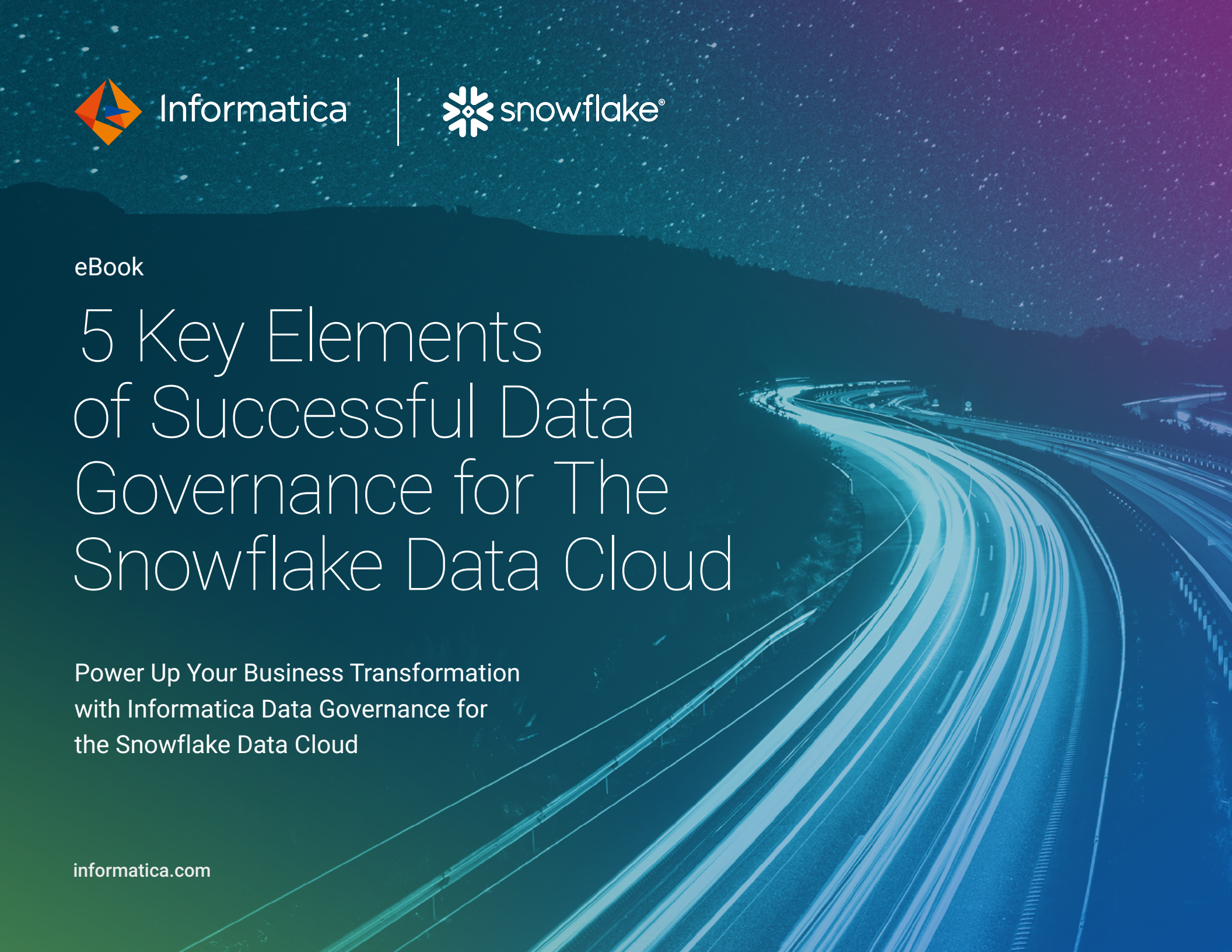
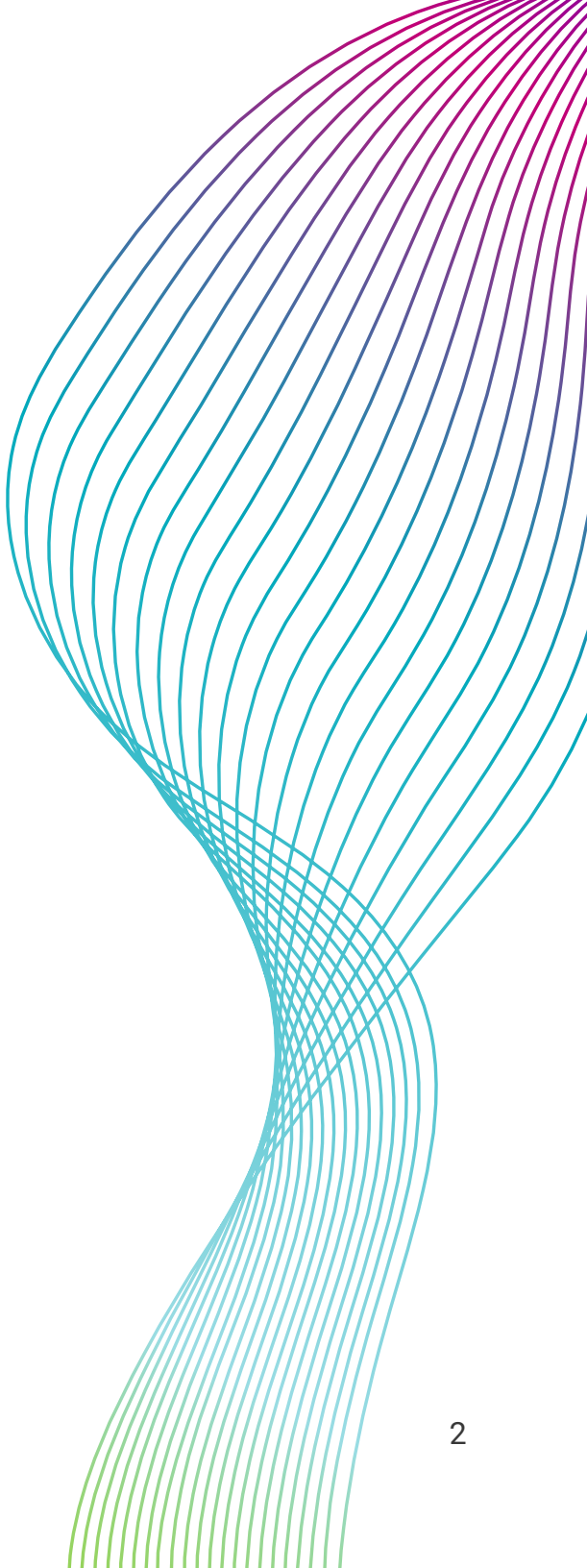

