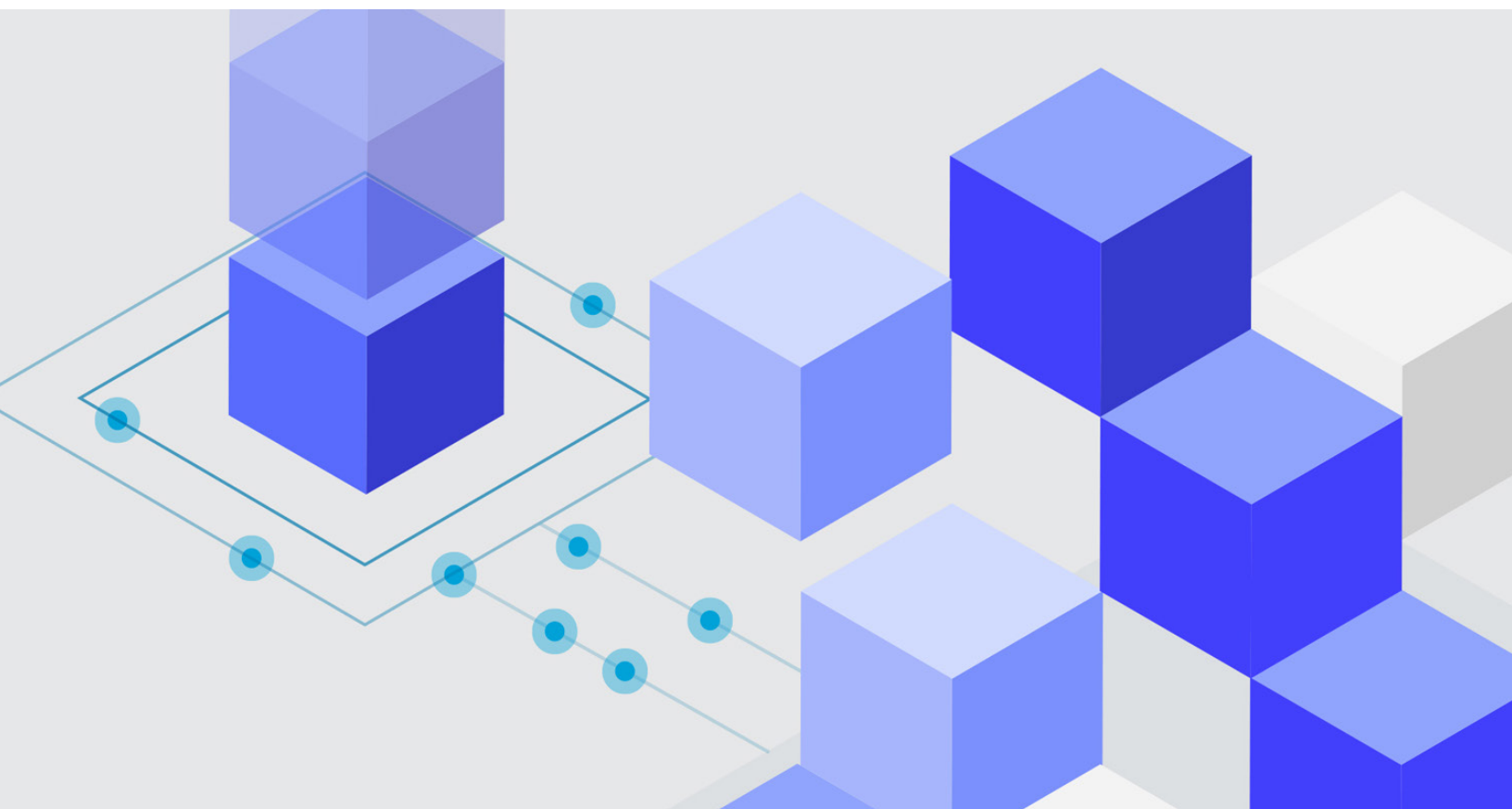


The data-driven enterprise of 2025

Rapidly accelerating technology advances, the recognized value of data, and increasing data literacy are changing what it means to be “data driven.”



