



# 7 Steps to Unlocking the Value of Data for AI

A Comprehensive Guide to Effective Data Management



# Building a Competitive Edge with Data

In today's rapidly evolving business landscape, data is a critical resource—the cornerstone of innovation and competitive advantage. Leaders who accelerate their business decisions through smart use of robust, accurate, and comprehensive data sets will set themselves apart. Nowhere is this more evident than in the adoption of [Generative AI \(GenAI\)](#).

GenAI promises unparalleled advancements and efficiencies, and perfect data is certainly not a prerequisite for the creative, iterative, and flexible outputs it can offer. But, at some point, you will have to assess your “data house” to ensure that it's in order. In other

words, you'll need to have more than just **access** to data; you'll need **strong data management practices**. This will become even more necessary as you scale.

**Strong data management is necessary, but it's not easy.**

Deploying [AI workloads](#) is the culmination of a complex and demanding process whereby data scientists and others identify appropriate datasets and ensure their cleanliness and comprehensiveness. This endeavor extends beyond technical challenges; it requires organization-wide processes and robust data management frameworks.



<sup>1</sup> Forbes Advisor: [How Businesses Are Using Artificial Intelligence in 2024](#)

<sup>2</sup> McKinsey Report: [The state of AI in early 2024](#)

<sup>3</sup> Fortune: [Want your company's AI project to succeed? Don't hand it to the data scientists, says this CEO.](#)



# The data management journey

To help navigate this intricate landscape, we have developed a comprehensive overview of the data management journey, distilled into seven essential steps. The culmination of extensive workshops and consultations conducted by Dell's expert data scientists with a diverse array of organizations, these steps present common challenges and successful strategies for creating scalable and effective [AI](#) models. In coming months, we will dive more deeply into each step so that you can overcome your organization's particular roadblocks and see your AI journey through.

This ebook offers a clear, actionable framework for understanding and implementing effective data management practices. By embracing these principles, your organization can go from experimenting with GenAI to scaling the transformation of data from a raw resource into a strategic asset, paving the way for innovation and sustained competitive advantage in the AI-driven future.

Download this [infographic](#) for an easy reference to these 7 steps.



## STEP 1

Identify the Business Need



## STEP 2

Accelerate Relevant Data Discovery



## STEP 3

Simplify Data Exploration and Access



## STEP 4

Optimize Analytics, ML Experimentation, and Modeling



## STEP 5

Scale Data and Analytics Productization



## STEP 6

Automate Data Management and Governance



## STEP 7

Evaluate Business Outcomes



## STEP 1

# Identify the Business Need



This step sets the direction for your data management journey, aligning efforts with strategic business needs. Without defined goals and metrics, achieving any meaningful value is unlikely.

- Begin by understanding the operational objective and the value it will unlock.
- Align across the organization on the desired outcomes and how success will be measured.
- Establish a clear vision of the value that will be created.

By knowing **what** you aim to achieve and **how** to measure it, you ensure that all subsequent data management efforts are purposeful and directed toward clear, achievable objectives.

## STEP 2

# Accelerate Relevant Data Discovery



With a clear roadmap, accelerate the discovery of relevant data. Not all available data is necessary; data scientists must quickly identify data pertinent to solving the problem. Establish a clear connection between data and its value by cataloging source datasets and creating metadata. This focused approach ensures that your data efforts are efficient and effective.

By pinpointing relevant data swiftly, you save **time** and **resources**, allowing your data scientists to concentrate on high-impact data that directly supports your strategic goals.

**68%** of data available to enterprises is left untapped<sup>4</sup>

<sup>4</sup> Seagate Technology: [Rethink Data: Put More of Your Data to Work—From Edge to Cloud](#).



## STEP 3

## Simplify Data Exploration and Access



If you're moving data to centralized locations for analysis, you're reducing the value of real-time use cases. Avoid this inefficiency by enabling data scientists to access data where it resides. Standardize processes and implement auto-discovery to arrange structured and unstructured data into easily accessible locations, facilitating real-time discovery and utilization. This approach not only saves time but also enhances the agility of your data operations, allowing for **quicker insights** and **more responsive decision-making**.

