



The Chief Data Officer's guide to digital transformation

How to push your business' data operations to the forefront of innovation and data-driven value.

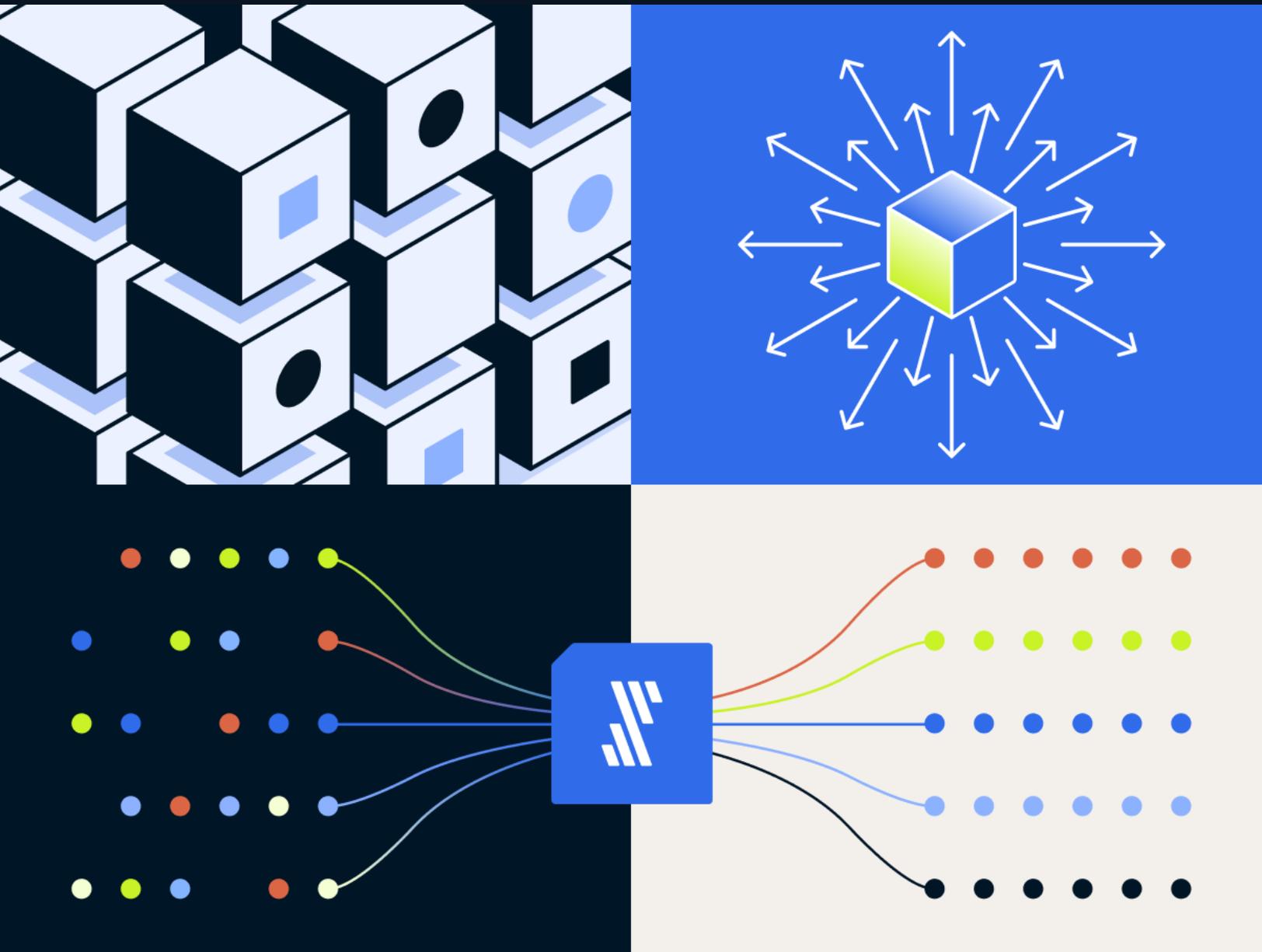


Table of contents

	Executive summary	3
	Introduction: A structured approach to data	4
	The CDO's mandate	5
1	Data centralization	7
	Why data centralization?	7
	Data centralization challenges	7
	How to solve data centralization challenges	8
	Enterprise spotlight: DocuSign	9
2	Infrastructure modernization	10
	Why modernize infrastructure?	10
	Infrastructure modernization challenges	11
	How to solve infrastructure modernization challenges	11
	Enterprise spotlight: Oldcastle Infrastructure	12
3	Data democratization	13
	Why democratize data?	13
	Data democratization challenges	13
	How to solve data democratization challenges	14
	Enterprise spotlight: Autodesk	15
4	Building data solutions	16
	Why build data solutions?	16
	Data solution challenges	17
	How to solve data solution challenges	17
	Enterprise spotlight: Saks	18
	A roadmap for innovation and data-driven value	19

Executive summary

Chief Data Officers (CDOs) are tasked with bringing digital transformation to their organizations. In the context of data, digital transformation means moving from simple, reactive uses of data, such as descriptive analytics, to innovative, higher-value uses of data, such as predictive analytics and generative AI.

Chief Data Officers face challenges such as institutional inertia, a lack of clear expectations and more:

- ◆ Teams must make quick, high-stakes data-driven decisions but decentralizing decisions carries its own risks.
- ◆ Competition and macroeconomic headwinds mean that organizations strive to do more with less headcount and budget.
- ◆ Data grows in the volume and complexity of its handling. Governments write increasingly stringent regulations, while consumers expect greater personalization and security.

