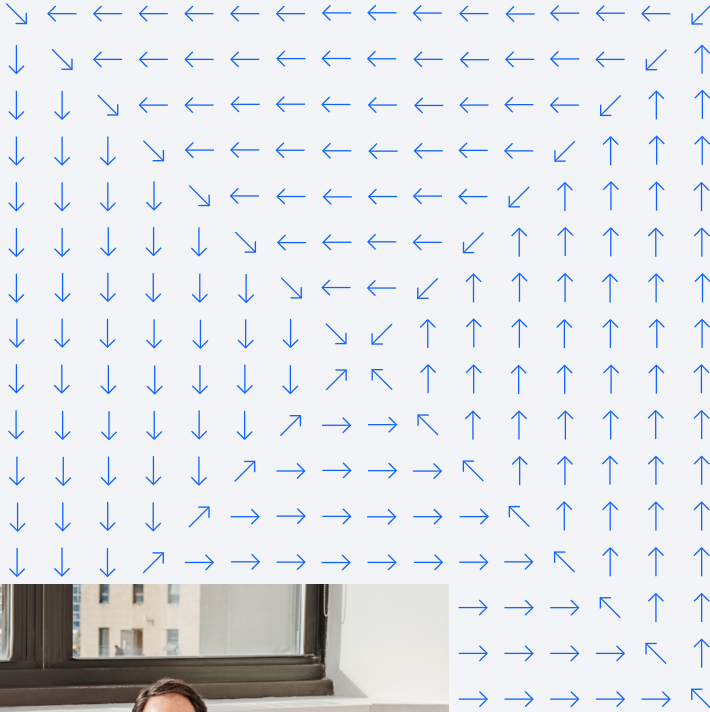




Data governance
and privacy for
↔ data-driven
organizations



Contents



01

Introduction—a data fabric approach to governance and privacy

02

Why establish automated data governance and privacy?

03

The building blocks of governance and privacy

04

Data fabric—a holistic approach

05

Data governance and privacy success story

06

Consider these components

07

Create your ideal governance and privacy solution

01

Introduction— a data fabric approach to governance and privacy

Data fabric is an architectural approach that helps ensure quality data can be accessed by the right people at the right time. A data fabric architecture supports all data and analytics use cases. Consequently, deploying this architecture requires good, automated governance capabilities to support varying scenarios such as reporting, dashboards, analysis, or automated data flow in support of operational functions.

Data governance is a foundational capability of a data fabric that optimizes data consumption and use in an organization. In addition to providing a strong foundation for multicloud data integration, 360-degree customer intelligence and trustworthy AI, the data governance and privacy capability of a data fabric strengthens compliance with automated governance controls, no matter where data resides.

Central governance applies the right level of data governance across the data estate, making quality data easier to find for those who should have access to it, while allowing sensitive data to remain hidden. Centrally enforcing governance across a hybrid and multicloud environment simplifies data access with appropriate security, privacy, lifecycle management and adherence to ethics.

Having insights into your business and customers is a competitive advantage. The Forrester Analytics Business Technographics® Data and Analytics Survey, 2020, found that advanced insights-driven businesses are more likely to have a data governance strategy that involves defining, executing, training, and overseeing compliance than beginner and intermediate firms. They're also more likely to have an executive in charge of their data governance and to use AI to crowdsource and embed data stewardship in everyday data engagements.¹

Strong privacy parameters help increase readiness for compliance and data protection anywhere, on-premises or across clouds. They allow businesses to understand their data and adhere to regulations that drive privacy requirements. The automated governance capabilities ensure the required level of privacy is enforced at the point of access no matter where the data resides.

In this guide, we'll look at the most common governance and privacy challenges modern organizations face, the building blocks of an effective solution or approach, and the technology components you'll need to build an automated, integrated data governance and privacy layer across critical data assets that enable your most strategic business outcomes. We'll also provide helpful resources such as a [data governance and privacy trial](#).

02

Why establish automated data governance and privacy?