With explosive growth of data, effective data management is necessary to successfully scale [AI workloads](#). Consider a tool like the [Dell Data Lakehouse](#) that provides a full solution stack for discovering, querying and processing all of your enterprise data, regardless of location or data source, with a single federated query engine.



<sup>5</sup> Seagate Technology: [Rethink Data: Put More of Your Data to Work—From Edge to Cloud](#).



## STEP 4

# Optimize Analytics, ML Experimentation, and Modeling



Encourage constant experimentation and modeling to identify the variables that can solve problems. Consider synthetic data creation, which can be especially helpful when facing data quality and privacy challenges, and can help to expedite the process, especially when beginning your AI journey. Leveraging pre-trained foundational models that only require augmentation and fine-tuning can be a great starting point.

Once scaling your project up, focus on easy access to data. This enables your team to optimize analytics through iterative testing. A platform that supports multiple iterations and algorithms can uncover key data variables, enhancing the effectiveness of generative AI and unlocking data value. This iterative process of testing, learning, and refining ensures that your **models are robust** and your **insights are actionable**, driving continuous improvement and innovation.

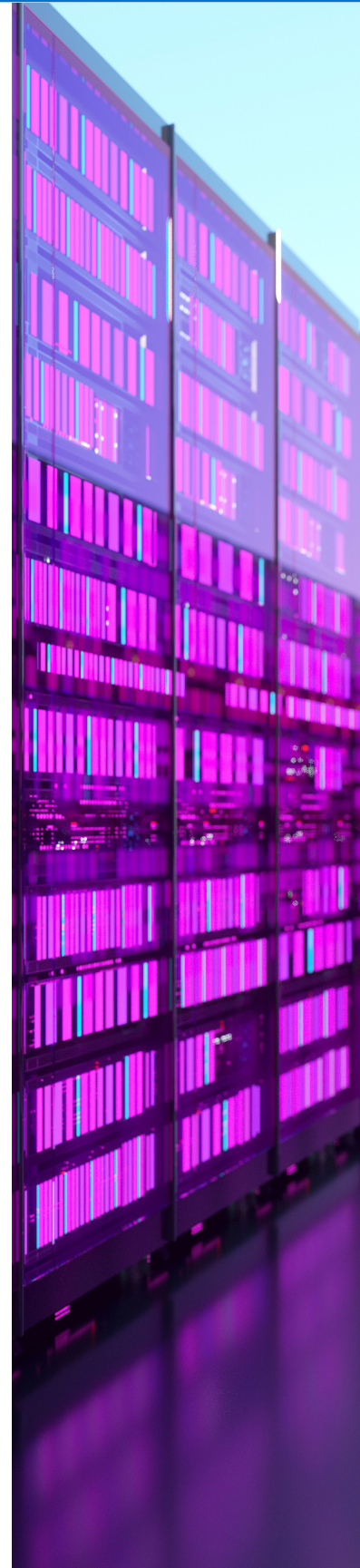


## STEP 5

# Scale Data and Analytics Productization



Transition from a data science **project** to a reliable, repeatable data science **product**. These products can operate independently and be periodically reviewed for improvements. An efficient data intake process accelerates time-to-insights, enabling data pipelines with built-in intelligence to analyze and use data in real-time, reaching predefined goals. By scaling data products, you ensure consistency and reliability, transforming **ad-hoc projects** into **sustainable, value-generating assets**.

