

CIO Guide

MODERNIZING YOUR DATA INFRASTRUCTURE





Introduction

Data has become much more than a business commodity—it is the beating heart of today's successful organizations. Data drives decisions and solves problems to help organizations become more agile, efficient, and innovative.

The high volumes of data being created by digital businesses have increased the urgency for CIOs to deploy a data infrastructure that goes beyond the on-premises databases and data warehouses of the past. A modern data strategy lets you store any amount of data you need at low cost and in open, standards-based data formats. It makes data more accessible while keeping it secure, empowering people to run analytics or machine learning (ML) using their preferred tool or technique to make the data actionable for specific tasks or use cases.

For organizations running legacy data infrastructure on-premises or self-managed in the cloud, overseeing this infrastructure is tedious, time-consuming, and expensive. IT teams spend time worrying about hardware and software installation, configuration, performance and availability, capacity planning and cluster scaling, and security and compliance issues.

Further, data modernization is critical to creating a foundation for new business and revenue opportunities, streamlining crucial decision-making, and improving efficiencies. CIOs and other IT leaders have come to view data and business analytics as a top driver of IT investments to deliver business objectives. In [IDG's State of the CIO study](#), data and business analytics technology initiatives topped the list of planned IT investments for 2021.

By modernizing their data infrastructure, organizations move away from on-premises data stores and onto the cloud. With the cloud, organizations access IT resources like storage, database, analytics, and ML over the internet instead of buying, owning, and maintaining physical data centers and servers. The cloud platform takes care of routine management tasks such as server provisioning, patching, configuration, and backups.

This e-book explores why data modernization is a defining moment for any organization looking to reinvent itself, and the infrastructure components CIOs can put in place to ensure a smooth modernization journey.

Contents

- 3 **[Business imperatives for modernizing data infrastructure](#)**
- 4 **[IT imperatives for modernizing data infrastructure](#)**
Benefits of data modernization in the cloud
- 5 **[Key components of a modern data infrastructure](#)**
 - Purpose-built databases
 - Data warehouse
 - Security and compliance*
 - Storage
 - Analytics systems
- 8 **[Take the next step in your data modernization journey](#)**
- 8 **[Sponsor viewpoint](#)**

Business imperatives for modernizing data infrastructure

Data has become a strategic asset and a catalyst for launching new business models, revenue opportunities, and other innovations. Nasdaq, for example, [built out its portfolio](#) of financial data and analytics services over the past decade to meet changing customer demands. That wouldn't have been possible without moving much of its existing data infrastructure to the cloud. In particular, customer demand for real-time data has helped drive the organization's digital transformation.

"Being able to get financial updates in real time is becoming table stakes," says Michael O'Rourke, Senior Vice President and Head of AI/Technology, Investment Intelligence at Nasdaq. "For trading applications, being able to visualize the analytics and assess the risk requires that you have real-time data in many cases."



Innovation takes many forms, including reducing risk for organizations and their customers. NuData Security, a Mastercard company, offers fraud-detection solutions to help banks, insurance companies, e-commerce sites and other businesses thwart unauthorized access and account-takeover attempts. Its NuDetect service [analyzes and correlates petabytes of data](#) each day, mitigating what amounts to millions of attacks daily.

Built in 2008 in a private data center, NuDetect migrated gradually to the public cloud to support its growing data and performance needs. Today, NuDetect runs exclusively on AWS cloud infrastructure, using a combination of machine learning and sophisticated rules engines to filter out malicious behaviors with minimal impact on legitimate users.

"Data is in an Amazon S3 data lake and everything just flows from one place to another," says Justine Fox, Director of Software Engineering at NuData. The environment frees NuData teams to focus on feature additions, cost optimizations, and other value-added activities, Fox says.

Becoming more data-driven is at the heart of digital transformation and business reinvention, and CIOs need to view it as a continuous exercise, not a point-in-time project.

"First and foremost, you need to be business-driven before you're data-driven," says Michael Gabriel, a Partner at Fortium Partners who has held CIO and IT executive positions at HBO, the National Basketball Association, and EMI Records. "Digital transformation isn't a goal, it's a process through which you can better achieve a goal. Technology doesn't stand still—and neither does business—so digital transformation is always ongoing and dynamic."



"Technology doesn't stand still—and neither does business—so digital transformation is always ongoing and dynamic."

Michael Gabriel
Partner, Fortium Partners

IT imperatives for modernizing data infrastructure

Technology most definitely does not stand still, which is why CIOs are under pressure to modernize their data infrastructures to support business-driven transformation.

For many years, IT teams successfully built and maintained large, general-purpose databases or data warehouses to store a variety of information that different departments could access to run the business. However, with the massive amounts of data that organizations are creating and collecting, this one-size-fits-all strategy no longer works.

