

Accelerate Analytics and AI Initiatives on Snowflake With an AI-Powered Data Catalog

Key Benefits

- Empower non-technical and technical users with rapid data discovery and collaboration on data at scale
- Easily visualize, trace, and understand your data from source to target with end-to-end data lineage
- Extract deep metadata and data lineage from your Snowflake and non-Snowflake data sources
- Enable robust enterprise-wide data governance, privacy, and regulatory compliance programs

Rapid Data Discovery and Governance Across Multi-Cloud Environments

Data modernization has fast become a key business imperative for enterprises of all types and sizes on their digital transformation journey. Whether it's cloud data warehouse and data lake modernization or migration to other forms of modern storage options, virtually every enterprise wants the ability to easily and cost-effectively store, access, blend, and analyze massive volumes and variety of data in real time. At the heart of this shift is the need to derive value from data to drive innovation, improve customer experiences, increase operational agility and speed, as well as democratize the use of fully governed data throughout the enterprise while enabling self-service analytics and artificial intelligence (AI) workloads at scale.

IDC estimates that by 2025 worldwide data will grow by 61% to reach 175 zettabytes.¹ While cloud data sources are prevailing in importance, it is predicted that nearly 90% of enterprises will be running their businesses across on-premises, dedicated private cloud and multiple public cloud environments, leveraging a host of modern and traditional data sources. This will further increase the complexity of the data landscape in which enterprises operate with data spread across hundreds of disparate data sources.

One of the biggest hurdles for enterprises when operating in a complex data landscape is the lack of end-to-end visibility and understanding of their data. With petabyte-scale data residing across modern data platforms such as Snowflake Data Cloud as well as other data sources, most enterprises lack in-depth information on what data they have, where it resides, who owns the data, whether the data is certified for use, if it meets data quality standards, whether usage of the data complies with governance and privacy policies, what transformations each dataset may have undergone throughout its lifecycle, and what data dependencies exist across the entire data ecosystem.

¹ [Forbes: 175 Zettabytes by 2025](#)

For enterprises that are migrating their data to Snowflake, this challenge is further amplified when the metadata and data lineage that resides in various data sources such as complex enterprise applications and systems, stored procedures for databases, data warehouses, and multi-vendor ETL and BI tools, is trapped and buried. It is often difficult to extract and even harder to understand. This poses a host of operational challenges and regulatory risks for enterprises. Furthermore, analytics and AI workloads depend heavily on the quality and veracity of the underlying data. By some estimates, 87% of data science projects fail to move to production due to lack of access to trusted data.² For instance, building trust in AI models and insights requires data scientists to have comprehensive visibility and understanding of the data. It requires a robust and unified data cataloging and governance foundation that spans across data sources, on-premises and multi-cloud environments enabling enterprises to overcome data and governance silos.

Informatica[®] Enterprise Data Catalog enables you to build a comprehensive inventory of all your metadata regardless of where it resides—inclusive of Snowflake Data Cloud and any other data sources. Powered by the metadata-driven intelligence in the Informatica CLAIRE[®] AI engine, Enterprise Data Catalog delivers advanced capabilities designed for rapid data discovery, curation, and collaboration on data and metadata at scale. With end-to-end data lineage and impact analysis, you can easily visualize, trace, and understand the flow of data within and outside Snowflake Data Cloud at a granular level. Enterprise Data Catalog is a foundational pillar for enabling a holistic and comprehensive data governance and data democratization strategy for all your data regardless of where it may reside—from on-premises, hybrid, to multi-cloud environments.

Key Capabilities

Rapid Data Discovery Powered by Advanced Machine Learning

Enterprise Data Catalog enables rapid discovery of data with powerful, Google-like semantic search—empowering data stewards, data scientists, analytics, and data governance leads as well as data architects to easily find the data they need across Snowflake Data Cloud and any other data sources. Users can quickly discover and profile data, identify its location, and obtain other key attributes about the datasets at scale. Semantic search is also applied to inferred data domains, including synonyms and concept matching, so that no data asset is left undiscovered across on-premises and multi-cloud environments.

Using advanced statistical and metadata-driven machine learning algorithms, Enterprise Data Catalog tackles the inherent complexity in data when it comes to discovering, tagging, clustering, and identifying similarities and patterns in data, enabling you to intelligently catalog all types of data at scale.

Enterprise Data Catalog also allows for easy import of business glossary assets such as terms, policies, and classifications from Informatica Axon[™] Data Governance as well as third-party tools. You can add rich business context to the data by automatically associating business terms with the right technical metadata.

² [Venture Beat: Why do 87% of data science projects never make it into production?](#)

Broad and Deep Metadata Connectivity With End-to-End Data Lineage and Impact Analysis

Enterprise Data Catalog is the catalog of catalogs with both broad and deep metadata connectivity. It offers the most comprehensive set of scanners that are purpose-built to extract deep metadata and data lineage from a myriad of widely adopted data sources across on-premises, hybrid, and multi-cloud environments.

End-to-end data lineage and impact analysis capabilities allow you to easily visualize, trace, and understand the flow of data within Snowflake Data Cloud as well as linked data sources such as those on Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), enterprise applications and systems, databases, and ETL and BI tools. You can perform detailed impact analysis of transformations within Snowflake Data Cloud as well as on third-party upstream and downstream data assets and linked systems. You can interactively trace data origin through lineage views at any level—from business-friendly, system-level views that highlight the endpoints to granular views that include all the complex details in between. Additionally, a drill-down lineage view expands any lineage path to show granular column- and metric-level lineage.