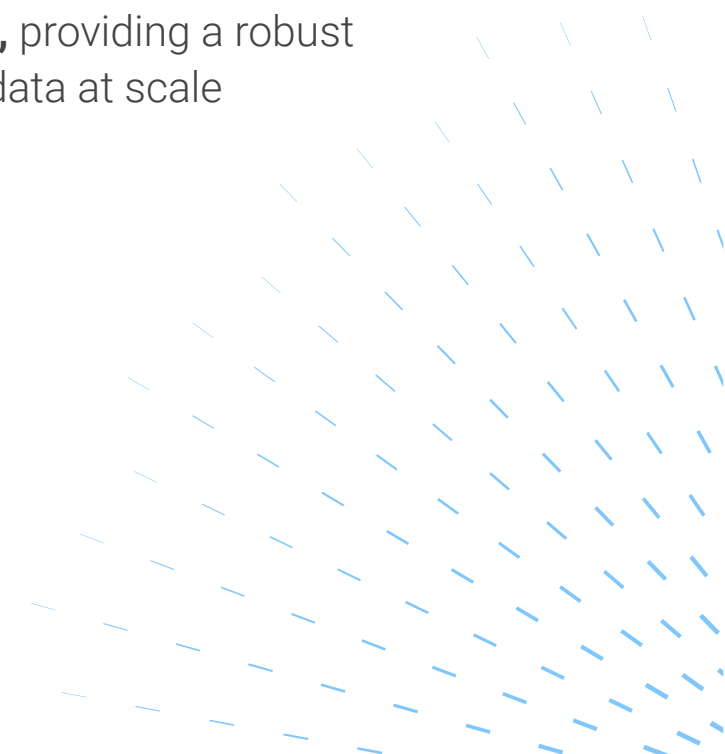


## STEP 6

# Automate Data Management and Governance



As products become repeatable, automate your organization's data management and governance processes. This allows the system to self-monitor and flag anomalies before they become problems. Embrace holistic observability to ensure workflows adhere to integrated data governance standards, policies, and security measures throughout the data management journey. Automation not only **reduces manual effort** but also **enhances accuracy and compliance**, providing a robust framework for managing data at scale





## STEP 7

## Evaluate Business Outcomes



Evaluate data from business outcomes and feed it back into the process, fostering a culture of continuous learning and adaptation. This ongoing evaluation captures actual results versus predicted ones, enabling refinements that improve outcomes and optimize results. Leveraging capabilities with minimal human intervention ensures sustained performance and value. By continuously analyzing outcomes, you create a feedback loop that drives **perpetual improvement** and **aligns your data initiatives** with evolving business objectives.







# Putting data to work with the Dell AI Factory with NVIDIA

Your current data center and IT operating model probably are not equipped to harness the speed and scale of GenAI. What you need is a new type of data center, purpose-built to meet the specific demands of AI.

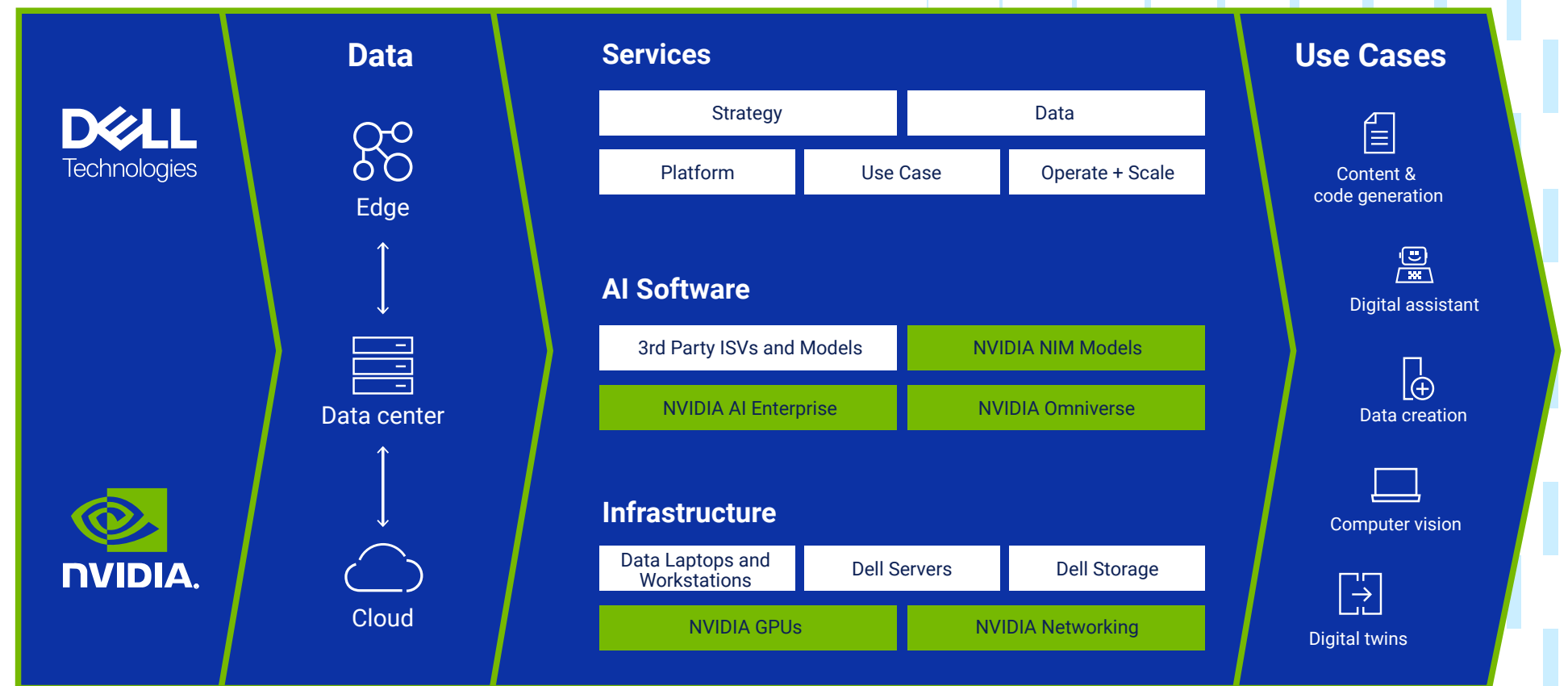
## You need the AI factory.