By 2025, smart workflows and seamless interactions among humans and machines will likely be as standard as the corporate balance sheet, and most employees will use data to optimize nearly every aspect of their work.

We know 2025 isn't too far off, but that's the point.

Seven characteristics will define this new data-driven enterprise, and we've already seen many companies exhibit at least some of them, with many more beginning the journey to do so.

Those able to make the most progress fastest stand to capture the highest value from data-supported capabilities. Companies already seeing 20 percent of their earnings before interest and taxes (EBIT) contributed by artificial intelligence (AI), for example, are far more likely to engage in data practices that underpin these characteristics.¹

This guide is intended to help executives understand the characteristics of the new data-driven enterprise and the capabilities they enable. It also provides resources to dive deeper on how to embed them in your organization.

The following are the seven characteristics of the data-driven enterprise:

1. Data is embedded in every decision, interaction, and process.
2. Data is processed and delivered in real time.
3. Flexible data stores enable integrated, ready-to-use data.
4. Data operating model treats data like a product.
5. The chief data officer's role is expanded to generate value.
6. Data-ecosystem memberships are the norm.
7. Data management is prioritized and automated for privacy, security, and resiliency.

Seven characteristics will define the new data-driven enterprise. Those companies able to make the most progress fastest stand to capture the highest value from data-supported capabilities.

¹ "The state of AI in 2021," McKinsey, December 8, 2021.

1. Data is embedded in every decision, interaction, and process

Today

Organizations often apply data-driven approaches—from predictive systems to AI-driven automation—sporadically throughout the organization, leaving value on the table and creating inefficiencies.

Many business problems still get solved through traditional approaches and take months or years to resolve.

By 2025

Nearly all employees naturally and regularly leverage data to support their work. Rather than defaulting to solving problems by developing lengthy—sometimes multiyear—road maps, they are empowered to ask how innovative data techniques could resolve challenges in hours, days, or weeks.

Organizations are capable of better decision making as well as automating basic day-to-day activities and regularly occurring decisions. Employees are free to focus on more “human” domains, such as innovation, collaboration, and communication. The data-driven culture fosters continuous performance improvement to create truly differentiated customer and employee experiences and to enable the growth of sophisticated new applications that aren’t widely available today.

Everyday applications²

- *Store managers* provide a differentiated shopping experience using real-time analytics to identify and direct loyalty-program customers to products they find interesting as they shop, and streamline or completely automate the checkout process.
- *Network operations* staff at telecommunications companies leverage autonomous networks that automatically identify areas requiring

maintenance and highlight opportunities for building out the network based on usage.

- *Procurement managers* regularly apply data-driven processes to instantly triage purchases for approval so they can focus on building out a more effective partner strategy.

Key enablers

- *a vision and data strategy* to highlight and prioritize transformational use cases for data
- *technology enablers* for sophisticated AI use cases, such as a cloud-based infrastructure; architectures that support real-time analytics; and flexible database/data-model tooling to support querying of unstructured data
- broad organizational *data literacy and a data-driven culture*, where all employees know and embrace the value of data

How to get started

- *Read* “Winning with AI is a state of mind” for more about making the shift to an AI-enabled organization, and learn how to harness the power of data from AI leaders.³
- *Begin upskilling your employees* for data use and AI, if you haven’t started already. Analytics academies can help.⁴
- *Learn* how to reimagine each workflow, journey, and function to leverage data and AI in “Getting AI to scale.”⁵
- *Articulate your vision* for a data-driven organization.

² Examples; not exhaustive.

³ See Thomas Meakin, Jeremy Palmer, Valentina Sartori, and Jamie Vickers, “Winning with AI is a state of mind,” McKinsey, April 30, 2021; and Mohammed Aaser, Jonathan Woetzel, and Kevin Russell, “Five insights about harnessing data and AI from leaders at the frontier,” McKinsey Global Institute, March 25, 2021.

