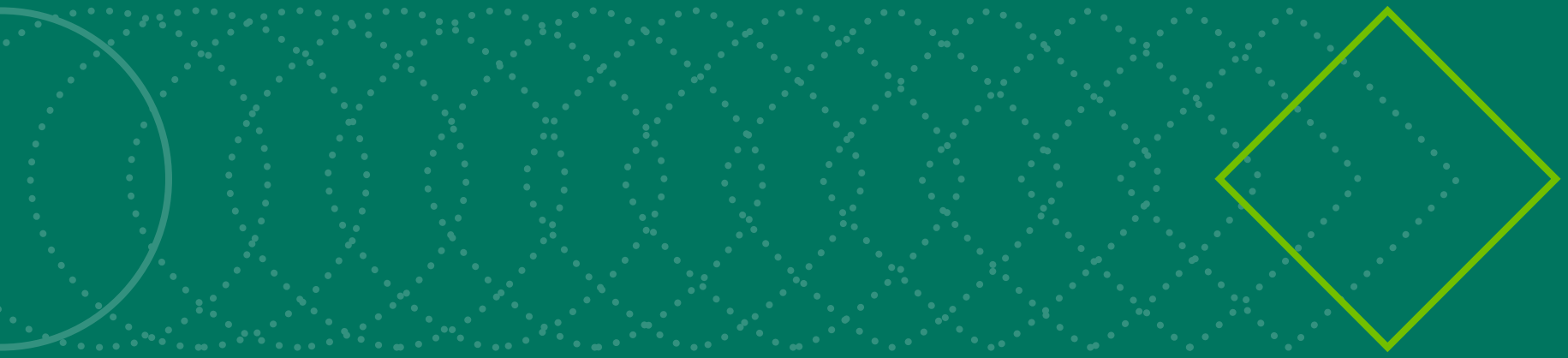
Collibra | snowflake®

Revised March 2023

Four steps to ensure a successful Snowflake migration

Accelerating adoption of governed enterprise data and analytics in the Data Cloud



The Business imperative

Digital transformation is impacting every industry and sector of the economy, and data lies at the heart of each of those transformation efforts. Taking a data-driven approach enables organizations to improve all aspects of their operations, but to fully realize that potential requires the right mix of capabilities.

Collibra has partnered with Snowflake to ensure mutual customers benefit from an integrated cloud platform to drive trusted insights.

The Snowflake Data Cloud is highly scalable, available, and simple to manage, enabling enterprises to aggregate, store and analyze data collected across all of their silos.

Those capabilities are perfectly complemented by Collibra's expertise in data intelligence and data governance. Collibra ensures that all of the data within Snowflake can be easily found, understood, trusted, accessed securely, and shared while maintaining compliance with all relevant policies.

A recent survey by Gartner found that the top two drivers for migrating to cloud run analytics were 'scalability to achieve performance' and 'operational effort (ease-of-use and time-to-deploy).' At the same time, more than 80% of respondents noted that their organization was pursuing a multi-cloud strategy (using at least two cloud service providers)

Gartner 2020

Collibra | snowflake

Migration challenges

Cloud data platforms like Snowflake offer clear and compelling benefits, including scalability, performance, availability, and ease of management. However, migrating enterprise data and analytics to the cloud can still be fraught with challenges.

Without the right foundations, organizations will typically struggle with:

Discoverability

Data security and privacy

Understanding data flows

People/process

Data quality and consistency

Top migration challenges

Discoverability. A cloud data platform is a powerful way to aggregate enterprise data assets. However, without curating necessary metadata, data consumers will struggle to find the right data for their analysis or understand its context.

Understanding data flows. To support the migration process, organizations need a detailed understanding of data lineage to map key dependencies in their data architecture and to easily view usage of all their data, including senstive PII.

Data quality and consistency. Consolidating data from multiple silos highlights the need to interpret data consistently, using clear definitions of key business terms and metrics. It also begs for a consistent approach to measuring data quality, providing consumers not only with insights into quality metrics but the ability to compare those metrics across sources.

