snowflake®

5 CRITICAL COMPONENTS
OF SUCCESSFUL
DATA GOVERNANCE
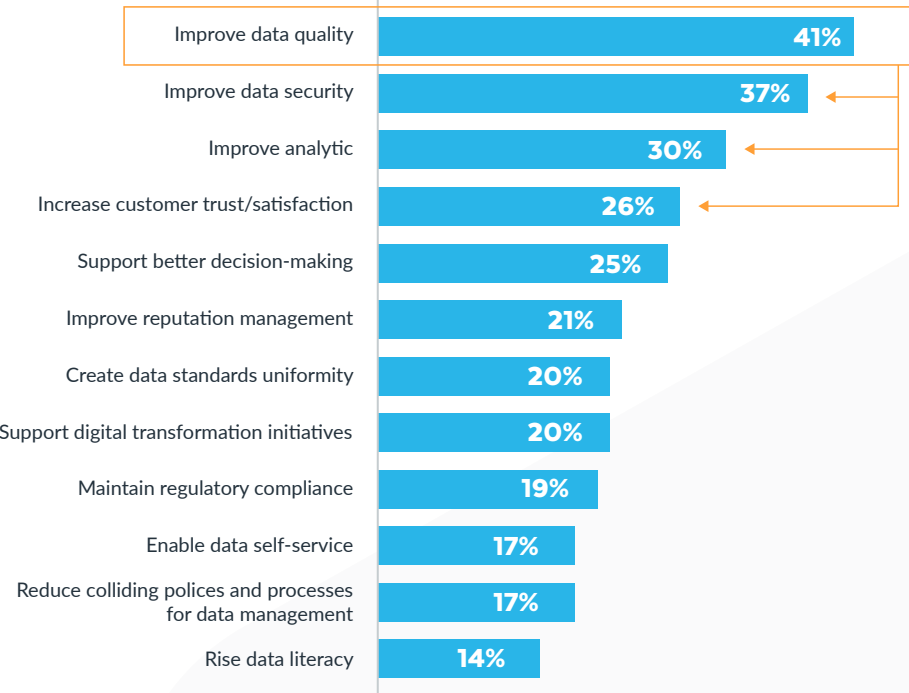
CHAMPION
GUIDES

EBOOK

TABLE OF
CONTENTS

INTRODUCTION

Organizational leaders worldwide agree that data governance is important; however, many organizations do not currently have a data governance program in place. In a 2020 Dataversity report, only 12% of companies had fully implemented programs, while 38% of programs were a work in progress, and 31% were just getting started.[1]

Historically, data governance was driven largely by regulatory compliance, and regulations such as GDPR made data governance a must-have for organizations. But recently, decision-making and analytics have become major drivers as well. This has changed the requirements for data governance programs: They must not only help companies follow guidelines and directives, but they must also empower companies to become data-driven.

The Snowflake Data Cloud provides the right foundation in support of data governance programs. Snowflake helps companies break down data silos and has features that enable companies to achieve compliance as well as better decision-making using secured, governed data. This includes availability on the three major cloud providers; elastic storage and compute; data encryption, access controls, and tracking capabilities; and integration with external data management tools. Here are the five critical components to keep in mind when building a successful data governance program, and how Snowflake's capabilities can help.

What are the top drivers of your organization's data governance program (i.e., what are the top outcomes it hopes to achieve)? (Percent of respondents, N=220, three responses accepted)

Driver	Percent
Improve data quality	41%
Improve data security	37%
Improve analytic	30%
Increase customer trust/satisfaction	26%
Support better decision-making	25%
Improve reputation management	21%
Create data standards uniformity	20%
Support digital transformation initiatives	20%
Maintain regulatory compliance	19%
Enable data self-service	17%
Reduce colliding polices and processes for data management	17%
Rise data literacy	14%

COMPONENT #1:
DATA ARCHITECTURE

According to a Gartner® report, "data and analytics leaders are quick to highlight the need to break down silos across business areas so that data and analytics can be treated more as an enterprise wide asset."[2] However, overcoming silo mentality in business areas is one of the most challenging aspects of data and analytics governance. Rigid legacy data architectures promote data isolation by hindering the sharing and dissemination of information throughout the entire organization.

Legacy architectures also make it difficult for companies to organize information coherently. Siloed, disorganized information makes it impossible to apply data governance, whether that be tracing data lineage, cataloging data, or applying a granular security model. Snowflake provides the right platform for successful data governance in three ways: silo reduction, secure data sharing, and flexibility.

Snowflake's centralized metadata repository provides a single source of truth for all your data in Snowflake, making it easier to find, view, and trace data, and to organize it both manually and with automation and AI. Data discovery, data inventory, and tracing data lineage are made easier with a platform that provides transparency into the source and granularity of data. Snowflake's Secure Data Sharing capabilities enable companies to instantly and securely share data without losing control of it, while keeping it centrally located and easily traceable. Customers can perform data cataloging through one of Snowflake's security and governance partners and can use any data modeling technique, enabling them to create a customized governance approach.

COMPONENT #2:
DATA QUALITY

Data governance involves oversight of the quality of the data coming into a company as well as its usage throughout the organization. Data stewards need to be able to identify when data is corrupt, inaccurate, or outdated, or when it's being analyzed out of context. They need to be able to set rules and processes easily. The ability to trust data is a cornerstone for data-driven organizations that make decisions based on information from many different sources. Fifty-eight percent of companies in the Dataversity report said understanding the quality of source data was one of the most serious bottlenecks in their organization's data value chain.