Just as physical factories fueled the industrial revolution, AI factories power the AI revolution. But instead of physical goods, AI factories produce intelligence. They are versatile and flexible, and operate anywhere your data resides—across clouds, within data centers, on workstations, and at edge locations.

2.4x

annual spending growth on training large-scale ML models—mostly in hardware like GPUs and servers<sup>6</sup>

<sup>6</sup> Epoch AI: [How Much Does it Cost to Train Frontier AI Models?](#)





Understanding your use cases and applying the right data management solutions is the foundation of optimizing your AI efforts.

The **Dell AI Factory with NVIDIA** brings together necessary building blocks to help you accelerate your AI innovation:

- a focus on data
- AI-optimized infrastructure
- AI software and models
- open ecosystems
- expert services
- and best practices

It's technology informed by your business requirements to deliver the right solution, fast, with a better total cost of ownership. Dell and NVIDIA have spent more than 340,000 hours validating these solutions and providing reference designs and benchmarking.

Together, NVIDIA and Dell deliver a full-stack solution that accelerates your AI adoption: the **Dell AI Factory with NVIDIA**.

With NVIDIA AI Enterprise, you can:



### **Access, manage, deploy and support**

GenAI models, applications, solutions, and accelerators across a range of infrastructure solutions.



### **Leverage NVIDIA NIMS**

to enable a robust data pipeline to fully take advantage of all of that data pipeline work



### **Easily transition**

from one model to the next while maintaining the sanctity of your data pipelines, providing investment protection and enabling you to always embrace the best available models



# Sustaining Success with Ongoing Data Management

The data management journey is not a one-time task. It's an ongoing process, utilizing a set of practices and tools that will help you continuously unlock value from your data. And it's critical for any successful GenAI initiative. Dell, in collaboration with NVIDIA, offers the world's broadest AI solution portfolio.



## End-to-end Solutions

From desktop to data center to cloud, our end-to-end solutions, powered by NVIDIA AI, provide security, accessibility, and scalability.