Data security and privacy. Aggregating enterprise data also poses greater complexity in controlling access. To keep sensitive data secure, organizations need a detailed understanding of relevant policies and restrictions on data usage. Without that knowledge, the tendency is either to err on the side of caution and overly restrict access, thereby limiting adoption; or fail to implement appropriate controls and risk incurring significant liabilities.

People/process. Cloud architectures have the potential to support more agile data operations. But that potential can only be fully realized when supported by the right people and processes. Knowing who is responsible for each data source, how to raise an issue relating to data quality, how to request new data items be added to the warehouse – all of these workflows are dependent on having the right governance framework in place, with clearly defined roles and responsibilities.

Collibra | snowflake

Four key steps for success

Organizations migrating enterprise data and analytics to the Snowflake Data Cloud need to be aware of the potential challenges and take preemptive action to address them. The four steps that we now outline provide a basic primer to support a successful migration and accelerated enterprise adoption of the Snowflake Data Cloud.

1. **Implement a unified approach to data quality**
2. **Catalog data ingested into Snowflake**
3. **Manage access controls**
4. **Supplement technical metadata with expert insights**

75%

By 2022, 75 percent of all databases will be deployed or migrated to a cloud platform.

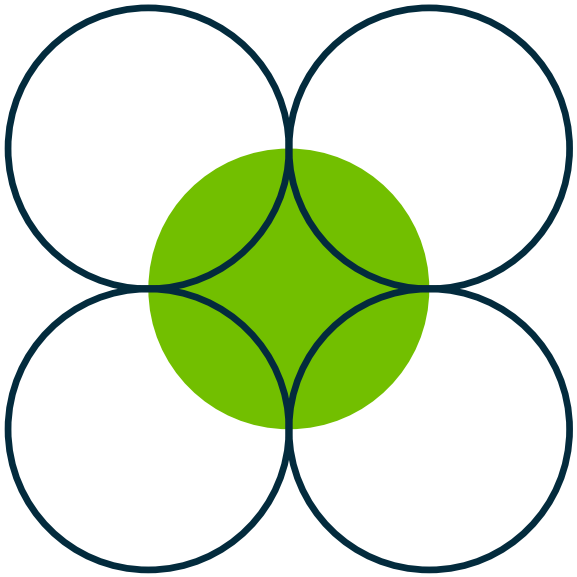
Gartner, 2019

Collibra | snowflake

Step 1

Implement a unified approach to data quality

Data quality can play an integral role in cloud data migrations. Refining the quality of data before moving it to a cloud data platform enables organizations to enhance usage of that data after migration. It also allows data reconciliation during data movement so identify missing records and broken relationships. With proper data quality rules in place pre ingestion, organizations can predict and prevent business disruptions.



By tracking quality metrics through the data lifecycle, and reconciling results between source and target, organizations can detect if any drift has occurred during the migration.

This kind of use case calls for a unified approach to data quality and the ability to monitor quality consistently across multiple environments. For that to happen, organizations need to be able to:

- Use machine learning to profile data, understand its unique properties and auto-generate adaptive rules to detect quality issues
- Measure quality across a full range of dimensions - including accuracy, consistency, completeness, redundancy and timeliness
- Make data quality metrics visible to data consumers to help highlight trusted sources