Informatica
snowflake®
eBook
5 Key Elements of Successful Data Governance for The Snowflake Data Cloud
Power Up Your Business Transformation with Informatica Data Governance for the Snowflake Data Cloud
informatica.com

Contents



Data Governance in a Multi-Cloud World

Organizations are moving their data to the cloud in droves. The benefits of successful cloud data transformation (such as agility, cost-effectiveness and scalability) are well-known and well-established. As the amount, types and speed of data being produced and collected increases, even highly cautious IT departments are shifting from using on-premises infrastructure to multi-cloud setups. The need to effectively manage and utilize large volumes of data at scale drives demand for cloud-native data platforms like the Snowflake Data Cloud.

But even if the benefits of cloud data transformation are readily apparent, achieving those benefits is easier said than done.

The Snowflake Data Cloud simplifies the process of centralizing data from different data stores and infrastructure types. It also promotes innovation by allowing self-service access and better concurrency for data democratization. However, there are technical and business-related complexities that must be addressed during the process.

A primary consideration is the question of data governance:

- How do you support a dramatic increase in data accessibility needs across the organization, enabling data consumers to derive value from data efficiently?
- How do you ensure complete visibility into data access and use and promote proper data handling across a complex, heterogeneous infrastructure?
- How do you maximize your investment to get the most from what Snowflake offers without introducing new risks and complexities?

In this eBook, we'll explore how to activate and operationalize an intelligent approach to data governance. We'll show you how to leverage the latest innovations in artificial intelligence (AI) and machine learning (ML)-powered automation to support the self-service and virtually unlimited concurrency enabled by Snowflake.

The Snowflake Data Cloud

Snowflake delivers the Data Cloud, a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency and performance. The Data Cloud allows organizations to easily unify and connect to a single copy of all their data. The result is a vast ecosystem of businesses and organizations that can effortlessly connect to their data and share and consume data and data services.

Data clouds are the future of enterprise data management because they break down the silos that limit data access, enabling everyone to work with a single version of the truth. They eliminate the duplication and versioning problems that undermine quality. Role-based access controls allow policy-administered security, and data can be safely and securely shared internally and with external constituents.

The Snowflake Data Cloud supports a wide range of workloads, including data warehousing, data lakes, data science and application development across multiple cloud providers and regions. It is built on cloud-native features and provides elasticity, secure collaboration and accountability controls.

Advantages of the Snowflake Data Cloud:

Data Democratization and Self-Service

- Allows for secure, governable collaboration across data from practically all sources.
- Democratizes analytics and accelerates time-to-value with faster data access for business users: Process data loading and unloading without hindering the performance of queries or other workloads that may be running, thanks to the separation between storage and compute.

Efficiency and Cost

- Offers usage-based "per-second" pricing for compute and storage, ensuring that you pay only for what you use.
- Allows flexibility in sizing warehouses up and down based on varying workload requirements for better utilization and lower costs.

Tools for Secure Data Transformation

- Robust data access and security rules allow you to determine who can access which data; you can trust that it is secure thanks to the highest industry-leading compliance standards and regulations.
- Secure data sharing and exchange mechanisms prevent the need to go through the cost or headache of manually creating subset databases or using static data-sharing methods to move data around.