"Given the explosion in data volumes and the types of data that customers are dealing with, organizations quickly reach a point where they need a different technology," says Rahul Pathak, Vice President for Analytics at AWS. "With a general-purpose database, you'll end up in a dead end at some point because you're trying to do everything reasonably well, which means you can't be excellent at any one particular thing. A more focused approach lets you build something where there's no compromise on performance, functionality, scale, or cost."

Technologies such as machine learning and artificial intelligence (AI) are also driving the need to rethink data infrastructure. On-premises data warehouses cannot handle the processing power and storage required to run machine learning algorithms on petabytes of data.

"While businesses have been using data analytics forever, the emerging areas of machine learning and AI are providing capabilities never before achievable," says Gabriel. "CIOs can, and should, be leading that charge by understanding how to work with the business to build those capabilities where they make business sense."





CIOs also need to help their organizations better understand all of the data they are collecting, since it likely lives across cloud services, in on-premises servers and storage systems, and on employees' work devices.

"Knowing the data you have available or could have available, and how it needs to be utilized to support analytics, is problematic," says Gabriel. "This has gotten far worse with the exponential increase in the data we can now capture and store relatively cheaply. The rise of SaaS availability has created muddy pools of data that exist within a company that are often unknown to IT or to the departments that could leverage that data."

While the cloud has amplified some of these challenges, it's also a big part of the solution. A tightly integrated, cloud-based ecosystem of storage, compute, and analytics provides better performance, increased scalability, and seamless access to data in ways that are unlikely or impossible with on-premises infrastructure.

Benefits of data modernization in the cloud

Modernizing your data infrastructure by migrating storage, data, and analytics services to the cloud can deliver many benefits. Here are a few examples:

-  **Operational and cost efficiencies**
Samsung's migration to a cloud-based relational database helped reduce its monthly database costs by 44% and maintenance fees by 22%.
-  **Performance and availability**
Cathay Pacific saw a 20% performance improvement after moving its on-premises passenger revenue optimization system to the cloud.
-  **Scalability**
Moving from an on-premises data warehouse to a cloud-based data infrastructure allowed Nasdaq to increase transactions from 30 billion records a day to 70 billion records with no disruption.
-  **Security and compliance**
Since moving to the cloud, NuData's fraud-detection service has thwarted millions of attacks daily—protecting more than 100 million accounts every month, with 99% accuracy and a sub-0.1% false-positive rate. ML services on Amazon SageMaker have resulted in a 60-70% velocity increase in determining fraudulent activity.

Key components of a modern data infrastructure

By modernizing data infrastructure, organizations can move away from on-premises data stores and into a cloud-based environment for databases, storage, and analytics. Here are the four primary components you need to modernize to deploy a data and analytics infrastructure that supports today's business and IT imperatives.

1. Purpose-built databases

With data volume requirements increasing from gigabytes and terabytes to petabytes and beyond, the same database solutions and tools that worked in the past can no longer keep up. Traditional relational databases, once the default choice, are now just one option for building highly resilient, scalable applications more economically.

Instead of monolithic, on-premises databases, consider moving to a model defined by building the right database for the right job. Choosing the right purpose-built database is determined by three factors:

▶ **Application workload:** Understand the type of data being stored and the access patterns. These fall into one of three categories: transactional (for a high number of concurrent applications), analytical (aggregating and summarizing large sums of data, which operate on many more rows per query), and caching (for read-heavy workloads that require faster load times to improve response times for end users).

▶ **Performance and scale requirements:** Consider not just the speed of the database, but also how it will serve end users. Developers should ask whether the database will be customer-facing versus internal-only. Geographically, if a database is closer to users, lower response times are achievable.

▶ **Type of data:** Understand the types of entities and their relationships.

- Relational databases normalize data into separate data tables.
- A key-value or wide-column database is designed for scale, with data split across multiple storage nodes.
- A document database is for large records that assemble heterogeneously for frequently accessed data.
- A graph database emphasizes relationships between data to find relationships. Examples here include social network relationships or fraud-detection services.
- An in-memory database is designed for caching, session management, leaderboards, and geospatial applications.
- A time-series database is ideal for internet of things (IoT) applications, devops, and industrial telemetry.
- A ledger database is ideal for systems of record and banking transactions.



Purpose-built databases are a critical part of data-driven transformation because they enable development teams to pick the right technology for specific business needs, without any tradeoffs on functionality, performance, or scale. This approach gives organizations the ability to innovate quickly to address changing business or customer demands.