These challenges contribute to the notoriously high turnover of CDOs, with an average tenure of fewer than **30 months**. This guide is meant to help CDOs identify critical junctures on their roadmaps to achieve digital transformation before the clock runs out.

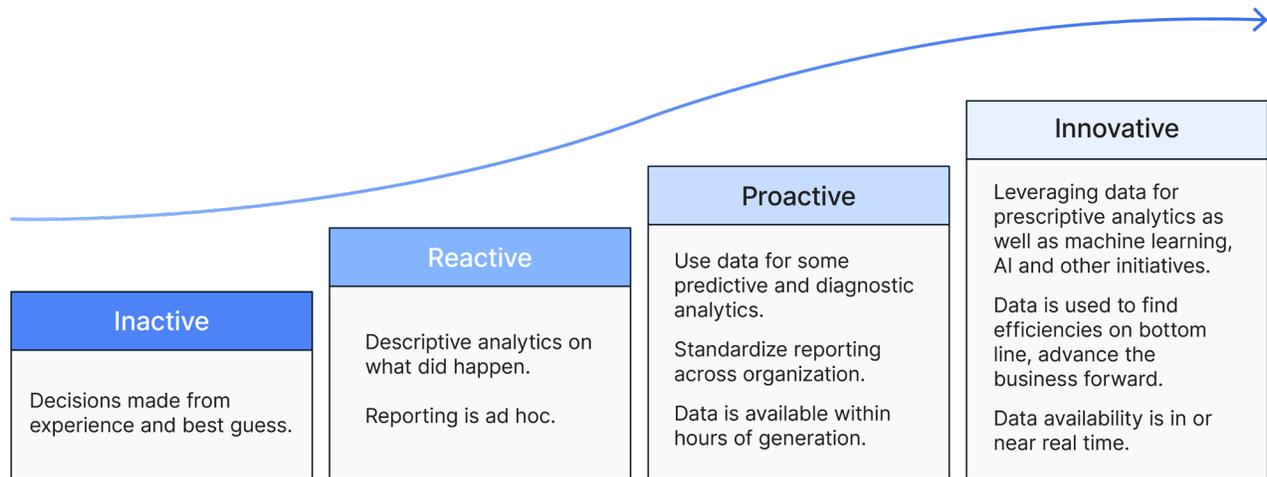
A roadmap for digital transformation must include four stages – data centralization, infrastructure modernization, data democratization and building data solutions. These stages share a technological solution, the automated **data movement** platform. A CDO is responsible for building a roadmap to digital transformation.

- 1. Data centralization**, the first step, is the classic data integration use case. As organizations adopt a growing range of applications and operational systems, they must make the data accessible. A comprehensive view of an organization's operations, customers and products unlocks all analytical and operational uses of data.
- 2.** Once data is centralized, **data democratization** brings data accessibility to stakeholders across an organization, enhancing organizational agility. However, it also poses security and governance challenges.
- 3. Building data solutions** consists of monetizing data. Raw data, insights, predictive models and artificial intelligence are valuable examples but require data sharing and programmatic control.

As an organization progresses through the above stages, it will necessarily modernize its infrastructure, such as by moving data repositories from on-premises to the cloud.

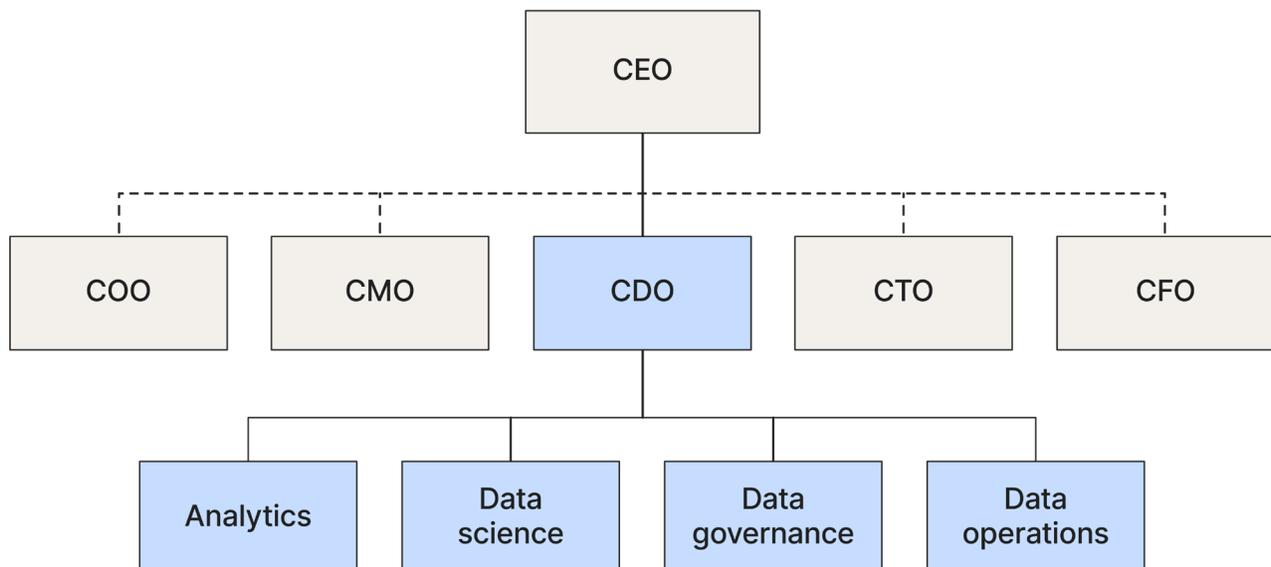
Introduction: A structured approach to data

Data is an essential resource for supporting decisions, automating processes and powering innovative products. Its value increases with freshness and scale. Data follows a maturity curve with four steps:



Data-driven companies that are proactive and innovative with data are far more likely to thrive than companies that rely on guesswork and ad hoc reporting.