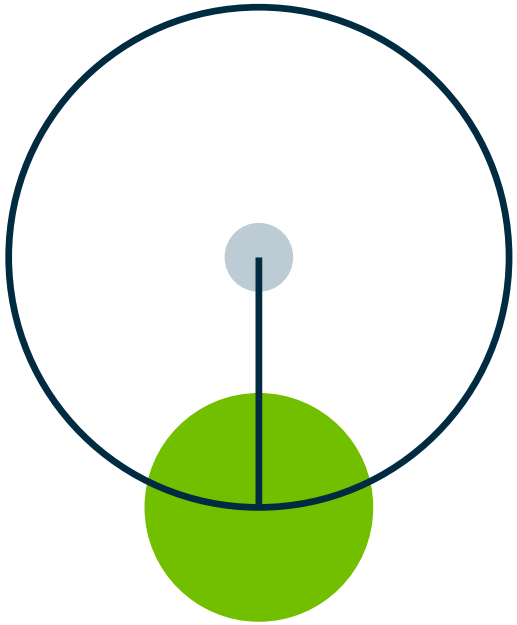
Step 2

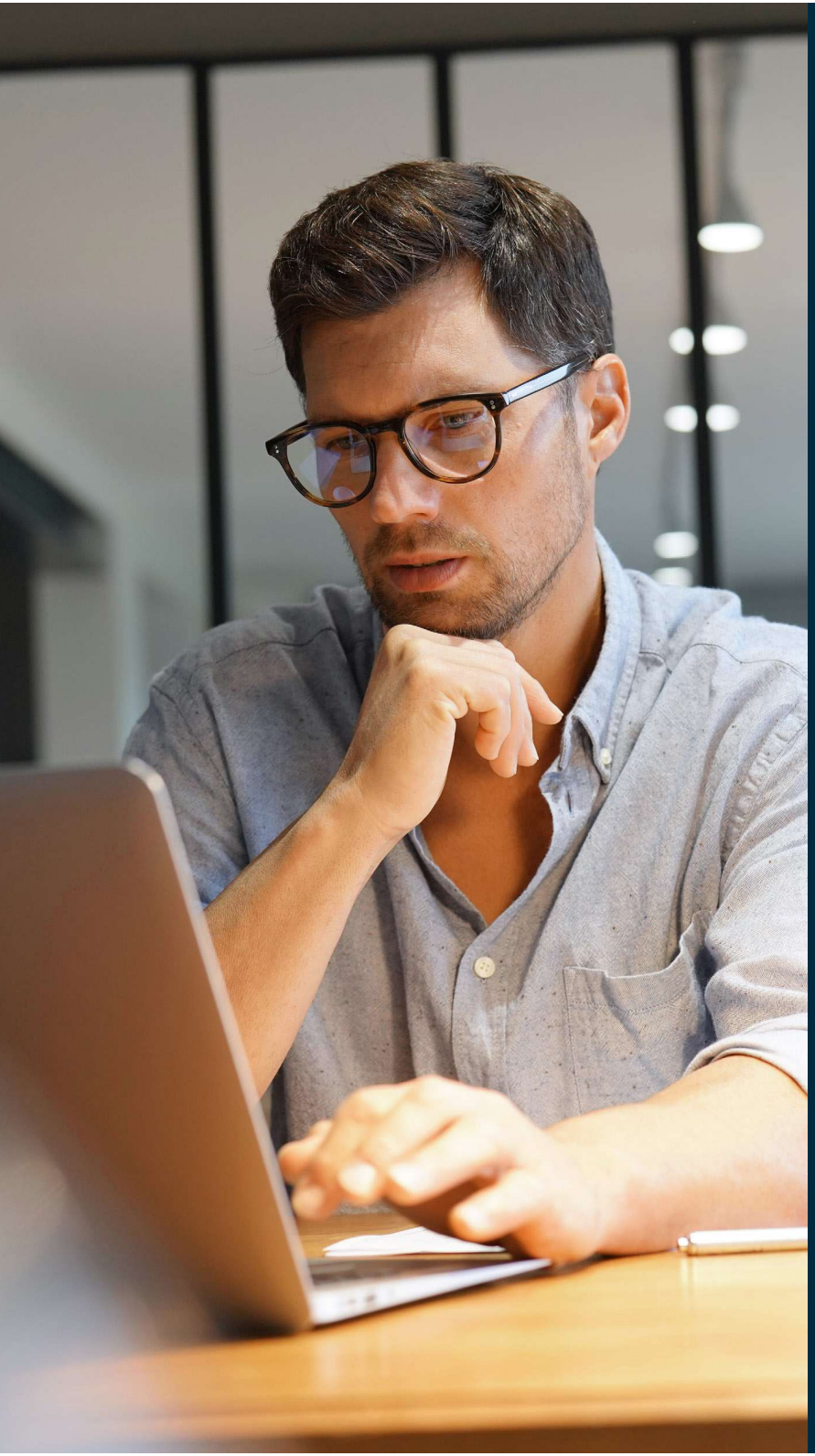
Catalog data ingested into Snowflake

It is crucial that data ingested into the Snowflake Data Cloud is properly cataloged, although this should not slow down the migration process.

For that to happen, organizations should:

- Automatically harvest technical metadata as data is loaded into the cloud. This will provide data consumers with the necessary context to support rapid discovery and understanding (through a data dictionary)
- Use machine learning to profile ingested data and classify sensitive information (including PII and PHI)
- Integrate data quality metrics in an enterprise data catalog as part of a broader governance framework to drive improvements
- Keep the catalog up to date by automatically tracking changes to tables and schemas, ensuring that data within Snowflake remains easy to discover and share (using accurate metadata)





Freddie Mac

Freddie Mac's data ecosystem transformation pivoted on Collibra

"We selected Collibra to meet our data governance and data stewardship requirements, as well as, to improve the collaboration within and across different Freddie Mac teams.

The metadata, data classifications, and technical metadata from Collibra power the data lake access management engine and the pipelines that hydrate our data lake. As the data lake gets hydrated with the approved data sets, the data in the lake is already curated and is ready for access by data consumers. In addition, the metadata from AWS S3 buckets, parquet files, and Snowflake is available in Collibra for data consumers.