⁴ Solly Brown, Darshit Gandhi, Louise Herring, and Ankur Puri, “The analytics academy: Bridging the gap between human and artificial intelligence,” *McKinsey Quarterly*, September 25, 2019.

⁵ Tim Fountaine, Brian McCarthy, and Tamim Saleh, “Getting AI to scale,” *Harvard Business Review*, May–June 2021, Volume 99, Number 3, pp 116–23.

2. Data is processed and delivered in real time

Today

Only a fraction of data from connected devices is ingested, processed, queried, and analyzed in real time because of the limits of legacy technology structures, the challenges of adopting more modern architectural elements, and the high computational demands of intensive, real-time processing jobs. Companies often must choose between speed and computational intensity, which can delay more sophisticated analyses and inhibit the implementation of real-time use cases.

By 2025

Vast networks of connected devices gather and transmit data and insights, often in real time. How data is generated, processed, analyzed, and visualized for end users is dramatically transformed by new and more ubiquitous technologies, such as kappa or lambda architectures for real-time analysis, leading to faster and more powerful insights. Even the most sophisticated advanced analytics are reasonably available to all organizations as the cost of cloud computing continues to decline and more powerful “in-memory” data tools come online (for example, Redis, Memcached). Altogether, this enables many more advanced use cases for delivering insights to customers, employees, and partners.

Everyday applications

- *Maintenance teams for physical assets*, such as those in factories, regularly leverage networks of connected sensors to detect maintenance needs in real time.
- *Product developers* use unstructured data and unleash unsupervised machine-learning algorithms on web data to detect deeply hidden patterns and develop a much richer understanding of customers than is possible today (for

example, by using internet-protocol data and website behavior to personalize web experiences for specific customers in real time).

- *Financial analysts* use alternative visualization tools, potentially leveraging augmented reality/virtual reality (AR/VR) to visualize analytics for strategic decisions involving multiple variables rather than being limited to the typical two-dimensional dashboards common today.

Key enablers

- *a view of the full business architecture* to understand integration across assets, processes, insights, and interventions and to enable the identification of real-time opportunities
- *more powerful edge-computing devices* (IoT sensors, for example), so that even the most basic devices generate and analyze usable data “at the source”
- *advanced-connectivity infrastructures*, such as 5G, to support high-bandwidth, low-latency data from connected devices
- *in-memory computing* for faster and more effective computations for intensive analytics jobs

How to get started

- *Take advantage of a road-tested reference data architecture* that enables the modularity, flexibility, and scalability needed to support these capabilities.⁶
- *Evolve to a cloud-enabled data platform* to meet future data and analytical needs, such as real-time capabilities.⁷
- *Learn about* the future of cellular-enabled computing devices.⁸

⁶ Sven Blumberg, Jorge Machado, Henning Soller, and Asin Tavakoli, “Breaking through data-architecture gridlock to scale AI,” McKinsey, January 26, 2021.

⁷ “Three actions CEOs can take to get value from cloud computing,” *McKinsey Quarterly*, July 21, 2020.

⁸ Ferry Grijpink, Kasia Jodlowska, Mark Patel, and Rutger Vrijen, “Reliably connecting the workforce of the future (which is now),” McKinsey, April 14, 2021.

3. Flexible data stores enable integrated, ready-to-use data

Today

Though the proliferation of data is driven by unstructured or semistructured data, most usable data is still organized in a structured fashion using relational database tools. Data engineers often spend significant time manually exploring data sets, establishing relationships among them, and joining them together. They also frequently must refine data from its natural, unstructured state into a structured form using manual and bespoke processes that are time-consuming, not scalable, and error prone.

By 2025

Data practitioners increasingly leverage an array of database types—including time-series databases, graph databases, and NoSQL databases—enabling more flexible ways of organizing data. This allows teams to query and understand relationships between unstructured and semistructured data easier and faster, which accelerates development of new AI-driven capabilities and the discovery of new relationships in the data to drive innovation. Combining these flexible data stores with advances in real-time technology and architecture also enables organizations to develop data products, such as “customer 360” data platforms and digital twins—real-time-enabled data models of physical entities (such as a manufacturing facility, supply, or even the human body). This enables sophisticated simulations and what-if scenarios using traditional machine-learning capabilities or more advanced techniques such as reinforcement learning.