The Challenges of Governing Data in the Cloud

Embracing the cloud for data analytics is a significant first step toward getting more value from your data. However, many organizations struggle to understand the following:

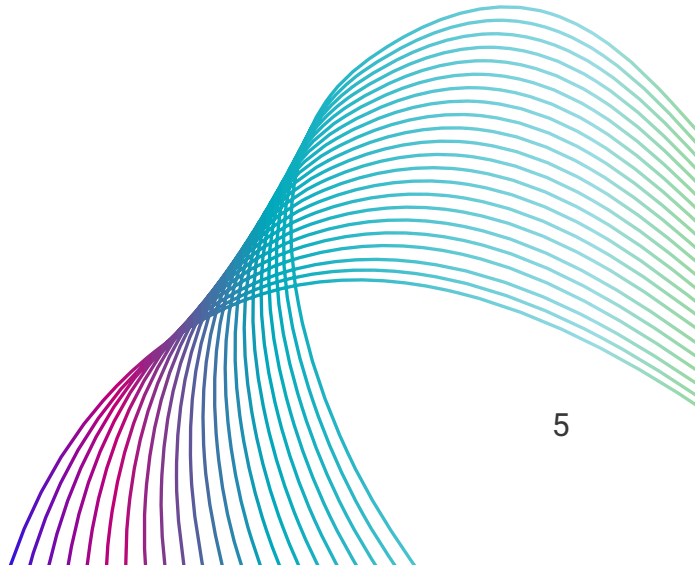
- Where their data originated
- Where it has moved
- Who is consuming it
- Its quality rating
- How and when it can be used appropriately
- Whether security and privacy controls are in place to help ensure data use with minimal risk exposure

This complexity makes it virtually impossible to deliver the right data to the right users for business-relevant purposes. An automated data governance solution would provide visibility and allow you to form a foundation for trusted data.

The primary challenges of governing data in the cloud include the following:

- **Ensuring appropriate access for self-service use**: Cloud platforms such as Snowflake offer the advantage of extending self-service data access to a wider range of users, including business users and external partners. However, without proper data governance policies in place to enforce appropriate access controls, managing data access can become difficult, leading to uncertainty and increased risk exposure. This can ultimately lower trust in the data and how it is utilized.
- **Managing the complexity of hybrid and multi-cloud data environments**: As the infrastructure underpinning your data sources becomes more complex and heterogeneous, so do the challenges of maintaining complete visibility and governing data pipelines. Finding relevant data, inventorying it and understanding whether it is fit for purpose becomes increasingly difficult as the amount and variety of data expand across your organization.
- **Maintaining data quality and trust**: When data quality is suspect or improper use occurs that compromises integrity, teams cannot confidently make data-driven decisions, and the data stores that inform analytics and ML models become perceived as unreliable. Once data stakeholders and data consumers no longer trust the data, rebuilding that trust is inevitably an uphill battle.

After taking the plunge and moving your data to the Snowflake Data Cloud, the next phase is to tackle these challenges, removing barriers to effective democratization. For that, you need a plan — a strategy built on the critical elements of successful data governance to automate and scale a trusted data foundation.



The 5 Elements of Successful Data Governance

You need a reliable and automated data governance framework to get the most out of the Snowflake Data Cloud and empower data consumers across your organization to find the data they need to generate value quickly. Not only can you lay a foundation of trust to deliver the data to those who need it for appropriate use that aligns with your organization's policies and business agenda. You can also accelerate data democratization and delivery by streamlining how you find, understand, trust and access high-quality, relevant data to generate value and achieve your business outcomes.

An effective platform to democratize data, build trust and deliver data intelligence for analytics insights, improved productivity and customer-centric value creation includes five key elements that support a complete enterprise data governance strategy:

1 Data cataloging

Maintaining an accurate inventory of data and metadata is critical to ensuring that your data management team can adequately contextualize your data. A **data catalog** assists with discovering, inventorying and defining what data your organization has available, automating the application of metadata for new insights and building a foundation for organizing data into tables and schemas to save time searching for assets across disparate sources. Related to cataloging, **data lineage** provides an understanding of data as its movement is traced across the organization for greater visibility.