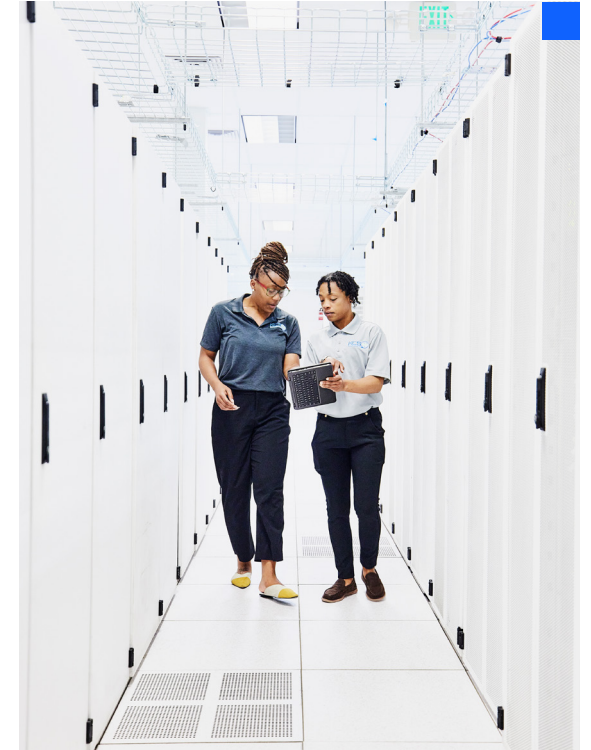
As organizations strive to establish cultures of data-driven decision making, a foundational requirement to drive data consumption is embedding automated data governance to establish data ownership, simplify data access, ascertain data quality, ensure appropriate knowledge of the data, and enforce the required privacy and security. Satisfying the outlined requirements is especially critical in dynamic regulatory environments. Such an approach allows organizations to deal with challenges such as:

The need to know your data

Establishing the source and ownership of data is an elementary step in using the data. The foundation of a well understood glossary of metadata helps users better understand the meaning of the data and determine how to use it.

The need to trust your data

In addition to understanding the data, users need to have confidence that the quality of the data is high enough to drive desired business outcomes. Data quality extends to having confidence in the source and a full understanding of the associated content and usability.



02

Why establish automated data governance and privacy?

The need to improve data access

Secure data sharing is a crucial factor when multiple teams require access to enterprise data. That data must be traceable and only visible to those who are authorized to use it. Yet 7 in 10 organizations are unable to secure data that moves across multiple cloud and on-premises environments.²

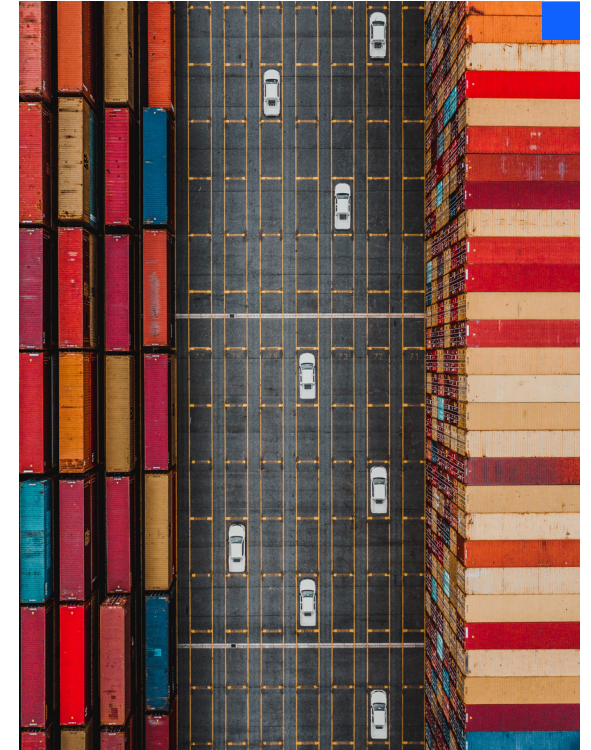
The data governance and privacy aspect of a data fabric rapidly delivers data into the hands of those who need it. The self-service capability enables the organization as a whole to find appropriate data quicker and spend more time using that data to provide tangible insights. Without being able to ensure centrally managed governance across their environment, teams hesitate to share data between business units, deepening silos. This forces IT teams to protect and ensure each data repository individually, spinning up their own repositories—shadow IT—leading to more complexity.

The need for data privacy at scale

A data fabric's governance and privacy capabilities keep data secure by applying compliance and data protection through a unified, consistent approach across an organization. Additionally, data policy enforcement at the granular level can be largely automated.

The risks of non-compliance (such as legal penalties, loss of customer trust, and loss of reputation) are real. More than 60 jurisdictions around the world have enacted or proposed privacy and data protection laws, and by 2023 more than 80% of companies worldwide will face at least one privacy-focused data protection regulation.³

Rather than responding to each challenge individually, a proactive approach to privacy and data protection is an opportunity for organizations to build customer trust. But to do it, data leaders need to build a holistic privacy program across the organization.



02

Why establish automated data governance and privacy?

