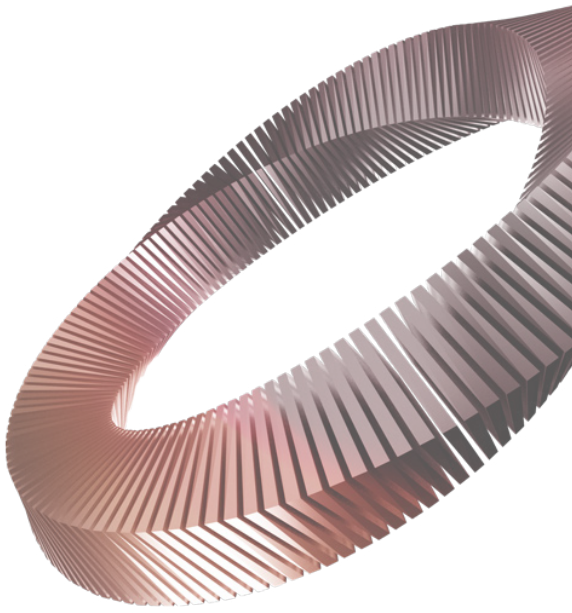
# The AI Data Factory

## Accelerating the Process from Ingest to Monetization



# Business Value of Information Assets Has Always Been About Monetization

Only a couple of decades ago, the process of finding obscure or unrealized profits required painstaking analysis of financial data across numerous spreadsheets and individual worksheets. A decade ago the techniques advanced through data mining and big data analytics. Today it is increasingly driven by – and is expected to soon be dominated by – ML and AI technologies, as explored in The Enlightened Leaders Guide to AI and Finance.

The process of extracting value from data so that it can be monetized by a company to improve its financial position is not new. But it has evolved, in terms of sophistication, scope and importance to the business. In the past, organizations were much more focused on using internal data – usually financials – to improve operating efficiencies and reduce expenses. And much of the analysis and monetization – in the form of cost-savings – had been confined to internal IT or business operations.

In the past several years, industry experts have been predicting and advocating for the application of these techniques. With broader analytics, the lens on monetization has become increasingly bidirectional, enabling organizations to go beyond the traditional goals of enhancing internal efficiency and cost savings. Data types selected for analysis have become more varied, market and customer information is more readily available, and analysis has become faster and more sophisticated.

Consequently, the scope of internal (or indirect) monetization has expanded outward, toward new sales and revenue opportunities and a proactive approach that seeks to take advantage of changing market conditions instead of being fearful of them.

**Data monetization is now led by market-driven business priorities and outcomes, expressed as measurable results that affect the bottom line.**

In addition to focusing on more targeted customer marketing and better customer experiences, organizations are now aligning with the industry experts had been recommending. They're looking to external data monetization, realized by sharing data or selling it directly to customers, partners or third parties. As IDC puts it: "What we see now is that enterprises that have successfully gained value from their internal data are beginning to provide their own data as a service to others who may be connected in the value chain – or in entirely different commercial spheres."

The information sold as a service includes raw data, derived metrics, insights, and recommendations, all of which will certainly become more valuable and actionable with advancements in AI models.

# Maturing Plans and Promising Tools Create a Period of Opportunity with Data as Hyper-Strategic Mechanism

As impressive as today's AI is, the technology is still in a nascent stage. There is, therefore, considerable upside and untapped potential for data monetization as AI-driven analysis enables faster, deeper and more valuable insights. Moreover, the evolution of AI-driven data monetization can be considered the next stage in the transition of IT from cost-center to the hub of strategic innovation and competitive advantage.

Therefore, the data management technologies that enterprises choose will likely have a more direct impact on business results.

Not surprisingly, data is the critical factor that makes monetization happen internally and externally for organizations in all industries. It is estimated that 84% of the market value of S&P 500 companies already comes from intangible assets, including data and software. Obviously this includes intellectual property, but it also increasingly includes data-driven insights to understand the status and health of the business as well as analysis to operate and grow the business efficiently.