Data users are able to get an integrated business view of the technical metadata, data quality, data movement controls, and business metadata in a single central universal platform.

Having an integrated business view of this rich content is very powerful — it empowers our business users so that they have the confidence they need to use the data they need."

Vikram Chopra, Senior Manager, Single-Family, Data & Decisions, Freddie Mac
Collibra Data Citizens '21

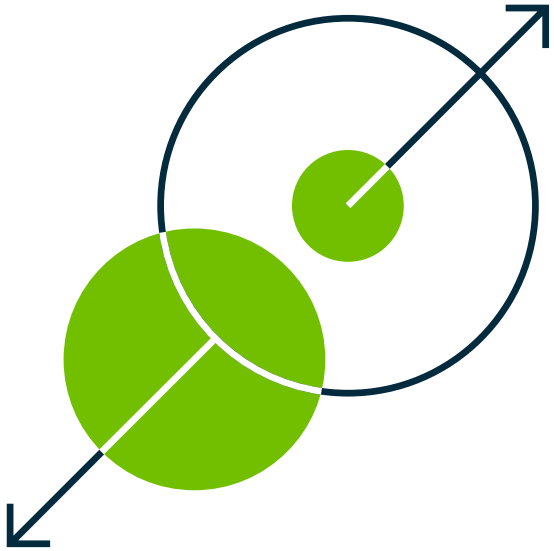
Step 3

Manage access controls

As organizations look to extract greater value from data, they must also be cognizant of relevant restrictions. Laws relating to data privacy, industry-specific regulations and internal policies all impact data usage and access rights in different ways. These laws not only limit internal use cases, but also the ability to share data with partners and customers.

Snowflake's scalability makes it an ideal platform to aggregate enterprise data and enables multiple teams to work concurrently on that data without impacting query performance. The platform excels at offering secure data sharing capabilities, enabling organizations to share access to data without replicating or moving it.

However, to take full advantage of those capabilities, organizations need to ensure they are controlling access in line with relevant policies. Collibra enables organizations to centrally manage data policies and accurately classify data from a compliance perspective. This provides the basis for understanding usage restrictions, which in turn, can help define access controls in a way that complies with all relevant policies.