The need to maintain data quality standards across the organization

Only 20% of business executives completely trust the data they get.⁴ Every year, poor data quality costs organizations an average \$12.9 million, according to a recent Gartner report. Gartner predicts that by 2022, 70% of organizations will rigorously track data quality levels via metrics, improving it by 60% to significantly reduce operational risks and costs.⁵

For all users throughout an organization to fully understand and have confidence in the data they are about to use, a data governance foundation of business definitions and metadata is essential. Trusting the data is predicated on establishing that the quality of the data is appropriate. Organizations can automate this process through metadata-driven automatic data profiling, classification and business term assignment.

The need for data lineage and traceability

Once analytics, operational, line of business or AI teams have built and deployed data products (such as dashboards, reports, and machine learning models), they need to be able to look back and see where the data product came from. For auditability and compliance use cases (often in regulated industries), an analytics team may be required to show all the steps taken in the life of the data as it has been transformed from the transactional system where it was originally created into its final form as it is used to support business decision-making. And for end users, being able to see the data sources and transformations can save a great deal of time as they build their own customized version of the dashboard.

The need to facilitate data consumption

To take advantage of the innovative and disruptive power of data, enterprises need to enable self-service data consumption. The ability to simplify data access and consumption is predicated on a robust framework and architecture that ensures data users in an organization can easily find and use the right data with a rich and metadata-driven index of cataloged assets. Data governance and privacy proactively enables enterprises to satisfy the need to drive innovation and meet business outcomes.

EU data protection authorities have handed out a total of \$1.2 billion in fines over breaches of the bloc's GDPR law since Jan. 28, 2021, according to law firm DLA Piper.⁶

03

The building blocks of governance and privacy

Ultimately, the goal of governance is knowing where data comes from, what it is, who can access it and when it should be retired. Several key technology building blocks exist to meet the need to integrate and improve data privacy, access, quality and traceability for all the data in an organization.

Let's look at what you'll need.

#1

Augmented data cataloging

A data catalog can help users easily find and use the right data with a rich and metadata-driven index of cataloged assets. The quality of your data determines how confidently you can act on insights. If low quality data enters your AI models, it could lead to inaccurate, noncompliant or discriminatory results. Low quality data also impacts operational and analytical use cases. Getting the best insights means being able to access data that is fresh, clean and relevant, with a consistent taxonomy.

[Read "A comprehensive guide for the modern data catalog" →](#)

#2

Automated metadata generation

Metadata tracks the origin, privacy level, age and potential uses of your data. Manually generating metadata is cumbersome, but with machine learning, data can be automatically tagged with metadata to mitigate human error and dark data. Automatic tagging of the metadata allows for policy enforcement at the point of access, so that more sensitive data can be used in a nonidentifiable and compliant way. In addition, metadata is used to establish a common vocabulary of business terms that provide context to data and to link data from different sources. This context adds semantic meaning to data so that it becomes more findable, usable and consistent within the organization, a key factor when seeking data for analytics and AI.

03

The building blocks of governance and privacy

#3

Automated data access and lineage

Data lineage shows, traces and analyzes how data is moved and consumed across all your applications and data sources. Knowing where data comes from is useful not only for compliance reporting but also for building trustworthy and explainable AI models. And it can be automated without complicating access. With restrictions built directly into access points, only the data users are authorized to access will be visible. Additionally, sensitive data can be dynamically masked so that models and data sets can be shared without exposing private data to unauthorized users. This clarity around what data can and can't be used supports self-service data demands and allows organizations to be nimble in responding to line of business needs.

#4

Data virtualization

Data virtualization connects data across all locations and makes the disparate data sources appear as a single database. This helps you ensure compliant access to the data through governed data access, regardless of where it lives, without movement. Using the single virtualized governed layer, user access to data is defined in one place instead of at each source, reducing complexity of access management.

#5

Reporting and auditing

Enterprises must comply with a wide variety of changing regulations that differ according to geography, industry and data type. They need to be broken down into a catalog of requirements with a clear set of actions that businesses must take. Regulatory information should be automatically ingested, deduplicated, and applied to workflows.

The secret to harmonizing all these data privacy and governance needs with business opportunity is aligning the technology components with a global data strategy and an open and holistic architecture.

04

Data fabric—a holistic approach

To harness data for insights and business growth—and ultimately create a data-driven culture—you need a holistic approach to data architecture and strategy that is efficient and doesn't involve manually patching together many solutions. This is why many organizations are adopting a data fabric.

A **data fabric** is an architectural approach to simplify data access in an organization to facilitate self-service data consumption. This architecture is agnostic to data environments, processes, utility and geography, all while integrating end-to-end data management capabilities. A data fabric automates data discovery, governance and consumption, enabling enterprises to use data to maximize their value chain.