Everyday applications

- *Financial institutions* regularly use graph-database technology and a common data model to stream and integrate customer data from multiple sources (marketing systems, enterprise-resource-planning systems, web data) into a single, unified, 360-degree view of the customer that can be modeled in real time.

- *Transportation and logistics companies* use real-time location data and sensors embedded into vehicles and transportation networks to develop digital twins of supply chains or transportation networks, enabling a range of potential use cases (such as what-if simulations, interaction monitoring, and real-time location insights).
- *Construction teams* crawl and query unstructured data from sensors on buildings to derive insights that allow them to streamline design, production, and operations; for instance, they can simulate the financial and operational impact of selecting different types of materials for construction projects.

Key enablers

- *a modern data architecture* to support more flexible data stores
- development of *data models and digital twins* to replicate real-world systems

How to get started

- *Implement culture and technology changes* to modernize your data architecture.⁹
- *Identify critical data sets* (such as customer purchase frequency, customer attributes) that could later be organized into data assets (for example, a complete view of the customer) and *develop a taxonomy* for these data assets (for example, a business-data product such as “customer 360”).
- *Explore flexible ontologies and knowledge graphs* to map the relationship between different classes of data and data points.
- *Upgrade existing digital simulators*, replatforming them onto a cloud environment and updating APIs, to support more sophisticated AI capabilities such as reinforcement learning.¹⁰

⁹ “Breaking through data-architecture gridlock,” January 2021.

¹⁰ Giacomo Corbo, Oliver Fleming, and Nicolas Hohn, “It’s time for businesses to chart a course for reinforcement learning,” McKinsey, April 1, 2021.

4. Data operating model treats data like a product

Today

An organization's data function, if one exists outside of IT, manages data using top-down standards, rules, and controls. Data often has no true "owner" ensuring that it is updated and ready for use in various ways. Data sets are also stored—sometimes in duplication—across sprawling, siloed, and often costly environments, making it difficult for users within an organization (such as data scientists looking for data to build analytics models) to find, access, and integrate the data they need quickly.

By 2025

Data assets are organized and supported as products, regardless of whether they are used by internal teams or external customers. These data products have dedicated teams, or "squads," aligned against them to embed data security, evolve data engineering (for example, to transform data or continuously integrate new sources of data), and implement self-service access and analytics tools. Data products continuously evolve in an agile manner to meet the needs of consumers, leveraging DataOps (DevOps for data) and continuous integration and delivery processes and tools. Altogether, these products provide data solutions that can more easily and repeatedly be used to meet various business challenges and reduce the time and cost of delivering new AI-driven capabilities.

Everyday applications

- *Dedicated teams at retail companies* develop data products, such as "product 360," and ensure that the data asset continues to evolve to meet the needs of critical use cases.
- *Healthcare organizations*, including payers and healthcare analytics firms, stand up product teams to develop, maintain, and evolve "patient 360" data products to improve health outcomes.

Key enablers

- a *data strategy* that identifies and prioritizes business cases for data
- understanding of the organization's *data sources and the types of data* they hold
- an operating model that establishes a *data-product owner and team*—which could include analytics professionals, data engineers, information-security specialists, and other roles as needed

How to get started

- *Embed AI teams in the business* and empower them to design, develop, deploy, and continually enhance new AI-driven products using these data products.¹¹
- *Employ a data-governance operating model* that ensures data quality and treats data like a product.¹²

¹¹ "Getting AI to scale," May–June 2021.

¹² Bryan Petzold, Matthias Roggendorf, Kayvaun Rowshankish, and Christoph Sporleder, "Designing data governance that delivers value," McKinsey, June 26, 2020.

5. The chief data officer's role is expanded to generate value

Today

Chief data officers (CDOs) and their teams function as a cost center responsible for developing and tracking compliance with policies, standards, and procedures to manage data and ensure its quality.