Organizations today use a wide range of systems that produce data, ranging from applications and databases to event streams and files. These digital footprints can produce invaluable insights. To lead efforts to gain control over the growing volume, velocity and variety of data created by business activities, many organizations have created the position of the Chief Data Officer (CDO).



The CDO's mandate

As leaders of a company's data organization, CDOs are responsible for the development and management of all data assets and their associated teams, technologies and workflows. All data-related operations begin with data movement:

- 1 Extracting raw data from sources
- 2 Loading raw data to destinations
- 3 Transforming raw data into data models for analysis and operational usage
- 4 Using data models to build products such as dashboards, reports, predictive models, automated processes and more

NOTE *Many organizations practice ETL, swapping steps 2 and 3, but ELT is a better approach. **More on that later!***

In turn, there are four major use cases for data movement:



Data centralization – consolidating data on a central platform for analytics



Infrastructure modernization – adopting more capable infrastructure



Data democratization – promoting self-service analytics across an organization

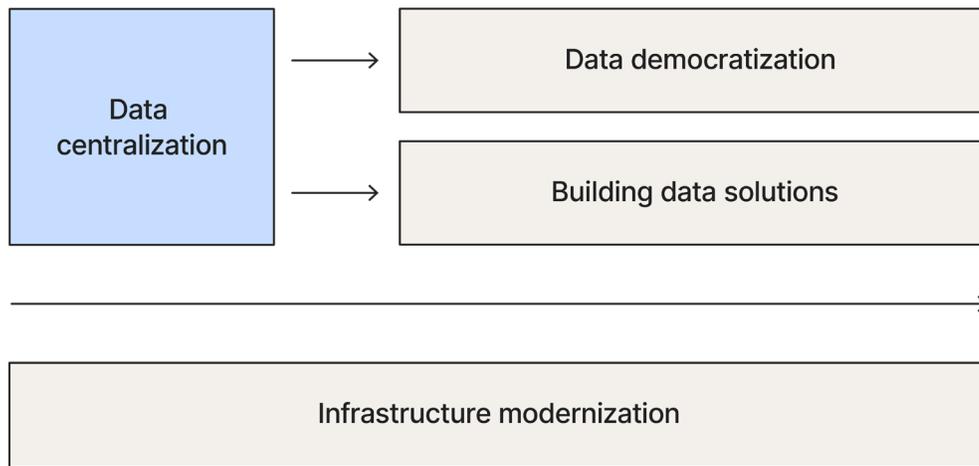


Building data solutions – sharing data and building systems powered by data

These use cases form a progression:

- 1 **Data centralization** comes first, as without data residing on a single platform it is difficult to systematically make use of data.
- 2 Once data is centralized, a profusion of data becomes available to users and **data democratization** leads organizations to weigh the tradeoffs between widespread access to data and preventing unsanctioned use.
- 3 With a solid grasp of data, an organization monetizes it by **building data solutions**.

As an organization progresses through data centralization, data democratization and building data solutions, it will revise its tools, technologies and platforms through **infrastructure modernization**.



These data movement use cases require a technological solution in the form of an **automated data movement platform**, which features three pillars:



Automation: The processes and technologies used to move data must minimize engineering time. Labor is a costly input for all organizations. Automation enables organizations to become agile by producing insights quickly.



Reliability: Data movement must involve minimal maintenance and downtime. Data pipelines must not be disrupted by schema changes, failed syncs or other issues.



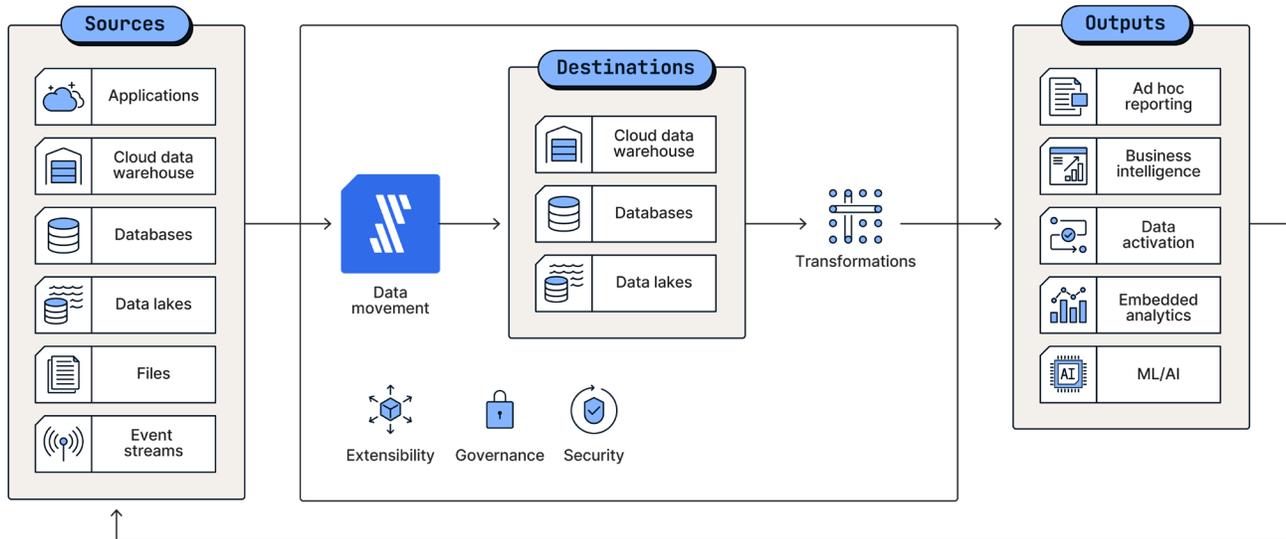
Scalability: Data operations must be able to scale seamlessly while supporting a huge range of data types and sources, governing the usage of data across departments and programmatically managing data movement.