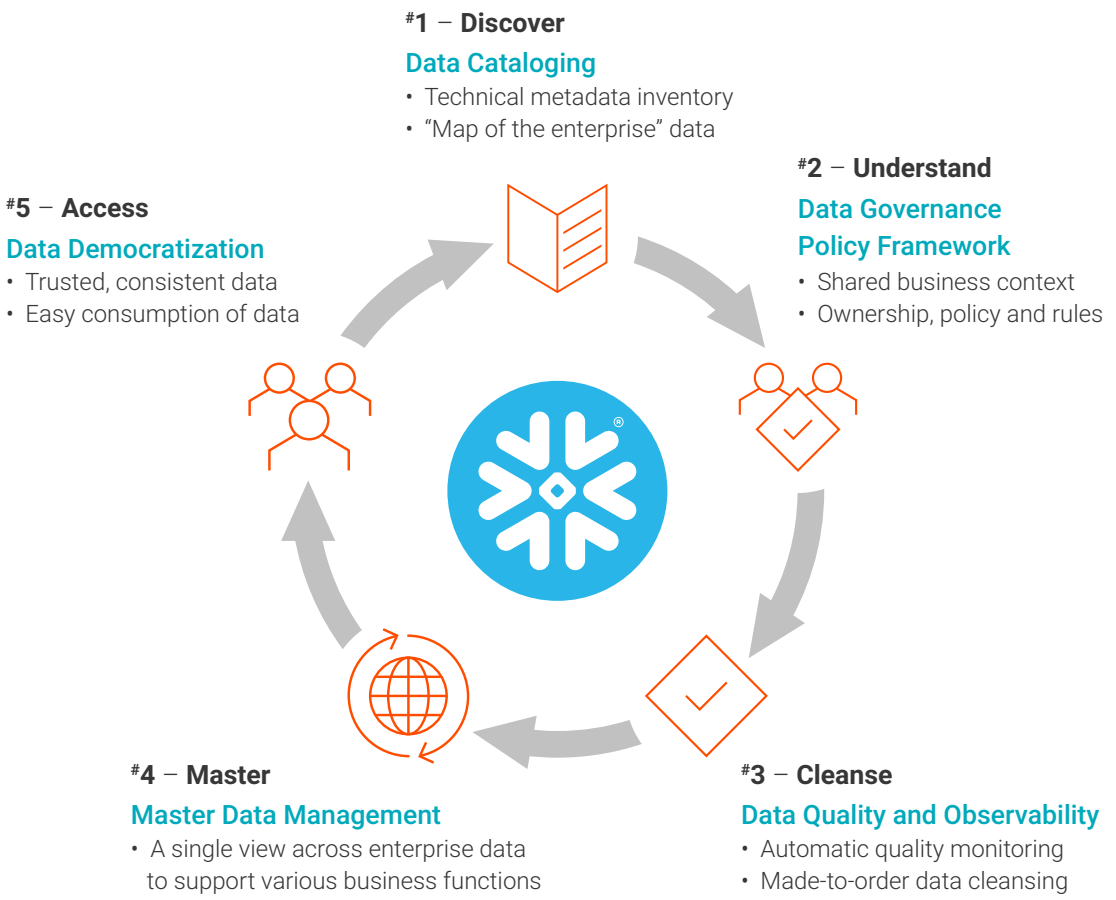
The 5 Elements of Successful Data Governance (continued)

2 Data governance policy framework

Automated tools that connect business and technical users to align on data access and use policies and agree on data purpose are fundamental for maximizing business efficiencies and productivity. Aligning policies also helps reduce risk exposure by operationalizing proper data handling processes and procedures, including applying privacy controls.

As data growth continues and organizations acquire more personally identifiable information (PII) about customers and grow a base of intellectual property (IP) through self-service data use and analytics, it's increasingly important to ensure that internal policies and regulatory compliance are enforced with reasonable controls in place to avoid violating customer trust or exposing confidential data in unintended ways.

Automated tools simplify appropriate access to reliable data to get it more quickly to those who need it to achieve business goals, while maintaining customer, business and shareholder trust.



The 5 Elements of Successful Data Governance (continued)

3 Data quality and observability

Organizations need the ability to identify **data quality** issues and data anomalies quickly. They also need to use **data observability** to measure compliance to help understand and improve the reliability and trustworthiness of their data.

Data quality and profiling capabilities help enable users to discover these issues and set data quality rules to take action. By implementing rules, organizations can refine their data through cleansing, parsing and verification processes. And with the right observability tools, organizations can monitor compliance and performance through various reporting such as threshold alerts and scorecards.

4 Master data management

Data accuracy is achieved when linking or merging data sources to create a master record that is normalized for reference. The master data record can then be considered a reliable and valid source of truth for further operations, such as providing a complete customer account that can be trusted for customer experience programs, clarity of financial transactions and more. Much like data quality and data privacy, **data mastering** contributes to the overall reliability and trustworthiness of data as fit and relevant for achieving desired business outcomes while minimizing friction and risk.

5 Data democratization

An emerging area of increased importance to fulfill the "last mile" of data access and delivery is **data democratization** and sharing through a data marketplace. By front-ending the data consumer with an easy-to-use data shopping experience, organizations can spend less time hunting for data, data sources and data owners across the business. They can instead accelerate recommendations for the datasets they need with automated delivery at their fingertips. A data marketplace is the missing link for connecting data consumers with the data intelligence they need for data-driven decision-making.

Why AI-Driven Data Governance Is Critical for Success

The five elements described in the previous section are necessary capabilities for operationalizing any successful data governance strategy — whether intended for traditional on-premises data warehouses and lakes or for supporting modern cloud-native solutions. Moreover, given the ever-increasing volume of enterprise data, fully understanding the data landscape is impossible with manual processes and procedures.

