

The journey to AI and business- ready data begins with information architecture

Explore a trusted analytics
foundation with governance
and cataloging at the core

Contents

- 2 Introduction
- 3 The building blocks of a trusted analytics foundation
- 4 Machine learning accelerates governance
- 5 A single foundation for many purposes
- 6 Conclusion

Key takeaways

- Compliance can encourage organizations to implement ongoing and beneficial data governance strategies.
- Machine learning automates governance and integration initiatives on a large scale, overcoming the limitations of human ability and the difficulties presented by large volumes of data.
- Data governance is effective on premises and in multicloud environments.

Introduction

The quantity and variety of data is increasing rapidly for enterprises of all kinds. In multicloud environments, a range of data sources is exponentially increasing the stream of incoming information—from the Internet of Things and social media, to mobile devices, virtual reality implementations and optical tracking. While organizations are readily investing in artificial intelligence (AI), most haven't done due diligence to understand their data or ensure the quality of data needed to benefit from AI solutions. In many organizations, data is inaccessible, unreliable, or non-compliant with data privacy and protection rules.

Global regulations such as the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA) and Brazil's Lei Geral de Proteção de Dados (LPGD) focus on personal customer and employee data. These types of regulations offer organizations an opportunity to transform and create new data-led business models, despite the severe penalties that may result from non-compliance, which can slow productivity or damage brand value. To meet privacy obligations and protect personal information, organizations must first discover and classify their various types of data. Businesses that struggle to gather or properly use customer data can experience severe problems. To address this challenge, organizations are implementing governed [information architectures](#) (IAs) that acknowledge regulations while continuing to support data-driven organizational performance and innovation.

The building blocks of a trusted analytics foundation

Data privacy regulations don't need to be seen as simply an obligation; they're also an opportunity to modernize data infrastructures offers Taking this approach can encourage organizations to implement data governance strategies that generate new business models and lead to data-driven insights. So-called "DataOps" initiatives orchestrate the structured and unstructured data in public and private clouds. Implementing DataOps has significant value for compliance efforts, but its value also affects other areas of an organization, particularly the governing of AI models for data scientists.

When an organization uses data governance to orchestrate and manage its data, users can have confidence the data came from a quality source. They know how the data is being used across the organization and they know how it will enhance any analytics project. No matter how advanced the analytic tools might be, a successful analytics initiative requires trusted data to work effectively. And the benefits of trusted, business-ready data seem limitless. Analytics can suggest new product designs and marketing programs, and improve sales, supply-chain or customer-service initiatives. Analytics can even uncover operational inefficiencies that can be eliminated to increase organizational agility and boost bottom-line revenue. Data and AI governance in your organization rests upon five building blocks, described in more detail below:

Data discovery and quality

Organizations can sometimes be unaware of the large amounts of data stored within their business. The first step in data governance is to inventory organizational data. Start by focusing on data sets in a specific project, then expand to other business cases for broader organizational coverage. Data that's redundant, obsolete or trivial (ROT) is not only costly to store and manage, but also clutters decision making and operations. It can also make compliance more difficult and thwart analytics efforts. Data must meet and maintain certain quality measures to make downstream usage successful.

Cataloging

Once data is discovered and profiled, it's cataloged using metadata tags to identify data types, usage, ownership, data lineage, and more. Because companies in certain industries share common needs, pre-built industry models can expedite the cataloging process by using readily available business terms and taxonomy. With advancements in machine learning (ML), business terms can be automatically mapped to build an enterprise catalog in a matter of hours. This cataloging foundation enables organizations to govern their AI models, notebooks and other data sources, creating a central library for organizational knowledge. This foundation is a resource for many data users in the organization including data engineers, data stewards, and line-of-business users such as analysts, data scientists and marketers.

Data movement, transformation and synchronization

Data from multiple sources can be easily integrated, transformed and shared with other systems as needed, physically or virtually. This process brings structured and unstructured data together and allows integration with open technologies such as Apache Atlas and Hadoop. Creating automated data flow and synchronization helps ensure that the most recent data is available in data lakes, data warehouses, data marts and point-of-impact solutions. As data quantities increase, replication technology supports large volumes with low latency. In addition, organizations can use virtualization technology to access the required data without moving it.

Master data management

Business users with self-service capabilities are approaching or exceeding the analytic output of professional data scientists, so it's essential for organizations to rely on an organization-wide, comprehensive, trusted and unified view of critical entities including customers, products and accounts. Modern master data management (MDM) implementations come with analytical graph-based exploration, a highly accurate matching engine, a data-first approach in selecting matching algorithms, and stewardship processes powered by machine learning. In addition, MDM solutions feature agile self-service access, governance tools and user-friendly dashboard capabilities.

Data privacy and protection

Organizations must proactively secure and protect their strategic and sensitive information assets. Management of the data lifecycle goes from creation to disposal, using such practices as records management, litigation and archival storage. By applying cognitive learning to an organization's documents, history and other context, risks can now be automatically identified.

Data that's governed both for business operations and compliance is therefore business-ready and can be readily used for any decision making, improvement or innovation.