This guide will discuss the central importance of an automated data movement platform and how CDOs (and CIOs and CTOs) can use it to transform data assets into value. Along the way, we will think through each key use case for data movement, specifically:

1. Why it matters
2. The challenges involved
3. How to address those challenges
4. Real-world examples of successful implementations

1. Data centralization

Data centralization means consolidating data from a variety of sources into a central destination. This destination may be a data warehouse, governed data lake or lakehouse. From there, data can be used for business intelligence, machine learning and artificial intelligence.



Why data centralization?

Data centralization ingests records from disparate data sources into a platform for analytical and operational uses. Data teams then create data models and use business intelligence platforms to produce visualizations and dashboards, supporting the discovery of insights to support business decisions and laying the groundwork for advanced uses such as AI.

Data centralization enables a 360-degree view of an organization's operations, customers and products, enabling improved decision-making of all kinds.

Data centralization challenges

Centralizing data involves solving the following problems:

- 1 Integrating a wide and growing range of sources
- 2 Ensuring that syncs run reliably and are resilient to upstream schema changes
- 3 Maintaining and upkeeping existing data connections as endpoints are updated
- 4 Guaranteeing data integrity and offering visibility into the status of syncs and data
- 5 Securing sensitive data as it moves from source to destination

Moving data from a source to a destination is a **deceptively complex engineering problem** involving considerations like designing the correct architecture, provisioning the appropriate compute and storage resources, ensuring timely updates with minimal disruption to source systems, building in resilience to failure, ensuring data security and more.

Centralizing sensitive data — financial, customer, operations, patient — can be particularly challenging. As a team builds pipelines they have to consider regulations around the data, keeping it secure in transit and ensuring access only to the authorized parties. Security teams may even mandate that this data can't be processed by third parties, so an organization must use and maintain self-hosted software or go the do-it-yourself route. But, building and maintaining a DIY pipeline demands considerable time, labor and money.

The DIY approach also jeopardizes organizational agility. There may be considerable turnaround time between the need for an insight and the availability of data. Disruptions, such as pipeline failures, cause downtime as well. Such delays kill the ability to make timely decisions supported with data. The inherent difficulty of moving data is multiplied by the fact that the data used by organizations is scattered and siloed across a wide variety of tools and platforms, each with its own **idiosyncrasies**.

Throwing engineering time at data centralization turns a data team into a cost center rather than a source of value and insight.

How to solve data centralization challenges

Data centralization is best solved using a fully managed, automated data movement platform. As you look for a solution, make sure to:

- ◆ Find a data platform that is easy to use out of the box, with minimal configuration and engineering time to start.
- ◆ Consider how the platform ensures reliability, including the ability to replay failed syncs without duplication, handle upstream schema changes and optimize pipeline and network performance.
- ◆ Ensure that the data pipeline uses an ELT architecture, rather than ETL. This simplifies the data pipeline, enables secure data processing and leverages the scalability of the destination system for transformations.
- ◆ Check that it supports your current sources and destinations as well as those you are likely to use in the future. This is critical to ensuring that your data infrastructure can keep up with organizational growth and scaling.
- ◆ Consider the product's security features, especially if your organization is in a highly regulated industry where sensitive data must not be accessed by unauthorized parties.

The most basic unit of data centralization is the data connector, a pipeline that connects a source to a destination. A well-functioning connector must be able to update from a source without disrupting operations, be resilient to failed syncs and other stoppages, accommodate changing data models at the source and be highly performant. Ideally, even non-technical stakeholders should be able to activate data connectors at will and largely disregard them thereafter. By contrast, DIY data connectors create an open-ended commitment to support bespoke systems with engineering hours.

DID YOU KNOW

Data engineers spend almost half their time maintaining data pipelines. **The total average cost? \$520,000 per year, according to Wakefield Research.**

Highly-skilled engineers and analysts shouldn't spend time reinventing the wheel. Rather, they should leverage a product that solves a known problem, supported by a provider with expertise dealing with specific data sources and destinations. This will enable your data team to focus on modeling, analyzing or operationalizing data.