This is especially true for cloud-native platforms such as the Snowflake Data Cloud. Support for increased user concurrence, infrastructure variety and self-service access makes systematic data governance simultaneously more necessary and even more overwhelming. As a result, taking full advantage of AI and ML for automation is a must.

The latest data management solutions incorporate a range of automated metadata-driven AI and ML capabilities to make data governance-at-scale a reality, including the ability to:

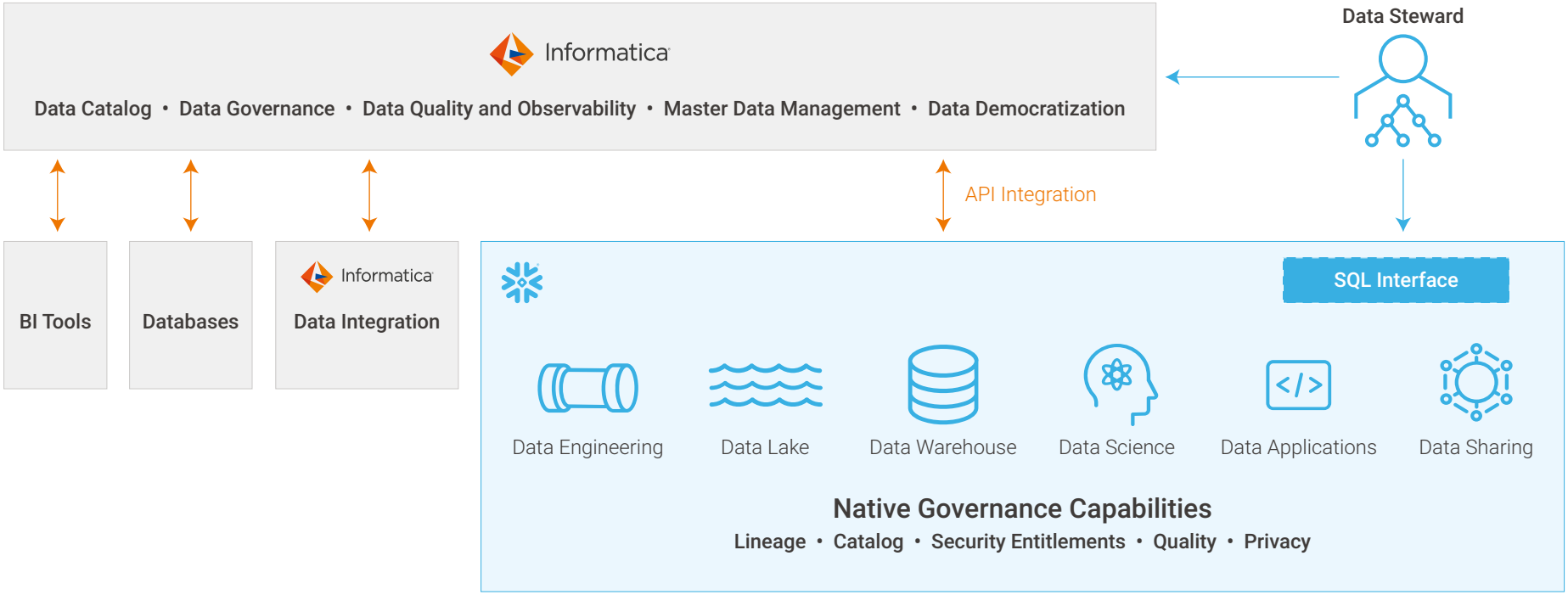
- Discover and auto-classify data to quickly understand its business context through metadata-driven insights, and demonstrate end-to-end data lineage for transparency that enables data stakeholders to understand data impact and take action.
- Automatically apply data quality rules to determine data reliability and trustworthiness for understanding fitness for purpose and appropriate use.
- Enable data observability to help identify data issues and anomalies to remediate and scale using trusted data.
- Enforce data privacy policies with rules for GDPR, CCPA and other mandates to protect data and expose it safely for analytics insights supporting business value creation.
- Eliminate labor-intensive manual and unreliable practices such as mastering for data curation. That way, data management teams can focus on higher-value activities while reducing errors.
- Improve consistency for business data visibility and reporting with transparency, reduce human errors to minimize risk and improve clarity for enterprise data consumption and auditing.

By integrating AI and ML-enabled capabilities across all key areas of data governance, you can scale your framework to keep up with expanding data volumes and new use cases enabled by the Snowflake Data Cloud.

AI-Powered Data Governance from Informatica

Informatica is the only vendor that can provide a broad range of end-to-end data governance capabilities through the Informatica Data Management Cloud (IDMC). These capabilities can help accelerate your cloud data workloads with a fully cloud-native solution that integrates seamlessly with the Snowflake Data Cloud to deliver data trust assurance.