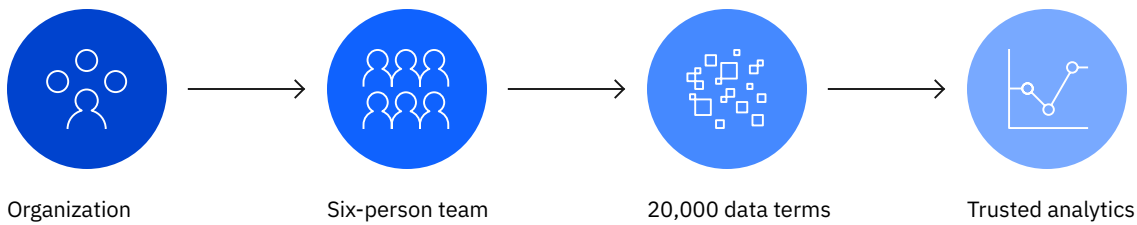
Machine learning accelerates governance

Thanks to recent technological advancements, machine learning can now augment human intelligence and complement the significant limitations of human ability. ML automates governance and integration initiatives on a large scale, overcoming the difficulties presented by large volumes of data, leading to sound data governance across the enterprise. For example, if an organization has 20,000 data terms, it typically takes six months for a six-person team to manually classify the terms to drive analytics in a reliable and trusted way.

With ML, the same process can be completed in a few days or even hours, depending on the quantity of the data assets. This level of acceleration removes a costly burden from the governance process. Machine learning can make compliance obligations more manageable, and it paves the way for productive analytics initiatives.

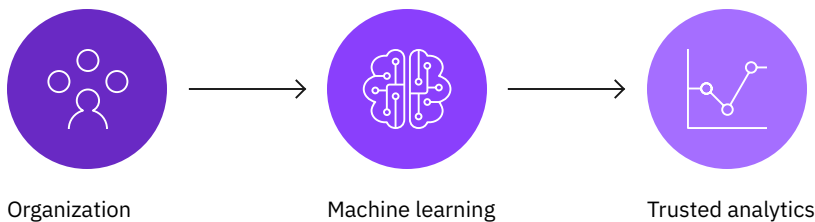
Manual classification

Six months to complete



Machine learning classification

Days or hours to complete



A single foundation for many purposes

When a governed data foundation exists it can be used across business units and throughout the organization, as described in the following common use examples.

Governed data lake

Many organizations in a wide variety of industries have undertaken big data predictive analytics projects. A key first step is to deposit large quantities of both structured and unstructured data in a data lake.

To achieve this, organizations have often used Hadoop or hand-coding solutions in their data lakes. But the absence of a governance structure paired with a lack of strategy for managing data standards, business terms, lineage, usage and quality can degrade the data lake into a so-called data swamp in which users neither understand nor trust their data. Organizations are increasingly realizing that their data lake needs a strategy and governance to be successful.

In a governed implementation, data in a data lake is mapped to business terms that are easy to understand by any data user and consistent across the organization. Having such data available for users can accelerate time to value and increase agility for virtually any self-service data science, data exploration or AI project. This access sets the foundation needed to support multicloud environments, on-premises architectures and a variety of datasources.

Application modernization

Organizations are making large investments in modernizing applications to increase efficiency, reduce costs and gain competitive advantages. Depending on the organization, app modernization ideas can manifest themselves in many ways. Top considerations include test data management, data virtualization and connectivity. Organizations now use agile methodology to test and develop concurrently with [virtual data access](#). They also rely on using flexible integration capabilities to connect business applications and data.

360-degree view of customer data

It's important that employees have trusted, up-to-date and accurate [single-view master information](#) with regard to customers, products or other entities. Erroneous or out-of-date information can damage customer interactions, erode confidence, lead to account turnover or increase supply chain costs. However, data that's sourced through DataOps processes can help customer interactions become ways to increase trust, brand loyalty and equity and increase supply chain agility.

Enterprise data warehouse optimization

Optimizing the architecture of an enterprise data warehouse(EDW) represents a dramatic upgrade to how data gets accessed, stored, prepared, governed and analyzed. One of the most [effective optimization](#) approaches is to offload extract, transform, load (ETL) jobs, data that's no longer used, and data that's required in exploratory models. This process not only reduces costs, it enables the data to be combined with other data types in environments such as governed data lakes for dynamic data exploration.

Regulatory compliance

A trusted analytics foundation empowers and [accelerates compliance](#) with regulatory mandates. Most importantly, the journey to personal data protection begins with defining what personal data is, so that an organization can discover what personal data it has. The foundational data catalog contains governing rules for data quality, enrichment and analysis, and policies for compliance.

Conclusion



As organizations go through digital transformation, business leaders are becoming aware of the benefits of governance across their data and AI models, whether on premises or in multicloud environments. By focusing on core governance practices, organizations are preparing their data and AI not only for analytical processing and insights, but also for compliance with the regulations they face. Although the data volumes are extensive, machine learning and artificial intelligence practices help augment human scale and intelligence in tasks such as data mapping, cataloging, matching large data volumes and sustaining data quality.

Business leaders with vision understand that taking the time to build a solid DataOps foundation will pay significant dividends today and in the near future. They realize their organizations will gain important advantages if they embrace governance as an enabler for business optimization, innovation and compliance across data and AI initiatives. It's critical to use solutions that encompass data operations management from creation to consumption. Streamlining these operations requires economies of scope, scale and sharing.

Learn more

IBM DataOps solutions can help you build a trusted analytics foundation to drive AI at scale.

Visit website



Explore more

Continue learning about cognitive data governance and how IBM solutions lead with machine learning and AI. Read the whitepaper.



© Copyright IBM Corporation 2021

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
September 2021

IBM, and the IBM logo are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on ibm.com/trademark.

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.

It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. 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.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.