By 2025

CDOs and their teams function as a business unit with profit-and-loss responsibilities. The unit, in partnership with business teams, is responsible for ideating new ways to use data, developing a holistic enterprise data strategy (and embedding it as part of a business strategy), and incubating new sources of revenue by monetizing data services and data sharing.

Everyday applications

- *Healthcare CDOs* work in partnership with business units to deliver new subscription-based services for patients, payers, and providers that can improve patient outcomes. Such services might include tailoring treatment plans, more accurately flagging miscoded medical transactions, and improving drug safety.
- *Bank CDOs* commercialize internal data-oriented services, such as fraud monitoring and anti-money-laundering services, on behalf of government agencies and other partners.
- *Consumer-products CDOs* partner with the sales team to use data to drive sales conversion and share responsibility for meeting target metrics.

Key enablers

- *data literacy* among business-unit leads and their teams to create energy and urgency to engage with CDOs and their teams
- an *economic model*, such as an automated profit-and-loss tracker, for recognizing and attributing data and costs
- *top data talent* with an eye for innovation
- adoption of *venture-capital-style incubator operating models* to support experimentation and innovation

How to get started

- For CDOs, *begin conversations with business-unit leaders* to identify opportunities for leveraging data to drive business value.
- *Develop holistic priorities*, underpinned by scorecards and metrics, that cover organizational health, talent, and culture, as well as data quality.
- *Reinforce the ethical use of data* to ensure that new revenue-generating data services align with corporate values and culture.¹³

¹³ *Tech: Forward*, "Ethical data usage in an era of digital technology and regulation," blog entry by Ewa Janiszewska-Kiewra, Jannik Podlesny, and Henning Soller, McKinsey, August 26, 2020.

6. Data-ecosystem memberships are the norm

Today

Data is often siloed, even within organizations. While data-sharing arrangements with external partners and competitors are increasing, they are still uncommon and often limited.

By 2025

Large, complex organizations use data-sharing platforms to facilitate collaboration on data-driven projects, both within and between organizations. Data-driven companies actively participate in a data economy that facilitates the pooling of data to create more valuable insights for all members. Data marketplaces enable the exchange, sharing, and supplementation of data, ultimately empowering companies to build truly unique and proprietary data products and gain insights from them. Altogether, barriers to the exchange and combining of data are greatly reduced, bringing together various data sources in such a way that the value generated is much greater than the sum of its parts.

Everyday applications

- *Manufacturers* share data with their partners and peers through open manufacturing platforms to build a more holistic view of worldwide supply chains.
- *Pharmaceutical and healthcare organizations* pool their respective data (for example, clinical-trial data gathered by pharmaceutical

researchers and anonymized patient data collected by the healthcare provider) so that each company can better achieve its goals.

- *Financial-services organizations* tap data exchanges to create new capabilities—for example, to support socially conscious investors by providing an environmental, social, and governance (ESG) score to publicly traded companies.

Key enablers

- *adoption of common data models* to facilitate ease of data collaboration
- *development of data alliances and sharing agreements*; several data-sharing platforms have emerged in recent years to facilitate the exchange of data both within and among institutions

How to get started

- *Read more* about the different types of data ecosystems and best practices for a successful ecosystem. There are examples in financial services, retail, and healthcare.¹⁴
- *Choose* the data-ecosystem archetypes that will be most important for your organization.¹⁵
- *Adopt* data-sharing tools, protocols, and procedures.

¹⁴ See Violet Chung, Miklós Dietz, Istvan Rab, and Zac Townsend, "Ecosystem 2.0: Climbing to the next level," *McKinsey Quarterly*, September 11, 2020; "Financial data unbound: The value of open data for individuals and institutions," McKinsey Global Institute, June 24, 2021; Julien Boudet, Jess Huang, Phyllis Rothschild, and Ryter von Difloe, "Preparing for loyalty's next frontier: Ecosystems," McKinsey, March 5, 2020; and Shubham Singhal, Basel Kayyali, Rob Levin, and Zachary Greenberg, "The next wave of healthcare innovation: The evolution of ecosystems," McKinsey, June 23, 2020.