With capabilities including data cataloging, mastering, quality and protection, Informatica's AI- and ML-powered solution enables enterprise-scale data governance to deliver consistent, trustworthy and compliant data in the Snowflake Data Cloud and beyond. Continue reading to learn why Informatica and the Snowflake Data Cloud are better together.



Why Informatica for the Snowflake Data Cloud?

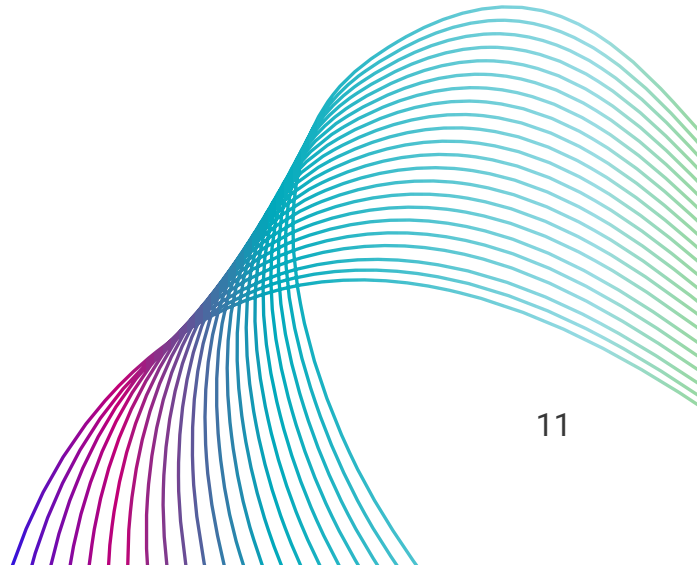
Cloud Data Governance and Catalog
Informatica Cloud Data Governance and Catalog, a service of IDMC, brings business and technical stakeholders together to collaborate on data policy and purpose. It enables rapid data discovery with powerful semantic search — empowering organizations to easily find the data they need across the Snowflake Data Cloud and other cloud and on-premises data sources. Users can profile data, identify its location, view its lineage and obtain additional useful information about data, such as ratings and reviews, at scale to better understand data assets and their value.

Cloud Data Governance and Catalog leverages AI/ML to help automatically discover and classify data, as well as to intelligently associate glossary terms to data. And this can help users quickly assess if data is relevant and appropriate for their needs. It can also help organizations manage and protect data assets per enterprise data policies and regulations.

Cloud Data Governance and Catalog also offers a single platform that allows users to govern policy creation, orchestration and execution across enterprise-level systems, which helps enterprises deliver data that is consistent and appropriate for use. Users can discover and associate Snowflake tags with their respective data elements, then enable push down to the Snowflake Data Cloud and apply privacy, quality and data sharing policies.

This IDMC service allows users to define and apply operational data governance policies in Snowflake without writing a single line of code or requiring SQL knowledge. It enables them to seamlessly deploy operational and policy-based data governance to provide an end-to-end automated “supply chain” for data.

Cloud Data Quality
Snowflake customers leverage Informatica Cloud Data Quality service on IDMC to centrally define data quality templates, continuously monitor Snowflake resources and remediate data quality issues. Cloud Data Quality helps ensure clean, complete, consistent and ready-to-use data. The service features standardization, matching, worldwide address cleansing and versatile data quality management and observability for analytics, AI and other data-intensive initiatives powered by Snowflake.



Why Informatica for the Snowflake Data Cloud? (continued)

Master Data Management
Informatica Intelligent Master Data Management can manage multiple domains of master data in a single SaaS solution. Preconfigured domain and industry content and low-/no-code configuration simplify administration and stewardship tasks — such as matching and merging — speeding deployment and streamlining ongoing management. AI-powered automation and modern user interfaces increase productivity and efficiency, and the ability to manage multiple domains in the Snowflake Cloud and beyond helps to ensure long-term return on investment.

With deep integration with the Snowflake Data Cloud, customers utilize Informatica Master Data Management services to connect and bring data from Snowflake's tables and other data sources to create a consolidated materialized view of domain entities (e.g., product, customer, etc.). This consolidated, high-fidelity, trusted master data is proliferated back to Snowflake for value-generating analytics and machine learning initiatives.

Cloud Data Masking
Informatica's data privacy solution helps you disposition and mask personal and sensitive data. Organizations can quickly implement a continuous, holistic process to operationalize data privacy that adapts to shifting priorities, evolving regulations and new privacy risks. This enables organizations to retain and enhance customer trust by continually securing sensitive data from critical data sources to help comply with evolving data security and privacy regulations as confidential data proliferates across the enterprise.

Cloud Data Marketplace
Trusted data for everyone, everywhere. The Informatica Cloud Data Marketplace for Snowflake makes data accessible and available to everyone, whether technical or non-technical users. By rethinking how you democratize, distribute, share and consume data through automation, you can save time, uncover insights quicker and enable data consumers to make better decisions by accelerating the connection with data producers. Your organization can increase productivity while minimizing time-consuming searches for relevant data used for analytics, customer experience programs and more to generate new value.