Key components of a modern data infrastructure

2. Data warehouse

On-premises data warehousing systems also cannot keep up with the rising volumes and complexity of data. With rigid architectures that require significant investment to maintain, update, and secure, they do not give organizations the opportunity to make the most of their data.

Moving to a cloud data warehouse removes these limitations. You can run queries across petabytes of data in your data warehouse and extend into a data lake.

A modern data warehouse automates most of the common administrative tasks to manage, monitor, and scale, freeing IT teams to focus on more productive, high-value endeavors. It also delivers fast query performance, improves I/O efficiency, and scales up or down as your performance and capacity needs change.

CIOs can opt for migrating an existing data warehouse to the cloud, but a better approach may be to start with a use case that is not well-served by the current infrastructure. "For example, many organizations decide they want to do something with machine learning to reduce their customer churn, so they run a proof-of-concept project," says Pathak. "That helps them understand how to work with the cloud and how to manage data in the cloud. Then that success leads to more momentum, which may then bring in legacy processes."



Security and compliance

A modern data infrastructure strikes the right balance between giving people access to data that they need, while also protecting the data for privacy or compliance regulations.

"A common mistake we've seen is that organizations end up creating overly rigid governance scenarios, where the central teams become a bottleneck to the lines of business trying to work with data to improve the business," says Rahul Pathak, Vice President for Analytics at AWS. "The key to good governance is figuring out how to define access, then getting out of the way. By that I mean

creating exception processes, rather than taking an approach that anytime you need data, you have to ask someone in a central organization for it."

A successful data governance framework includes:

- ▶ **access control policies** for people, applications, and servers
- ▶ **a zero-trust security model**
- ▶ **geographical compliance rules**
- ▶ **a central mechanism** for managing data and enforcing governance policies.

Key components of a modern data infrastructure

3. Storage

On-premises storage can be costly and complex, with expensive hardware refresh cycles and data migrations required to support system upgrades. In addition, gaining insights from data is difficult if data is trapped in silos or other “muddy pools” across the organization.

Moving storage to a cloud consumption model lets companies adjust on the fly and use whatever storage they need now—without being locked into a hardware refresh. It can also keep organizations agile, reduce costs, and provide for unlimited scalability while also eliminating data silos.

Moving storage workloads from on-premises systems to the cloud can reduce total cost of ownership through a flexible buying model that helps to eliminate over-provisioning, shorten refresh life cycles, and reduce the cost of maintaining storage infrastructure.

4. Analytics systems

Managing open-source analytics software like Apache Hadoop/Spark, Elasticsearch, and Apache Kafka on-premises is complex, time-consuming, and expensive. Challenges to this approach include keeping a dedicated team of experts to manage hardware and software configuration, patching and backups, tuning and optimizations for performance, and capacity planning for future growth. A move to managed analytics in the cloud can save time, reduce costs, and significantly improve productivity.





Take the next step in your data modernization journey

Deploying a modern data infrastructure requires alignment between business and IT goals. As businesses look to reinvent themselves, CIOs have an opportunity to steer the journey with a data infrastructure that makes it easier to access, and act upon, the information that drives business growth and innovation.

A cloud-based data infrastructure ultimately positions organizations to adapt more quickly to changing markets and customer needs. "This approach will endure because it's the right way to build things," says Pathak. "It's decoupled. It lets you scale. It gives you option value for the future and lets you bring in new technology choices."

In other words, a modern data foundation will accelerate your ability to capture value from ever-growing volumes of data, while keeping it secure, to help you focus on innovation, not infrastructure.

Sponsor viewpoint

Leveraging data as a strategic asset can help your organization meet customer expectations and remain competitive. But the insights and truths data offers must be accessible to every facet of your business.

If you're committed to reinventing your organization to a data-driven one, you can empower everyone to unlock data's potential and innovate in new ways. The move from on-premises and self-managed data solutions to fully managed cloud data services is a critical first step. Teams that are now free from managing complex and expensive infrastructures can spend time innovating and building new applications.

With the right cloud partner, you can make this initial move immediately. Organizations want to be confident their choice of technology will deliver value from their data while keeping it secure and compliant across a broad and ever-changing set of regulations. More organizations partner with AWS to do this than anyone else, with over 450,000 databases migrated from on-premises to the cloud using our database migration service. We've also completed this data modernization ourselves: Amazon.com migrated 75 petabytes of internal data stored in nearly 7,500 Oracle databases to AWS databases and reduced costs by over 60%.



AWS provides you with an easy path to increased data access for all, greater visibility of your data, and immediate peace of mind. Migration to fully managed databases, storage, and analytics services on AWS allows you to offload infrastructure management tasks to AWS and focus on building applications. Join the dynamic AWS community of customers and take the first step of your data modernization journey.

▶ [Learn more](#)