With a data fabric, enterprises elevate the value of their data by providing the right data at the right time regardless of where it resides. It brings together capabilities like those listed previously as part of a unified architecture,

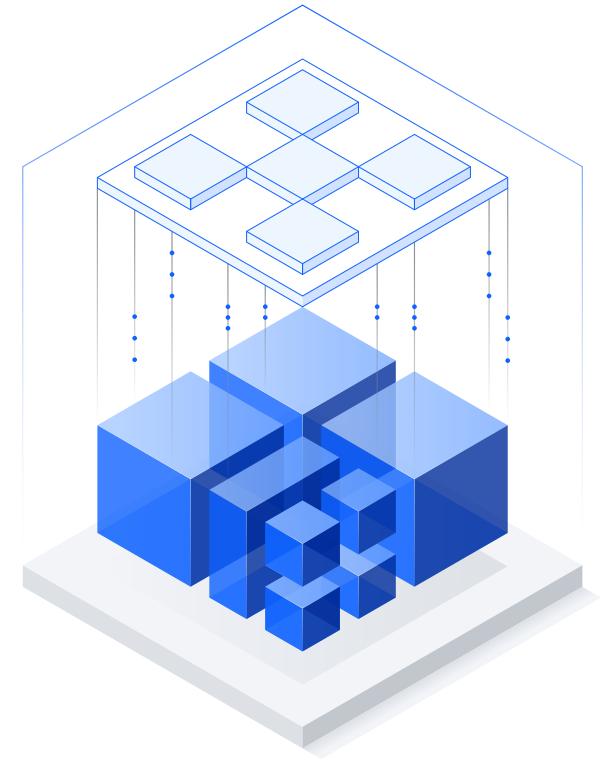
avoiding the cost and complexity of integrating a plethora of point solutions. Instead of a fragmented group of products that have been stitched together, a data fabric offers a single, holistic solution that is built to work seamlessly.

In addition, a data fabric can address three separate use cases beyond data governance and privacy. These include multicloud data integration, 360-degree views of customers, and MLOps and trustworthy AI (covered in separate ebooks).



A data fabric can elevate the value of existing data by providing access to the right governed data at the right time regardless of where it resides.

[Get a more detailed view →](#)



By 2024, data fabric deployments will quadruple efficiency in data utilization while cutting human-driven data management tasks in half.⁷

05

Data governance and privacy success story

Financial services: ING ↻

ING is a Dutch bank with over 57,000 employees serving around 39.3 million customers, corporate clients and financial institutions in over 40 countries. To bring his vision of data governance to life at ING, Ferd Scheepers, ING's Chief Architect, wanted to implement a data fabric approach in the company's hybrid cloud environment.

ING needed to govern its data in the cloud consistently with its on-premises environment. As the data leader, Scheepers had specific goals:

- Empower ING's data citizens with fast and simple access to governed data and toolsets
- Ensure strong governance and privacy parameters across a complex global ecosystem
- Comply with business policy and multi-jurisdiction regulations with changing requirements

ING created a data fabric solution to help implement a single corporate operating model and streamline data management and applications across all operational countries.

- ING sought an approach to orchestrate data governance that could scale for a global operation, reducing friction for data scientists.
- IBM and ING piloted a common, smart metadata tier to discover, govern, and secure data more easily. This metadata tier enabled users (e.g., business analysts, data scientists) to collaborate with data stewards who declaratively write data access policies.
- IBM and ING leveraged an orchestrator that ensures proper data protection, privacy, and compliance policies are consistently enforced.

It runs across an open hybrid cloud environment that adapts to ING's multi-platform, heterogeneous landscape. Applying data virtualization across existing on-premises investments, it removes data silos, enabling just-in-time access to the right data across any cloud and on-premises, at the optimum cost, with the appropriate level of governance.

Using their data fabric, ING can provide a consistent user experience to increase collaboration, streamline application management, and optimize licensing and IT costs.

[Read the case study →](#)

06

Consider these components

An effective data governance and privacy strategy is dependent on a technology stack designed to gain end-to-end governance, deliver quality data, and ultimately accelerate collaboration. In an enterprise, the value of data governance is amplified when this capability is integrated with data integration, providing a comprehensive view of an entity and enabling maximum data utilization to drive business outcomes.

As part of a modern data fabric, data governance and privacy creates an end-to-end user experience rooted in metadata and active policy management that allows users to view, access, manipulate and analyze data without the need to understand its physical format or location, and without having to move or copy it.