¹⁵ Mohammed Aaser, Kumar Kanagasabai, Marcus Roth, and Asin Tavakoli, "Four ways to accelerate the creation of data ecosystems," November 23, 2020.

7. Data management is prioritized and automated for privacy, security, and resiliency

Today

Data security and privacy are often viewed as compliance issues, driven by nascent regulatory data-protection mandates and consumers beginning to realize how much of their information is collected and used. Data-security and -privacy protections are often either insufficient or monolithic, rather than tailored to individual data sets. Providing employees with secure data access is a highly manual process, making it error prone and lengthy. Manual data-resiliency processes make it difficult to recover data quickly and fully, creating risks for lengthy data outages that affect employee productivity.

By 2025

Organizational mindsets have fully shifted toward treating data privacy, ethics, and security as areas of required competency, driven by evolving regulatory expectations such as the Virginia Consumer Data Protection Act (VCDPA), General Data Protection Regulation (GDPR), and California Consumer Privacy Act (CCPA); increasing consumer awareness of their data rights; and the increasingly high stakes of security incidents. Self-service provisioning portals manage and automate data provisioning using predefined “scripts” to safely and securely provide users with access to data in near real time, greatly improving user productivity.

Automated, near-constant backup procedures ensure data resiliency; faster recovery procedures rapidly establish and recover the “last good copy” of data in minutes rather than days or weeks, thus minimizing risks when technological glitches occur. AI tools become available to more effectively manage data—for example, by automating the identification, correction, and remediation of

data-quality issues. Altogether, these efforts enable organizations to build greater trust in both the data and how it’s managed, ultimately accelerating adoption of new data-driven services.

Everyday applications

- *Retailers* with an online presence specify the data from consumers that they collect and develop consumer portals to obtain consent from users and allow them to “opt in” to personalized services.
- *Healthcare and governmental institutions* with highly sensitive data institute advanced data-resiliency protocols that automatically back up data multiple times daily and, when needed, identify the “last good copy” and restore it seamlessly.
- *Retail banks* automatically provision credit-card data needed to support customer-facing applications, specifically during development or testing, to improve developer productivity and provide access to data more efficiently and securely than is possible with traditionally manual efforts today.

Key enablers

- *elevating the importance of data security* throughout the organization
- *rising consumer awareness* of, and active involvement in, individual data-protection rights
- *adoption of automated database-administration technologies* for automated provisioning, processing, and information management
- *adoption of cloud-based data-resiliency and -storage tools* to facilitate automatic backup and restoration of data

How to get started

- *Consider adopting a data-ethics framework* to understand and evaluate potential ethical and regulatory ramifications of data and analytics activity, especially involving consumer data.¹⁶
- *Consider leveraging cloud tools to store, manage, and secure priority data*, and, for data already residing on the cloud, leverage automated backup and resiliency capabilities and tools as part of cybersecurity policies.¹⁷
- *Create a road map* for migrating to new automatic provisioning and resiliency capabilities as they evolve.
- *Adopt a frequent, iterative approach to developing, reviewing, and revising governance and control protocols* to take advantage of forthcoming opportunities to automate database administration—for example, by setting up a self-service provisioning portal and mandating automated backup and restoration procedures on compatible data platforms.

Neil Assur is an associate partner in McKinsey's Philadelphia office, and **Kayvaun Rowshankish** is a partner in the New York office.

Designed by McKinsey Global Publishing
Copyright © 2022 McKinsey & Company. All rights reserved.

¹⁶ See "Ethical data usage," August 2020; and Venky Anant, Lisa Donchak, James Kaplan, and Henning Soller, "The consumer-data opportunity and the privacy imperative," McKinsey, April 27, 2020.

¹⁷ "Security as code: The best (and maybe only) path to securing cloud applications and systems," McKinsey, July 22, 2021.