Informatica Intelligent Data Management Cloud

Informatica's data governance services are hosted on the Informatica Intelligent Data Management Cloud (IDMC), the industry's first and most comprehensive cloud focused solely on data management. IDMC is designed to cut through the complex challenges of dispersed and fragmented data, enabling organizations to truly innovate with their data on virtually any cloud, multi-cloud and multi-hybrid platform.

IDMC provides complete cloud-native management capabilities, with seamless Snowflake integration and a full complement of AI-powered tools — including data cataloging, data integration, data quality, privacy controls and much more — to facilitate a robust, efficient and scalable enterprise-grade cloud data governance framework.

Organizations can access data management services within a single metadata-driven platform with flexible, consumption-based pricing, unlike licensing and maintaining multiple point solutions. Start small and then expand as your business and data management requirements evolve.

IDMC helps organizations deliver on their digital-first initiatives and build a competitive edge with these critical benefits:

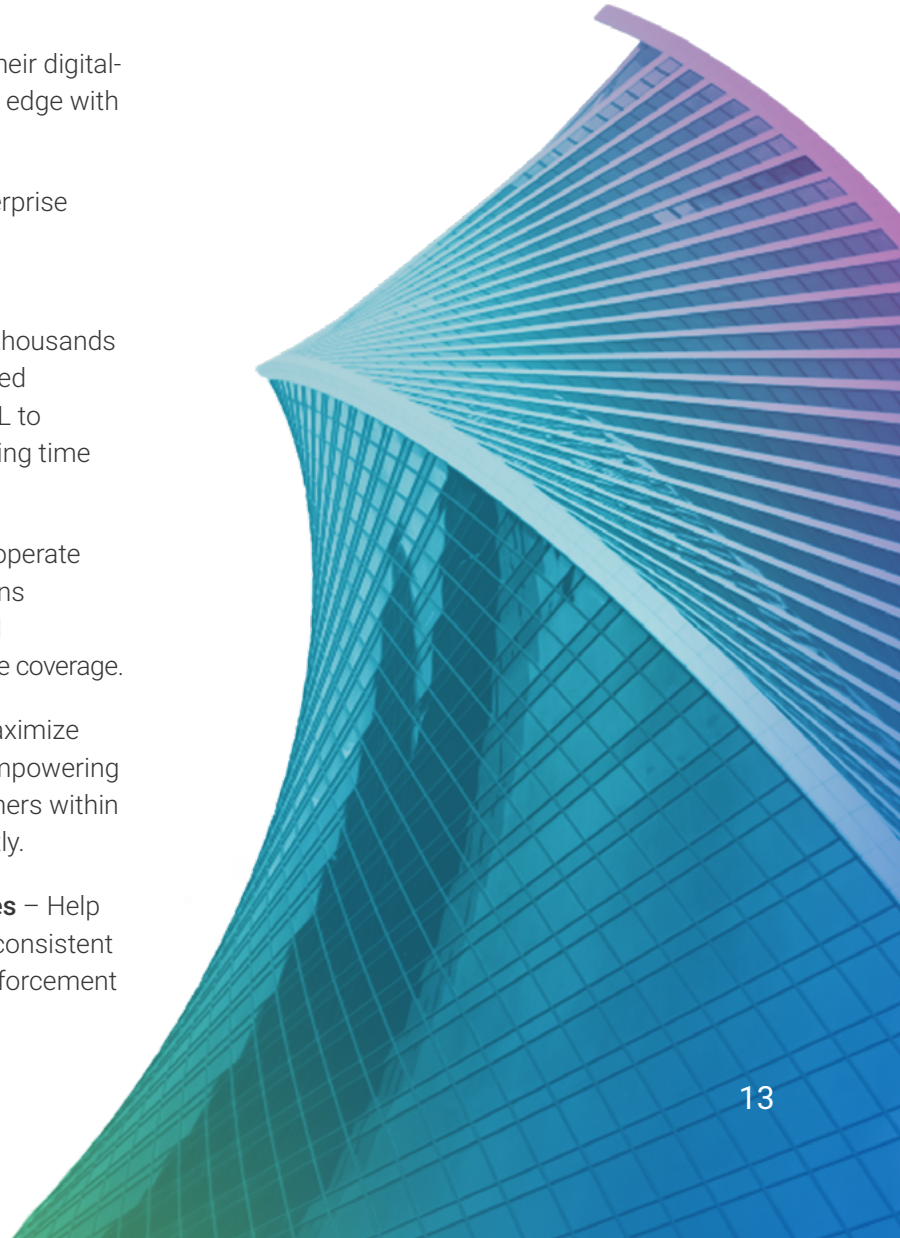
Cloud-native at scale – Scale for enterprise workloads with elastic and serverless processing with unrivaled flexibility.

AI-powered automation – Automate thousands of manual tasks and accelerate data-led transformations by applying AI and ML to data and metadata management, saving time and effort.