Data and analytics are increasingly important factors in determining an organization's financial performance against its industry peers. And in order to survive, let alone thrive, businesses also need a disciplined approach toward data management and processing to meet the demands for data monetization.

The challenge, however, for nearly every midsize to large enterprise is that there's too much data to manage. It's coming in too fast, from all directions, with patterns that change very quickly. Most organizations simply cannot efficiently manage the volume and variety of data for effective analysis and monetization using existing resources.

# AI Expands the Playing Field for Enterprise Data Analytics & Decision-Making

AI is the game-changer that will both improve and determine how well organizations use, manage and capitalize on the value of data. Industry opinions also seem to be converging:

> [ QUOTE ]
>
> 84% of executives believe they won't achieve their growth objectives unless they scale AI...
>
> Scaling AI: Giving Data Its Due
>
> **Data Its Due**

> [ QUOTE ]
>
> When asked to select the top initiative for increasing their competitive advantage from AI, respondents singled out "modernizing our data infrastructure for AI.
>
> Machine Data Revolution: Feeding the Machine
>
> **Data Its Due**

> [ QUOTE ]
>
> At every level of your business, investing in data and AI should be the highest priority to improve net margin, free up working capital, improve customer satisfaction, anticipate changing demand, [and] manage costs…
>
> New Era of Value Realization is Here – Put Your Data to Work with AI
>
> **Microsoft**

Until recently, only massive technology companies that considered data to be part of their core value (think Apple, Amazon, Google) were able to collect large volumes of data and invest in AI resources. But we're in a new AI reality today, as noted by the Harvard Business Review in 4 Steps to Start Monetizing Your Company's Data: "The days of AI and ML being a luxury only accessible to major tech companies are over. But companies need to learn how to use [these powerful new tools] strategically — and how to think about the data that powers them. Learning to do that is where you'll really find the competitive advantage of AI."

Companies inside and outside of tech can now leverage AI tools to create greater opportunities for faster, more sophisticated insights, analysis and recommendations. The improved speed and accuracy of ML training has created opportunities to attract more customers and increase market share. In fact AI-based analysis is going to be used by ALL organizations that want to monetize their data.

But it's not enough for organizations to have access to AI- and ML-based apps and tools. They need an intelligent, systematic data-centric approach and AI infrastructure that can deliver the speed and scale across sequential data management stages to handle today's AI-based processing. A technology solution that supports end-to-end AI data lifecycle management is critical to fast, intelligence-driven data monetization.

# What's Holding Back Monetization for Most Organizations?

While more that 50% of organizations are using AI in some functional area of the business, very few have a systematic process in place that strategically optimizes AI data.

For example, despite their stated interest in data monetization, 72% of corporations have no data and analytics strategy. And only 14% of companies who have stalled on their AI journey think data and analytics are vital to successful AI adoption.

Yet interestingly (and perhaps strangely), 53% of senior executives identify data and analytics as their top investment priority over the next two years and 75% of executives believe they risk going out of business in 5 years if they don't scale AI operations.

The collective and somewhat confusing market data seems to suggest that most enterprises agree that AI and analytics are important for data monetization, but many are either unable to understand the importance of an AI data management and analytics strategy or are unable to implement one.

It's also possible that both shortcomings are connected; i.e. the inability to formulate a strategic approach to AI data management is due to technology limitations that restrict an organization's ability to take advantage of AI-driven data monetization.

The very real technology challenge today is that IT systems and enterprise storage solutions still struggle to provide the necessary power and scale for AI data processing. In addition, most enterprise storage solutions aren't designed for a data-centric future.

They cannot support all the different data and file types or integrate with all sources of data. Data is frequently isolated in islands and the ongoing proliferation of data sources also makes it harder to discover what information could be most helpful.

All of these factors complicate the data ingest process.

Furthermore, an increasing amount of data for ML and analytics is coming from IoT, social media, video and audio files, etc. This data is unstructured and doesn't fit neatly into database tables, rows and columns. Yet all of this disparate data needs to be captured, because some of it is likely to be valuable – if not today, then sometime in the future.