CUSTOMER SPOTLIGHT

[Read the full customer story →](#)

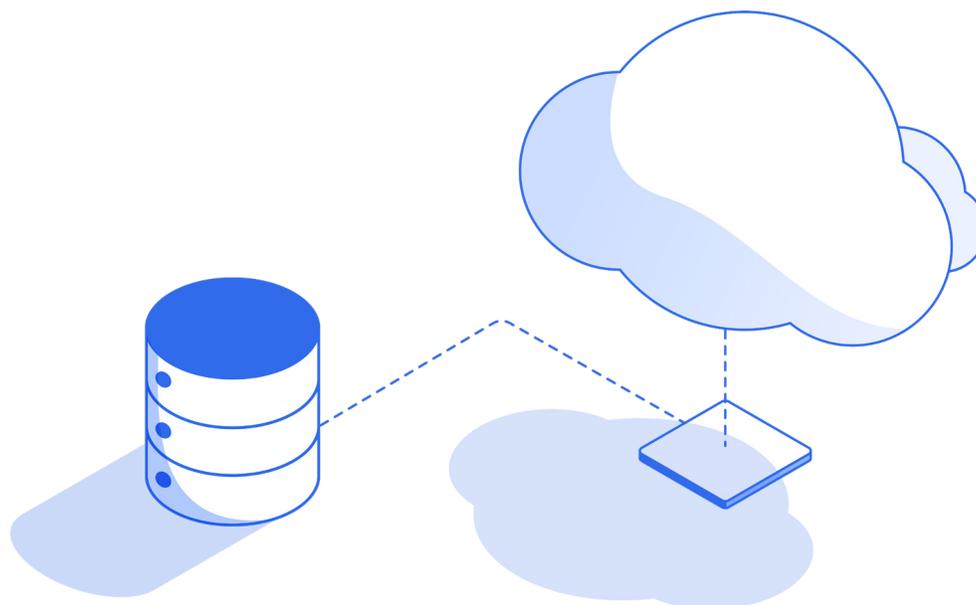


DocuSign, the electronic agreement and signature provider, offers an example of successful data centralization enabling greatly expanded data operations. Originally, DocuSign made the common engineering choice of duplicating a SQL Server database for analytics. People were confined to analyzing data from within each silo or extracting CSVs into spreadsheets.

As the company grew rapidly, the business intelligence team realized it needed a new, systematic approach to sustainably add new data sources. Marcus Laanen, Senior Manager of Business Intelligence, estimated "It would take a highly paid engineer anywhere from three to six months to build out a data pipeline and up to 20 hours a week afterward to keep things running. Our team would have to double in size."

To avoid manual data integration, DocuSign migrated its analytics environment to Snowflake and began using Fivetran. DocuSign has been able to triple the number of sources it analyzes data from and maintains over 100 active dashboards that are regularly used across its organization. Its engineers now work on core projects rather than data engineering, while analysts engage in data modeling and cataloging.

2. Infrastructure modernization



Infrastructure modernization involves changing the tools, technologies and platforms an organization relies on for data operations. It can take place along several dimensions:

- ◆ An organization may move from on-premise operational systems, pipelines and destinations to the cloud for greater flexibility and performance (and outsourced maintenance).
- ◆ Data teams may switch a data pipeline from ETL to a more modular, flexible ELT-based architecture.
- ◆ An organization may **migrate from one type of destination to another** for a better cost structure or performance.
- ◆ A data team might upgrade its data movement capabilities from batch updating to streaming.

For smaller or newer organizations, this is typically greenfield – a change from a crude or nonexistent data stack to a cloud-based data stack. Without an incumbent solution, it is relatively straightforward to adopt and start using a new solution. For larger and more established organizations rip-and-replace is seldom viable due to the importance of workflows that depend on the existing infrastructure.

Why modernize infrastructure?

Fundamentally, modernizing infrastructure is about improving the capabilities of your data organization by improving its tools. Important benefits include improved flexibility in terms of compute and storage, lower costs, reduced engineering workloads and interoperability with new technologies.

Centralizing data is the essential first step to enable further uses of data, including democratizing data and building systems that monetize data. Infrastructure modernization plays a major role in all of these stages.

Infrastructure modernization challenges

A fundamental stumbling block to infrastructure modernization is that ongoing, mission-critical analytical and operational processes depend on existing infrastructure that cannot be shut down without jeopardizing an organization's existing commitments. Furthermore, the migration itself must not interfere with the performance of operational systems. This means that migrations can be costly; often, both legacy and new environments will be simultaneously active as old data is migrated to the new platform. In this regard, smaller, leaner organizations without existing obligations have a real agility advantage and can simply start using a new system with minimal disruption.

Another key challenge is security and compliance. Once data is in the cloud, it is no longer confined to proprietary infrastructure (although it may be secured via a private cloud). Data leaders must also prevent unauthorized access while data is in transit. Regulatory compliance and brand risk can be managed with strong authentication and authorization protocols, end-to-end encryption and other security features.

Finally, it is easy to offer too much access to too many parties too quickly, leading to ballooning compute and storage expenses. Fortunately, data governance – an issue we will address in the next section – offers a means to get these expenses under control.

How to solve infrastructure modernization challenges

If your organization has existing obligations, you must keep existing processes running while you set up the new ones. There are several ways to contain the costs of modernization.

You don't need to migrate all parts of your infrastructure at once and it can make sense to modernize infrastructure piecemeal. One possibility is hybrid cloud architectures in which some elements of the infrastructure remain on-premises or in private clouds while others are hosted in the cloud. This is especially relevant in sensitive, highly-regulated industries. As a starting point, however, many organizations choose a managed, cloud-based data platform as a destination.

