infrastructure.



Unified data access

Data silos are a major obstacle to the productive use of data, particularly when it comes to AI. Collecting data can be the most time-consuming phase of an AI project. The skills investment in data set organization and classification should be leveraged across all AI projects. It requires a data and storage architecture that minimizes redundancy, improves efficiency and enables common, shared data for multiple projects and supporting the range of data analytics tools.

Data throughput performance

AI model accuracy is a function of good data input and sufficient compute resources to analyze it. Graphics Processing Units (GPUs) are often used for AI because they analyze large data sets quickly. The IT infrastructure must be paired with storage performance to match the compute resources' ability to consume data. Similarly, streaming data may be used for real-time insights requiring attributes that properly distribute data workloads.

Agility with container support

AI projects are typically managed in containers because they are lightweight, quickly deployed, and can combine multiple programs and scripts. To quickly scale from initial experiments to production-grade AI, persistent storage that works with Kubernetes and Red Hat OpenShift is required. Containers not only simplify development, but also add agility to the IT infrastructure to accommodate growth in the demand for enterprise AI services.

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 enables you to securely collect and organize data, providing data-anywhere access with a unified data foundation that simplifies AI adoption.

Learn about IBM Spectrum Scale

IBM Cloud™ Object Storage

IBM Cloud Object Storage delivers performance and scalability for cloud native applications and AI frameworks. It is a secure, software-defined storage platform that easily scales capacity and throughput from terabytes to exabytes. IBM Cloud Object Storage is the ideal solution for teams using the latest cloud development environments that also need data security or high-performance local data.

Learn about IBM Cloud Object Storage

IBM Spectrum Discover

IBM Spectrum Discover is 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 scientists and storage teams to efficiently unify, catalog, and enrich metadata to increase insights from their growing, diverse stores of unstructured data.

Learn about IBM Spectrum Discover

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 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 in 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 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 analysis and long-term 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

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 [COMPARE] 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 can quickly grow across the organization. Navigating that journey successfully starts with creating a robust, agile IT foundation optimized for the unique data requirements that drive productivity and adoption. The right storage platform must deliver 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 the data pipeline 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 insights that create better business outcomes.

Learn how IBM Storage for AI can put your organization on the fast track from ingest to insights.

Discover IBM Storage for AI

Resources

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