As Deloitte pointed out in Machine Data Revolution: Feeding the Machine:

> [ QUOTE ]
>
> The ability to process larger volumes of diverse data in real time is the secret sauce of ML-based data decisioning. The faster that big data systems can capture and process data, feed it into ML and analytics platforms, and then serve up insights to users, the more impactful your data investments can be.

**Storage & Data Management Limitations that Can Hinder Monetization Include:**

- Inadequate performance and scale required for AI and analytics. This slows down several processes including data ingest, AI training, user jobs and time-to-production
- Inability to guarantee secure data governance, sharing and management
- Inflexible solutions that can't accommodate future enhancements (e.g., to support different data types and storage or network protocols) for long-term growth
- Delayed or incomplete analytic results which fail to keep data scientists productive

# How an AI Factory Connects the Dots Across the End-to-End AI Data Lifecycle

The following is a summary of salient points from above. It serves as both a recap and segue to the next topic, which describes an effective approach to AI-based data monetization:

- Inadequate performance and scale required for AI and analytics. This slows down several processes including data ingest, AI training, user jobs and time-to-production
- Inability to guarantee secure data governance, sharing and management
- Inflexible solutions that can't accommodate future enhancements (e.g., to support different data types and storage or network protocols) for long-term growth
- Delayed or incomplete analytic results which fail to keep data scientists productive

The data-centric approach toward AI data requires more than just technology. As stated in a recent Harvard Business Review article, "Getting the right [AI] tools isn't enough. To gain long-term profits and a competitive edge, companies need to optimize their data — and learn to use it strategically. Their leaders should prioritize developing data processes as a core component of the business."

And as noted by Deloitte: "..ML algorithms will deliver little ROI in companies with outdated data infrastructure and processes..."

What is required is an AI and data platform; one that can operationalize the end-to-end AI data lifecycle process to achieve AI at scale for future-resistant performance and capacity.

According to Accenture, "Data is the new capital. The most important investment a company should consider making is in building a Data and Analytics Platform. The platform should be foundational and enterprise-wide to allow for interoperability of cross-functional data to maximize power.

These platforms can create analytics reports, train artificial intelligence or machine learning models to hone predictive capabilities or be used for operational applications and can convert data for easy consumption and enable agile data processing. The platforms are built for speed and adaptability, enabling the processing of complex, data-driven insights in real-time. The platform is comprised of numerous capabilities to enable the entire data and analytics ecosystem."

DDN likens this processing of AI data to a factory that starts with various data as its raw materials and delivers services as products that elevate the value of data and facilitate its internal and external monetization.

In this AI lifecycle model, as shown in Figure 1 below, data feeds the multi-stage AI process, powered by the DDN Data and AI Platform, to produce multiple AI models and services. The contributors to – and stakeholders for – the AI Factory include IT and support staff, data engineers, data scientists, data journalists, business analysts, marketing, operations and sales.