According to the report, "Automating and matching business terms with data assets and documenting lineage down to the column level are critical steps to optimizing data quality."

With Snowflake, companies can increase trust in the validity and accuracy of their data by breaking down silos. Snowflake Streams and Tasks enable identifying and managing changes to underlying data within a pipeline. As inbound source data changes, pipelines in Snowflake can be built to spot specific patterns, and when such a condition is identified, Tasks can be triggered to track and trace changes as they occur, and drive transformation logic for dependent processes when needed. Snowflake external functions allow third-party services, such as address standardization or natural language processing, to be invoked within a pipeline, helping to maintain a high quality of data in the platform. Finally, when robust data quality is needed, Snowflake's large ecosystem of data integration partners offers capabilities to further achieve a high level of data quality and trust.

COMPONENT #3:
DATA MANAGEMENT

Data governance requires companies to answer an important set of questions: What data do I have, and where does it reside? Who has access, and how do they use it? Data management is key to performing this sort of data inventory; a strategy and methods are needed for accessing, integrating, storing, transferring, and preparing data for analytics. According to Forrester Research, "Effective data governance grows out of data management maturity."[3] However, IDC says almost half of organizations struggle with data management deficiencies.[4]

Snowflake helps companies to classify columns containing sensitive personal data and to tag objects for reporting and access control. Customers can further track the lineage of and access to their data. Access History is an account usage view that provides an audit record of all tables and columns accessed by all users, and it also houses records of all data movement between tables. Additionally, object dependencies is an account usage view that provides records of all objects related by metadata, such as the tables backing a view. Snowflake's ecosystem partners such as Informatica and Collibra are embracing these new views to provide end-to-end lineage, access audit, and data popularity insights to their end users. For example, integrating Snowflake with Collibra, companies can provide complete business context around data with Collibra's governed data catalog and enable IT and business teams to collaborate on, share, and access trusted, governed data in a central location.

COMPONENT #4:
DATA SECURITY

Along with the proliferation of data sources both inside and outside enterprises, data breaches are on the rise. Data governance is vital to improving data security. Similar to successful data management, data security hinges on traceability: knowing where your data comes from, where it currently is, who has access to it, how it's used, and how to delete it. Data governance sets rules and procedures, preventing potential leaks of sensitive business information or customer data so data doesn't get into the wrong hands. However, legacy platforms create siloed information that is difficult to access and trace. Those silos are often exported, sometimes to spreadsheets, and duplicated to combine data with other siloed data, making it even harder to track that data.

Snowflake's cloud data platform has several features that enable traceability and accountability. With role-based access control, Snowflake provides granular control over access to objects to identify who can access what objects, what operations can be performed on those objects, and who can create or alter access control policies. Row access policies dynamically control access based on user authorization. Dynamic Data Masking allows security administrators to create policies to limit the visibility of data at a column level; data can be unredacted, partially redacted, or fully redacted. This capability also includes integration with external tokenization services for added layers of security. Conditional masking enables masking of protected data based on values in the other column(s) of a table. With tag-based masking, masking policies can be automatically assigned to columns based on the presence of a tag on the column. With secureviews and secure userdefined functions, administrators can ensure that end users have no access to hidden underlying data or the logic used to mask it. Moreover, Snowflake offers end-to-end encryption (E2EE) so that only authenticated users and the runtime components can read the data. No third parties, including Snowflake's Data Cloud or the underlying cloud vendor, can see your stored data. In the event of a security breach of the cloud platform, E2EE protects stored data because it is always encrypted.

COMPONENT #5:
DATA COMPLIANCE

Businesses often begin thinking about data governance when they need to comply with regulatory policies such as GDPR, HIPAA, PCI DSS, and the U.S. Sarbanes-Oxley (SOX) law. In the Dataversity report, 48% of companies ranked regulatory compliance as their primary driver for data governance. These regulations require organizations to be able to trace their data from source to retirement, identify who has access to it, and know how and where it is used. Data governance sets rules and procedures around ownership and accessibility of data. Without it, sensitive information can get into the wrong hands or be improperly expunged, leading to governmental or regulatory financial penalties, lawsuits, and even jail time.

Snowflake offers features that can set controls on data ownership and access, enabling the implementation of rules and procedures for data governance. These include Dynamic Data Masking and secure views. Data providers retain ownership of their shared data through Snowflake's collaboration capabilities, with the ability to revoke access at any time. By retaining control of the data, it is much easier for companies to properly and securely dispose of data. Snowflake's data integration partners offer data lineage tools for auditing and tracing to assist in these types of compliance scenarios.

Because compliance and security are so important, Snowflake's portfolio of security and compliance reports is continuously expanding. The current list includes certifications for ISO 27001, AICPA/SOC, SOC 1 Type 2 and SOC 2 Type 2, HIPAA, FedRAMP Authorized (Moderate), PCI DSS, FISMA (Moderate), NIST 800-171, FIPS 140-2, and ARS 3.1.

CONCLUSION

Data governance is a continuous organizational commitment that needs to grow and adjust to new compliance, business, and security challenges. Focusing on data infrastructure, data quality, data management, data security, and data compliance will help ensure that data governance programs can scale and adapt to meet these challenges. Whether a company is seeking to comply with regulations, become more data-driven, or both, Snowflake provides a strong foundation for successful data governance.

ABOUT SNOWFLAKE

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 510 of the 2022 Forbes Global 2000 (G2K) as of July 31, 2022, use Snowflake Data Cloud to power their businesses. Learn more at snowflake.com.

Learn more at **snowflake.com**