Another consideration is that modernization, specifically the design, construction and maintenance of data connectors, is costly and complex. The solution is to adopt a data movement platform that features off-the-shelf data connectors and supports both on-premises and cloud-based data sources.

It can be a good idea to start a migration effort with applications, which produce valuable data (especially in sales and marketing) but tend to be less technically demanding than operational databases. Migrating applications is a low-risk, low-cost approach to building a strong proof of concept. Building on initial successes, your team will be able to approach subsequent efforts with more confidence, activating connectors to more data sources, including operational databases, over time.

For the sake of security, compliance, privacy and access control, look for regulatory compliance features, end-to-end encryption and the ability to designate different levels of access by role (i.e., role-based access control or RBAC) within the organization.

CUSTOMER SPOTLIGHT

[Read the full customer story →](#)



The building materials company **Oldcastle Infrastructure** maintained both an on-premises operational database and cloud-based NetSuite ERP. This hybrid architecture limited its ability to view transactional, manufacturing and production data in a single environment. To solve this problem, the team initially tried to migrate its SQL Server database to Azure cloud. This effort took five months of planning and was three months into execution when Nick Heigerick, IT Manager of BI, realized “We didn't have the expertise to manage so many moving parts in-house, so we had to decide: Are we going to pay consultants to do this forever or is there a better way?” In short, Nick and his team realized that this approach would struggle to accommodate new data sources and scale the company's usage of data.

Oldcastle Infrastructure replaced Azure out for Snowflake and connected Fivetran to both its on-premise SQL Server and NetSuite instances. In just 10 days, the entire data set was migrated to the cloud, resulting in the following benefits:

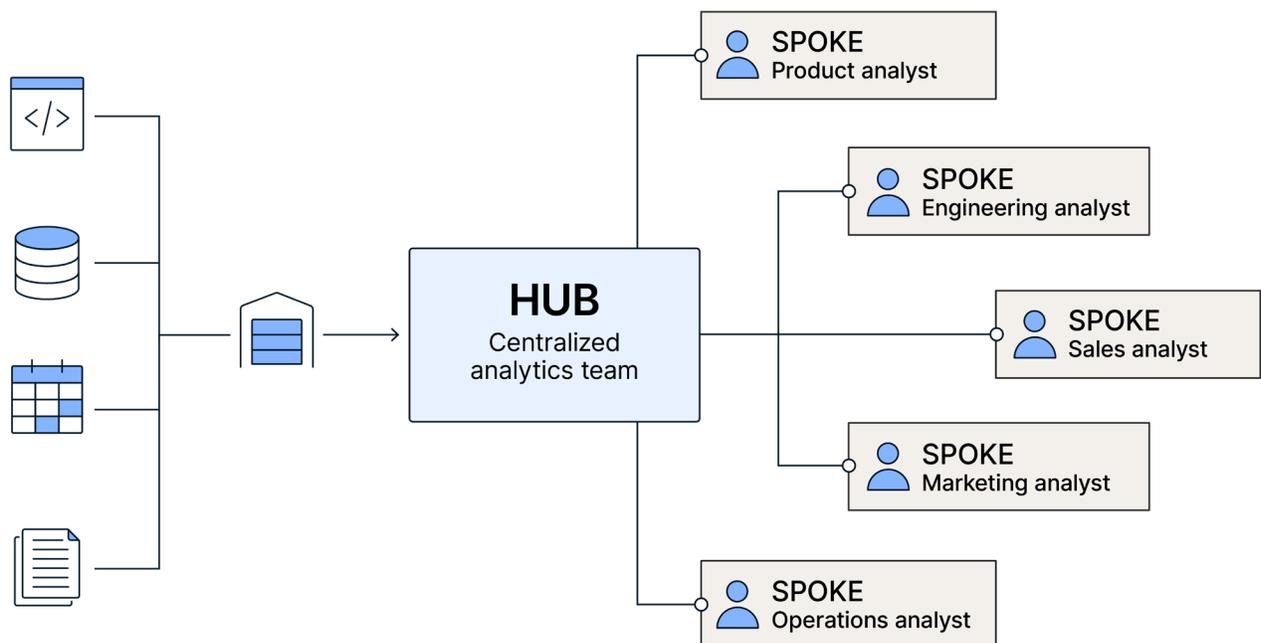
- ◆ \$360,000 in annual infrastructure savings
- ◆ 21.5% growth in operating profit within one year
- ◆ Projected ROI of at least \$25 million in the next 12-18 months
- ◆ The ability to sync data from Salesforce, Coupa, Box and other SaaS applications

3. Data democratization

Data democratization is a key goal of CDOs and consists of promoting data literacy and self-service analytics across an organization to become truly data-driven. This requires trust in the quality, accuracy and reliability of data. With robust security and governance measures, different teams across an organization are empowered to own their pipelines and access their data from a centralized platform. Data democratization enables everyone from individual contributors to executives to routinely consult and use data to support decisions.

Why democratize data?

Data democratization offers data-driven decision making to business users who know their domain best, enabling an organization to become more agile and responsive to changing market conditions. This approach is especially suited to companies with decentralized data teams.



Data democratization challenges

Every organization contains three key groups of stakeholders:

- ◆ **Data consumers** such as analysts want unrestricted access to data to facilitate their projects. They are frustrated by slow turnaround and stale data and often circumvent existing data service processes, accessing and producing data products in unapproved ways.