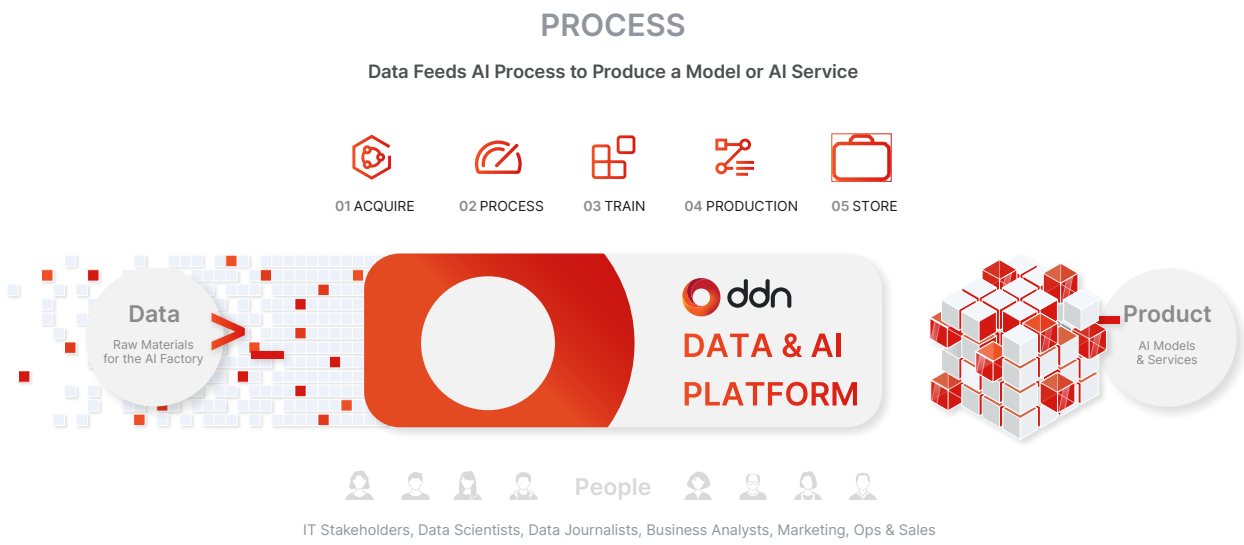



Figure 1: DDN Platform as AI Data Factory

The table below highlights the challenges that hinder the implementation of an end-to-end AI data and monetization lifecycle process. It also highlights the operational benefits associated with the successful implementation of a DDN AI Factory solution: simplicity, speed, scale and intelligent automation for AI data and processes at each lifecycle stage.

| | 01 ACQUIRE | 02 PROCESS | 03 TRAIN | 04 PRODUCTION | 05 STORE |
|---|---|---|---|---|---|
| **Challenge** | Novel data sources and finer granularity increase strain on storage ingest capabilities | Mixed file types and multiple access methods aren't accounted for in current design. | Slow and unreliable systems that delay results or keep them from being delivered altogether. | The move to production is blocked because of scaling issues. | Current storage design isn't flexible enough to meet future capacity requirements |
| **Solution** | Ingest 10s of millions of docs and audio and video files in real time rom dozens of sources. High write performance enables real-time AI even with high volume bursts. | Quickly label, clean and transform incoming media files. 10x more IOPs from flash significantly speeds up pre-processing. | Specific DDN optimizations for GPUs and AI frameworks time to production to a accelerate deep learning workloads so you can train across all your data. | Built-in automation and DDN's AI team reduce few hours and provides expert help when you need it. | Current storage design isn't flexible enough to meet future capacity requirements |

Table 1: The AI Data Factory: Implementing an End-to-End AI Data Lifecycle

# Results: Operationalizing the AI Data Lifecycle Provides a Faster Route to Internal & External Data Monetization

Even sophisticated enterprise IT organizations can benefit from a trusted partner

with proven AI data execution experience at scale. As a global leader in at-scale data management for over two decades, DDN has helped thousands of customers successfully maximize the business value of their data. DDN delivers the industry's highest performance storage solutions for faster and more reliable AI data-driven insight and action.

DDN A$^3$I is the first storage solution to be fully validated in production with NVIDIA DGX™ BasePOD and DGX SuperPOD™ at multiple customers worldwide at any scale.

By implementing the AI Factory model, DDN customers are experiencing results that directly enhance data monetization efforts across the following areas:

### 01 SIMPLICITY

Unified data and simplified access from dozens of sources. Organizations can build their data store and control their AI models and services – all from a single silo-free environment.

### 02 SPEED

Accelerated deep learning with 100x GPU rates. DDN systems ensure that AI-driven analysis, insights and recommendations are generated with maximum velocity, so data scientists are always fully engaged and productive.

### 03 RELIABILITY

Stronger service levels, data accessibility and security for better data-availability, proactive resources and performance optimization.

### 04 SCALE

Ability to grow to supercomputer class without unexpected surprises and bottlenecks.

### 05 GOVERNANCE

Secure data sharing and management, enabling organizations to confidently audit and govern the data that is ingested into their AI model(s).