Advanced Scanner for Snowflake

Enterprise Data Catalog Advanced Scanner for Snowflake is purpose-built for enabling deep extraction of metadata and detailed data lineage. The Advanced Scanner enables you to scan both static and dynamic code as well as perform language parsing to obtain automated data lineage. Extracted data lineage provides full visibility into the procedure calls with parameter tracking and more. It enables you to automatically track the movement of data within, to, and from the data warehouse on Snowflake Data Cloud. Users can also track data that's moving out of the data warehouse through the Snowpipe³ capabilities.

The Advanced Scanner for Snowflake automatically connects database objects with underlying storage including Amazon S3, Azure Data Lake Storage Gen2, and Google Cloud Platform accessed using Stages and Pipes. Moreover, it generates detailed lineage from stored procedures supporting JavaScript parsing, variable binding, and dynamic SQL lineage extraction. It allows you to easily visualize and gain quick access to every expression being applied to the data from source to target at a granular level.

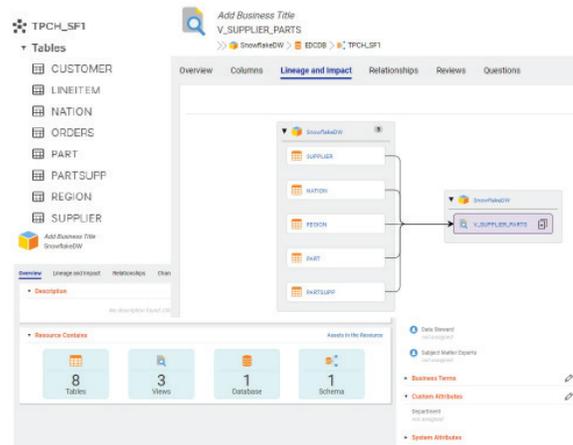


Figure 1: Advanced Scanner for Snowflake.

³ [Snowflake: Introduction to Snowpipe](#)

Data Collaboration and Social Curation with Intelligent Crowdsourcing and Annotations

Enterprise Data Catalog empowers data stewards, data scientists, and data governance and analytics leads to easily find the most relevant and trusted data for analysis by harnessing the combined power of sophisticated machine learning algorithms, human expertise, and collaboration. Data owners and subject matter experts can certify datasets and provide ratings and reviews, enabling social curation of data. A Q&A platform enables subject matter experts to answer common questions from users.

Integrated Data Quality

View data profiling statistics, data quality rules, scorecards, and metric groups alongside technical metadata to understand the quality of data assets within Snowflake Data Cloud before using data for analysis. Profiling statistics include value distributions, patterns, and data type and data domain inference.

Advanced Data Asset Analytics

Data Asset Analytics provides prepackaged reports and dashboards on data asset inventory, usage, enrichment, level of collaboration, and more. Reports are extensible and can be exported, enabling data leaders to share business adoption and value metrics with stakeholders. Automated Data Value Calculator, a first-of-its-kind capability, allows an enterprise to measure and optimize the value of its data assets based on key factors that impact data value. For instance, you can obtain information on what percentage of your data inventory resides in key data sources as well as the types of data your users are accessing. This will help you proactively prioritize, manage, and optimize the value of your data assets when migrating to Snowflake Data Cloud.

Key Benefits

Enable Comprehensive Data Governance at Scale

The Informatica Data Governance and Privacy solution brings together advanced capabilities delivered by Informatica's AI-powered Enterprise Data Catalog, Axon Data Governance, and Data Privacy Management using a consistent metadata-driven platform to share data intelligence. The intelligent, integrated, and modular solution allows you to democratize data use rapidly and cost-effectively with trust assurance, encompassing all your data—whether it resides in Snowflake Data Cloud or alternative platform data sources across on-premises, hybrid, or multi-cloud environments.

Capture and Enforce Privacy Policies

The solution leverages comprehensive, user-defined privacy policies to align stakeholders and workflows while rapidly discovering data that is subject to privacy regulation compliance that may reside in Snowflake Data Cloud as well as other data sources. For example, using Boolean match conditions and acceptance thresholds, users can search any of the multiple data elements controlled by privacy policies (e.g., the CCPA, the GDPR, BCBS 239, HIPAA, and more).

About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category, or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities, or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

Empower Non-technical and Technical Users

Data stewards, analytics and data governance leads, data scientists, and data architects can rapidly discover, certify, and collaborate on data at scale. Users can easily identify which datasets may contain PII (Personal Identifiable Information), and in which data source systems it originates. End-to-end data lineage enables users to trace and understand the movement of sensitive data sharing at a granular level, for the intelligence needed to make informed decisions on data exposure and value to the organization.

Achieve Faster Time to Trusted Insights

You can build a comprehensive and unified view of all your analytical data that resides within and outside Snowflake Data Cloud. Armed with end-to-end visibility and in-depth information, data science teams can build accurate AI models with trusted data and enable self-service analytics with confidence.

Next Steps

To learn more, please visit the web pages for [Informatica Enterprise Data Catalog](#) and [Enterprise Data Catalog Advanced Scanners](#).



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

IN17_0921_04199

© Copyright Informatica LLC 2021. Informatica, the Informatica logo, Axon™ and CLAIRE® are trademarks or registered trademarks of Informatica LLC in the United States and other countries. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners. The information in this documentation is subject to change without notice and provided "AS IS" without warranty of any kind, express or implied.