- ◆ **Data producers**, typically data engineers, are responsible for managing a growing queue of data integration projects. They are forced to balance the interests of the security and legal teams with those of analysts. For security and legal teams, they must have answers for audits. For analysts, they must onboard and deploy new data sources quickly and reliably.
- ◆ **Security and legal teams** prioritize regulatory compliance, especially in the face of continued regulatory changes. Their main goal is to minimize data misuse and the risk of data breaches. Security and legal teams want to prevent data mishandling that risks fines and reputational damage.



Data consumers



Data producers



Security and Legal



Given the high stakes involved with improper exposure and misuse of data, organizations have strong incentives to impose stringent limitations on access in order to guarantee compliance. However, red tape can cause analysts and business users to experience long turnaround times and stale data, and miss opportunities.

The core challenge of data democratization is governance, and its core tension is between access and compliance.

How to solve data democratization challenges

The solution to simultaneously ensuring access and compliance comes down to observing, controlling and scaling authorized access to data.

Observing data requires full visibility of all the data that is being generated across the organization. Consolidating a view of all the data in a data catalog via metadata sharing allows data stewards to track column-level lineage and schema changes to see what data exists and how it changes over time.

To control data, an organization needs to be able to enforce policy and compliance requirements. This requires basic **security features** as well as role-based access control to enforce a hierarchy of permissions to prevent insecure actions on pipelines or destinations.

Finally, scalability requires programmatic control. Programmatic control reduces the risk of human error and can be ensured through an API. Another aspect includes automated user provisioning using System for Cross-domain Identity Management (SCIM) such as Okta or Azure AD, allowing an organization to quickly and safely onboard many users and manage the user lifecycle journey.

Data democratization requires resolving the conflicting needs of data consumers, data producers and security and legal teams, allowing organizations to safely provide widespread self-service analytics.

CUSTOMER SPOTLIGHT

[Read the full customer story →](#)



Autodesk, one of the world's premier industrial software companies, leveraged Fivetran to scale access to data-as-a-service to over 13,000 employees. The Analytics Data Platform at Autodesk is the go-to center for teams to understand critical product usage – everything from subscriptions to **customer behavior** to the health of the business. But with their previous architecture, accessing data for analytics was complicated and fragile.

“We couldn't even come close to enabling self-service with any of our previous tools and processes,” said Mark Kidwell, Chief Data Architect of Data Platforms and Services. “We had a central ingest team that solely took ingest requests – they were the only team that could do that work.”

To enable self-service data to more than 13,000 employees across the enterprise, Mark knew the analytic data sets would need to be easily queried. Mark looked to modern data tools, including Fivetran, **Snowflake** and **dbt**, to revamp the ingestion, transformation and orchestration.

“It's as simple as choosing between an unmanaged, low-level pipeline that breaks, or a pipeline [like Fivetran] that handles schema evolution, back-fill and other cases that we know we can't support.”

— MARK KIDWELL, CHIEF DATA ARCHITECT OF DATA PLATFORMS AND SERVICES

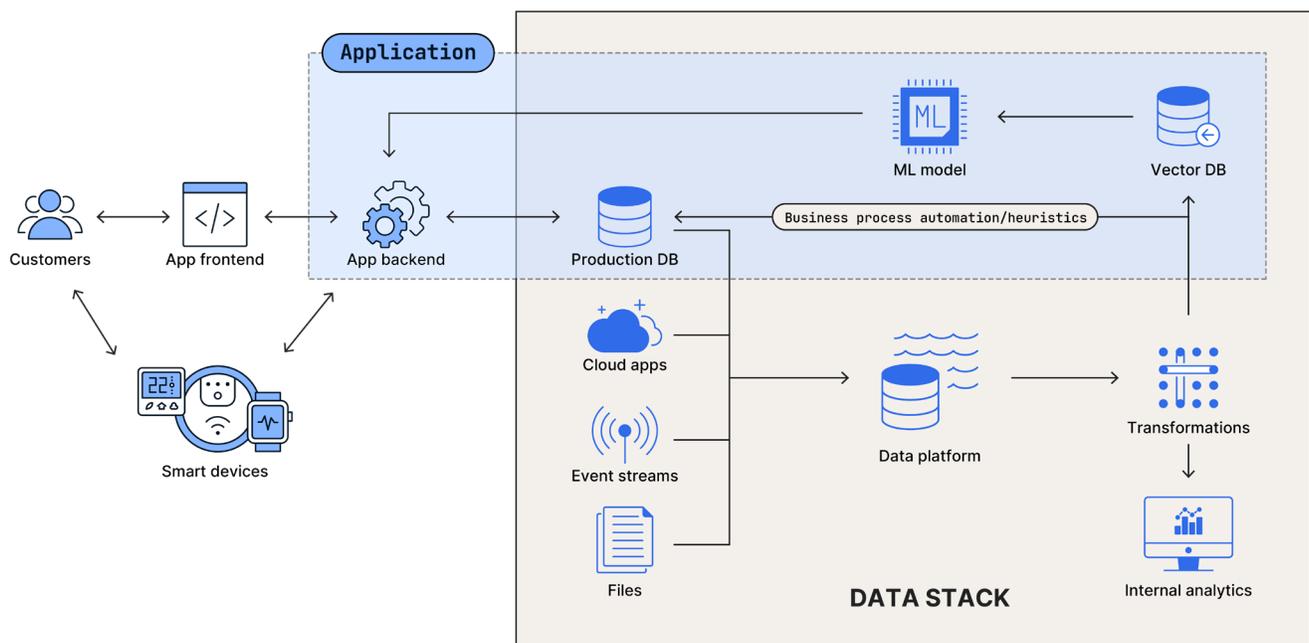
This move towards a modern data stack was a large business effort involving multiple business units, including two construction product acquisitions: **BuildingConnected** and **BIM 360**. The Analytics Data Platform team moved all of its business-critical data to run on Snowflake.