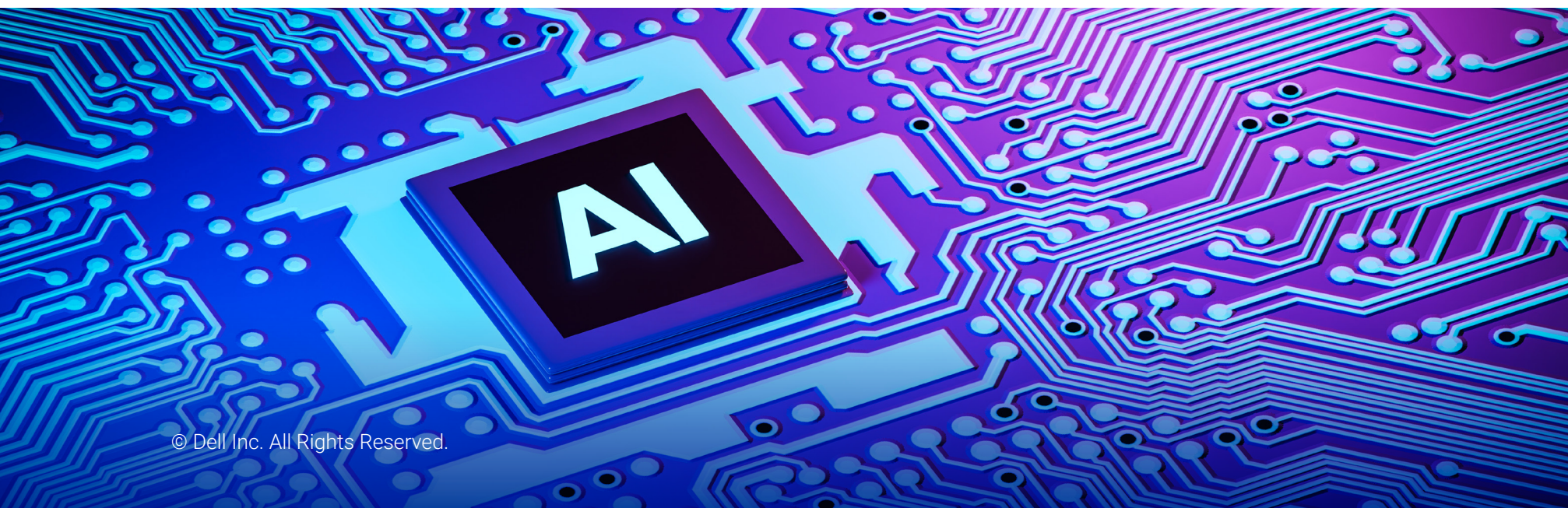
## Committed to AI

The Dell AI Factory with NVIDIA embodies Dell's commitment to embracing and implementing AI, helping customers accelerate their AI initiatives to maximize their data and achieve their goals.



## Simplified Data Navigation

With Dell and NVIDIA, you can navigate the data management journey with ease. And in the realm of AI-driven data workloads, the journey is just as important as the destination.





## Data Management is not a one-time task.

The data management journey is an ongoing process, utilizing a set of practices and tools that will help you continuously unlock value from your data. And it's critical for any successful GenAI initiative.

## There is no “one size fits all” approach.

You have a unique data foundation—your own specific ecosystem of hardware and software and data location and types, and your own distinct use cases. The steps outlined in this ebook serve as an overarching guide to help navigate your organization's specific needs.

## And you are not on your own.

[Dell Services](#) offers deep expertise at every stage of the lifecycle to accelerate tangible time to value, tailored to your organization. From aligning a winning strategy and validating data to quickly implementing your data platform and ensuring secure, optimized operations, our trusted experts are here to assist you.

No matter where you are in your journey, let us help you leverage enterprise data effectively to power AI projects and accelerate time to value.

[VISIT DELL.COM](#)



## Glossary of Terms

---

### Generative AI

Generative AI refers to a subset of artificial intelligence that focuses on creating new content or data. This type of AI uses algorithms and models, such as Generative Adversarial Networks (GANs) and transformers, to generate text, images, music, and other types of media. Generative AI is unique in its ability to produce original outputs that mimic the style and content of its training data, making it highly valuable for creative tasks, content creation, and simulating complex scenarios.

### AI

Artificial Intelligence (AI) is the broad field of computer science focused on creating systems capable of performing tasks that typically require human intelligence. These tasks include learning, reasoning, problem-solving, perception, language understanding, and decision-making. AI encompasses various techniques and approaches, such as machine learning, natural language processing, and robotics. It aims to create intelligent agents that can adapt to new situations, improve performance over time, and assist or automate human activities across a wide range of applications.

### AI Workloads

AI workloads refer to the specific tasks or processes that are handled by AI systems. These workloads can vary widely depending on the application and may include data processing, model training, inference, and analysis. AI workloads are characterized by the intensive computational demands required to train models, process large datasets, and perform real-time decision-making. Managing AI workloads involves optimizing hardware, software, and algorithms to efficiently handle these demanding tasks, ensuring that AI systems perform effectively and at scale.

## References and Further Reading

---

[Infographic: The data management journey](#)

[Dell Professional Services for Generative AI](#)

[Dell Data Management Solutions](#)

[The Dell AI Factory with NVIDIA](#)

[Top 5 Security Considerations for Gen AI](#)