COX AUTOMOTIVE™

Advancing the lake: Using Collibra for the Cox Automotive data marketplace

"We use Collibra as our one-stop-shop for data consumers and Collibra became part of our data marketplace suite along with AWS and Snowflake. Ensuring that we were governing data safely, but also enabling people throughout the organization to find and use the data they need to maximize value to our customers.

We need to also make sure people can get their hands on data. And so we've embedded the entitlement request process into Collibra for all of the data in our data platform. And so, as you can see, as users check out data sets and the data basket, we have tied our usage policies into that, where we can actually have you read and acknowledge acceptance of how to use the data before you actually get your hands on it. This process enables us to govern data access both in S3 and in Snowflake and enables kind of that full circle process for an end user. So you're not swivel chairing between screens and applications."

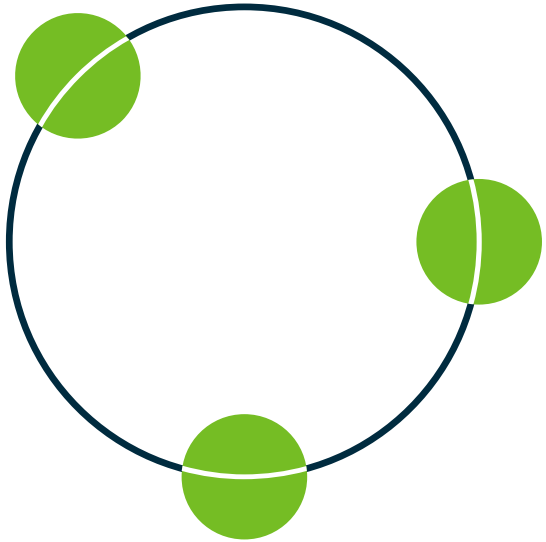
Vikram Chopra, Senior Manager, Single-Family, Data & Decisions, Freddie Mac
Collibra Data Citizens '21

Collibra | snowflake

Step 4

Supplement technical metadata with expert insights

Automation and machine learning can drive productivity, but to maximize the value of enterprise data, organizations also need to capture expert insights. This value-added metadata serves a much broader range of functions — not only enabling consumers to find and understand data, but to interpret that data consistently, gauge its trustworthiness, and understand whether restrictions apply to its access or use.



Business glossary. A business glossary clears up ambiguity in the definition of key business terms and metrics, helping to promote consistency.

Certification. Certifying data and analytics is a crucial process in assuring trust. Certification helps capture the intelligence of data experts and makes it easier for data consumers to find trusted sources.

Ownership and stewardship. Assigning clear roles and responsibilities helps to promote accountability and ensures the right processes are in place to aid data discovery, promote understanding and drive improvements in data quality.

Data lineage. Lineage can be crucial in supporting pre-migration planning, helping organizations understand logical data flows and dependencies. Once data has been migrated, lineage also serves as a valuable resource in supporting ongoing data operations.

Conclusion

Snowflake is a highly popular choice for organizations that want to implement a cloud platform for enterprise data and analytics. But while Snowflake offers numerous technical benefits, there are still several obstacles to overcome to ensure a successful migration and widespread adoption of cloud data platforms.

We have highlighted four simple steps to ensure successful migration and immediate adoption of Snowflake Data Cloud:

1. Measure data quality consistently across on-prem and cloud environments to detect drift. Leverage machine learning and parallel processing for scalability.
2. Ensure all of the data ingested into the cloud is cataloged, using automation where possible so as not to delay the process.
3. Control access to data in a way that complies with relevant policies.
4. Supplement basic technical metadata with expert insights, leveraging tools and capabilities such as a business glossary, certification, ownership and stewardship, and data lineage.

These steps demonstrate how Collibra and Snowflake offer technology solutions that naturally complement each other. Snowflake provides a scalable platform to aggregate, store, analyze, and share data. Collibra ensures all of that data can be easily found, understood, and trusted, helping to promote data quality, consistency, and compliance.