The Analytics Data Platform team quickly and securely onboarded teams and users to Fivetran to add their own connections and access pipelines and destinations for analytics. Enterprise **data governance** features like team and connector permissions, custom roles and team access, and single sign-on enabled the Autodesk team to scalably enforce platform security with its large user base.

4. Building data solutions

Data centralization facilitates the creation of a wide range of novel products and services. At Fivetran, we often liken data to electricity – an enabling technology with unlimited potential for innovation. There are three ways to think about data solutions:

- 1 **Enterprise pipeline management** involves programmatic management of data assets at a massive scale and complexity. Data travels in a dizzying range of directions for both analytical and operational purposes.
- 2 **Analytics products** are derived from data and can range from dashboards and reports all the way to machine learning and generative AI.
- 3 **Data sharing systems** make real-time data available externally to customers and third parties.



Why build data solutions?

Data can be monetized as products ranging from views of key metrics to recommendations, automated decisions and artificially intelligent models and agents.

Data solution challenges

Enterprises face several challenges related to building data solutions.

- 1 **Integrating technologies** can be difficult, as enterprises use dozens or hundreds of data sources, tools and platforms.
- 2 Data solutions may involve handling data on behalf of customers, which requires **onboarding customer data** that an enterprise doesn't produce or own.
- 3 Data solutions must simultaneously **ensure access and compliance** so that valuable data is only exposed to the appropriate parties.
- 4 In order to **monetize data assets**, an enterprise needs some mechanism for pricing and selling data, requiring visibility into the unit economics of the data.

How to solve data solution challenges

Each challenge can be addressed with the right technology.

The ability to integrate disparate technologies largely depends on a data movement platform with the following characteristics:

- ◆ It is compatible with a wide range of vendors and solutions, including orchestration technologies for custom workflows.
- ◆ It supports a wide range of data sources and destinations that can support the appropriate use cases.
- ◆ It features an API that enables large-scale, programmatic management of data connectors and users.

Access and compliance depend on data governance and security features. Some we have previously discussed, such as various certifications and role-based access control. Another important security feature is localized data residency. It is also important for a data platform to avoid commingling of customers' data through process isolation and multitenancy. Finally, features that support visibility into the unit economics of data include metadata APIs and the programmatic ingestion of logs. These enable a user to observe consumption, spend and the general health of data connectors.



Leading luxury retailer **Saks** used Fivetran, Snowflake and dbt to replace a legacy data infrastructure with a modern data stack in a matter of months. Most importantly, automated data integration has enabled three important data solutions:

- 1. Elevating the customer experience:** Fivetran provides access to near real-time data that enables Saks to more efficiently optimize the use of data from the full customer journey, which, in turn, enables the high-touch and personalized experience that customers expect when shopping for luxury fashion while ensuring security.
- 2. Improving customer service through experimentation with large language models:** With Fivetran, Saks can easily ingest multiple new data sources and feed them into its LLM to determine if Saks should integrate it into its operations. For example, Saks utilizes Fivetran in the contact center to ingest data from natural language processing tools. Saks then uses its LLM to review agent-customer interactions and create call summaries with actionable insights, leading to improved customer experiences.
- 3. Productizing data to power new applications:** Saks has created several data marts from its data ingestion. Customer, inventory and product data marts each have dedicated teams of data analytics engineers, data scientists and business intelligence analysts supporting dozens of use cases. The Fivetran-powered data marts have enabled Saks to advance from basic reporting dashboards to powering internal applications, as well as providing near real-time, actionable insights for business leaders across the organization.

By deploying a modern data stack, Saks was able to rebuild its data approach from the ground up, empowering its data engineers with new levels of efficiency. With Fivetran, Saks has been able to scale its data infrastructure, ingesting data from about 20 databases, allowing its team of data analysts and scientists to devote time to data as a product – rather than managing data movement – to drive the business.

A roadmap for innovation and data-driven value

The four use cases we have discussed are all essential elements of a cohesive strategy for digital transformation. Each of these elements can be addressed using a modern data movement platform that leverages automation, security, governance and extensibility features to provide reliable, secure and scalable data integration. With continuing developments in artificial intelligence and machine learning, the most innovative uses for data have yet to be invented. In the meantime, your organization stands to benefit from following a roadmap predicated on progressing through the four data use cases.

[Sign up for a 14-day free trial of Fivetran](#)



Fivetran, the global leader in data movement, helps customers use their data to power everything from AI applications and ML models, to predictive analytics and operational workloads. The Fivetran platform reliably and securely centralizes data from hundreds of SaaS applications and databases into any cloud destination — whether deployed on-premises, in the cloud or in a hybrid environment. Thousands of global brands, including Autodesk, Condé Nast, JetBlue and Morgan Stanley, trust Fivetran to move their most valuable data assets to fuel analytics, drive operational efficiencies and power innovation.

For more info, visit [Fivetran.com](https://fivetran.com).



[Start your free trial](#)

©2024 Fivetran Inc.