The technology components of the IBM data fabric approach allow companies to automatically apply industry-specific regulatory policies and rules to their data assets, securing them across the enterprise, with:

- An AI-augmented data catalog allowing business users to easily understand, collaborate, enrich and access the right data
- A metadata and governance layer for all data, analytics, and AI initiatives that increases visibility and collaboration on any cloud
- The ability to dynamically and consistently mask data at a user defined granular level
- The ability to create anonymized training data and test sets while maintaining data integrity



IBM Cloud Pak for Data

IBM Cloud Pak® for Data is a platform built specifically with a data fabric architecture in mind to predict outcomes faster and allow you to collect, organize and analyze your data, no matter where it may reside. The platform thus helps to improve productivity and reduce complexity by building a data fabric that connects siloed data distributed across a hybrid cloud landscape.

[Learn more about IBM Cloud Pak for Data →](#)

06

Consider these components



IBM Watson Knowledge Catalog

IBM Watson® Knowledge Catalog provides intelligent cataloging, with automated metadata collection and policy management to ensure the details of a model are automatically collected and stored for maximum transparency and repeatability. It ensures that models are impartial, address bias, are explainable and adapt to changing model parameters.

[Learn more about IBM Watson Knowledge Catalog →](#)



IBM Watson Query

Applying sweeping governance rules across data lakes, databases, and warehouses is time consuming, and often leaves users with long delays to get access to the right data. Watson Query enforces governance policies when data is accessed across multiple sources, quickly providing data to your end applications through one view without manual changes, data movement or replication.

[Learn more about IBM Watson Query →](#)

07

Create your ideal governance and privacy solution

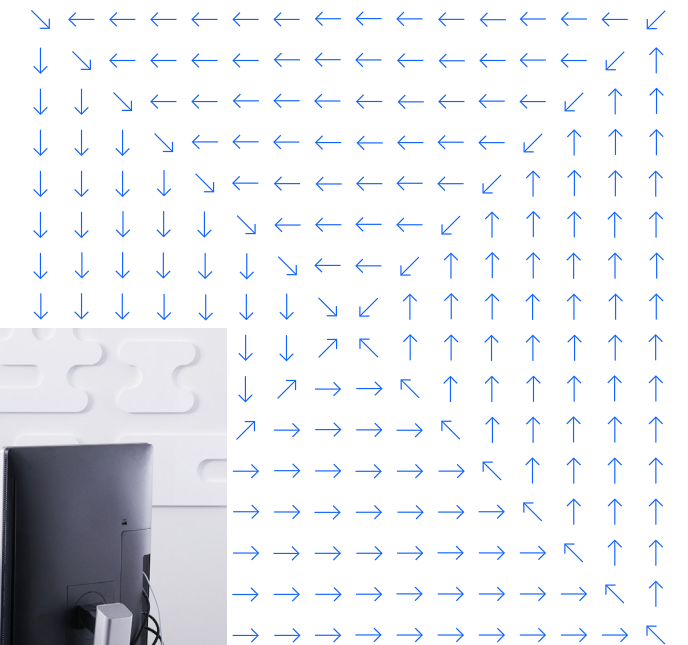
If you're ready to embrace a unified strategy and architecture approach to improve the accessibility, security and compliance of your data of all types and sources, we encourage you to take advantage of a few resources.

Foremost is the free [data governance and privacy trial](#), which will provide hands on experience with a data fabric designed for this use case. Second, review the information on the [IBM data governance and privacy website](#) to learn even more about the benefits. And, finally, reach out to one of our experts either by [scheduling a time online](#), talking with your IBM representative or reaching out to one of our business partners.

Learn more about the role of a data fabric in building a data strategy and architecture with [The Data Differentiator](#).

Check out the other three data fabric use case ebooks:

- [Multicloud Data Integration](#)
- [Customer 360](#)
- [MLOps and trustworthy AI](#)





© Copyright IBM Corporation 2022

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
October 2022

IBM, the IBM logo, IBM Cloud Pak, IBM Watson, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

- 01 Break Through Data Governance Fatigue, A Framework for Effectiveness and Sustainability, Forrester, 25 February 2021.
- 02 Getting started with zero trust security: A guide for building cyber resilience, IBM Institute for Business Value, 22 July 2021.
- 03 Hype Cycle™ for Data Privacy, Gartner, 13 July 2021.
- 04 Ryan Brown, Fines for breaches of EU privacy law spike sevenfold to \$1.2 billion, as Big Tech bears the brunt, CNBC, 18 January 2022.
- 05 How to Improve Your Data Quality, Gartner, 14 July, 2021.
- 06 Fines for breaches of EU privacy law spike sevenfold to \$1.2 billion, as Big Tech bears the brunt, CNBC, 18 January 2022.
- 07 Top Strategic Technology Trends for 2022: Data Fabric, Gartner, 18 October, 2021.