Multi-cloud, multi-hybrid – Run, interoperate and support wide-ranging combinations of multi-cloud and on-premises hybrid infrastructures to enable comprehensive coverage.

Low-code / no-code experience – Maximize agility and accelerate time to value, empowering the large community of data practitioners within your organization quickly and efficiently.

Security and trust as design principles – Help ensure appropriate levels of security, consistent data quality and end-to-end--policy enforcement for data privacy across the enterprise.

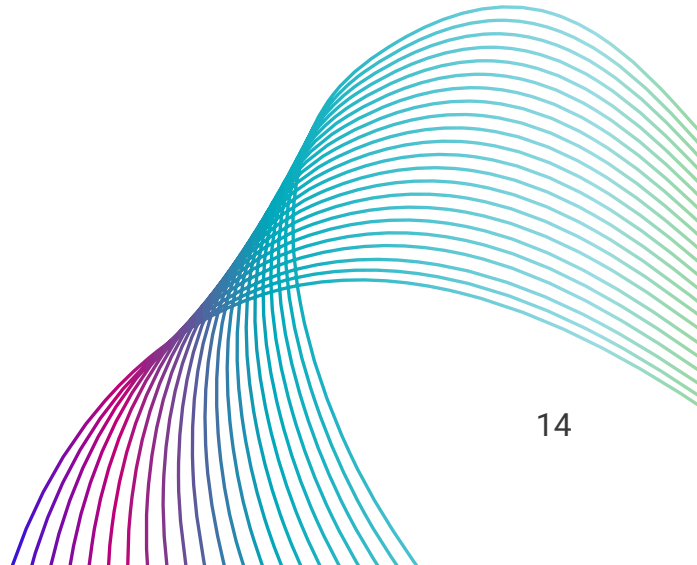


Informatica and Snowflake: Better Together

Informatica capabilities for Snowflake empower organizations to discover, manage, integrate, curate, cleanse and govern relevant and trusted data for analytics in a fast, efficient and scalable way.

Advantages include:

- **Seamless Integration between Informatica and Snowflake Data Cloud**: Track data as it is moved into and consumed from Snowflake and enable critical transparency — the foundation of trusted data. Faster, trusted data means accelerated adoption and time-to-value.
- **Automated, AI-powered cloud data management**: With industry-leading AI and ML capabilities built in at every level, data governance services that run on the Informatica IDMC give you the tools you need to operationalize your data governance framework and democratize your data securely for data consumers to use at cloud-scale.
- **Cloud-native design for future-proof analytics**: With Informatica and Snowflake's scalable data cloud, customers gain a modern, unified cloud data architecture that's built for the future, runs on virtually every cloud and combination of clouds and delivers insights from almost any data source at any scale.



Conclusion

In the past, data governance policies were created to comply with regulations and limited access to data. This approach hindered data consumers from using data to create new value and business outcomes. However, a cloud-native data platform can balance risk management with opportunities by enabling trusted data use. With a transparent, agile and automated approach, organizations can collaborate, experiment and innovate confidently with data. Modern data governance protects data while maximizing its utility to help organizations make smarter data-driven decisions. This approach also enables transparency to scale out for the future.

Snowflake opens the floodgates for sheer data volume and diversity, enabling who can access what data and when. The potential value is enormous. So too, are the data governance challenges. Still, thankfully, the days of manual data inventory, best guesses on appropriate uses and fitness for purpose and ad hoc pipeline monitoring are behind us.

By implementing a sound data governance framework and utilizing automated AI- and ML-powered tools, organizations can effectively leverage democratized data that can be easily shared with data consumers while minimizing risk. As Snowflake adoption continues to grow, organizations that adopt this approach will benefit the most. With this framework in place, individuals across the enterprise can easily access valuable insights from previously siloed customer, financial and product data.

Further Reading

Informatica Solutions for Snowflake

Informatica is an elite Snowflake partner with hundreds of joint enterprise customers. We enable customers to ingest, transform and govern trillions of records monthly on Snowflake Data Cloud to uncover meaningful insights using AI and analytics at scale.

LEARN MORE

Informatica Experience Lounge

To see our solutions for Snowflake in action, visit the Experience Lounge on the Informatica Success Portal. Explore self-guided product demos and simulations organized by role, ecosystem and more.

EXPLORE NOW

About Us

At Informatica (NYSE: INFA), we believe data is the soul of business transformation. That's why we help you transform it from simply binary information to extraordinary innovation with our Informatica Intelligent Data Management Cloud™. Powered by AI, it's the only cloud dedicated to managing data of any type, pattern, complexity, or workload across any location—all on a single platform. Whether you're driving next-gen analytics, delivering perfectly timed customer experiences, or ensuring governance and privacy, you can always know your data is accurate, your insights are actionable, and your possibilities are limitless.

Worldwide Headquarters
2100 Seaport Blvd,
Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/company/informatica
twitter.com/Informatica

CONTACT US

IN19-0723-4155