

Trusted Data on Amazon Web Services with Cloud Data Governance and Catalog

Key Benefits

- Generate trusted insights with integrated data and AI model governance
- Enable analytics and self-service for all users
- Empower effective collaboration across the organization
- Get started instantly, scale quickly and lower operational costs with cloud-native service on AWS

Trusted Data: The Key to Analytics and AI Initiatives on Amazon Web Services (AWS)

Organizations are seeing greater adoption of cloud and automation. As a result, they are looking to analytics and artificial intelligence (AI) to help them modernize their processes, increase relevance and create more value for their stakeholders. In a recent survey of executives from top, industry-leading organizations, more than 90% of respondents indicated that they plan to increase their investment in data and AI.¹

Despite this increased investment, organizations have struggled to reap the full benefits of this transformation. According to a recent survey, only 53% of organizations have implemented AI beyond pilots and proofs of concept, and a mere 13% have successfully scaled their AI efforts, citing a lack of governance models as a primary challenge to expanding deployment.² One ongoing challenge is how to provide stakeholders with the access they need to the trusted data that enables insights, enhances customer experience, provides new services or ensures greater compliance with regulatory authorities. One critical key to the success of these strategic initiatives that rely on data is to have the governance capability that improves operational efficiency, assures data quality and provides trusted access to empower data professionals in their decision making.

Govern and Share Your Data at Scale on AWS

Informatica® Cloud Data Governance and Catalog is a unified, cloud-native SaaS solution that enables you to find, understand, govern and trust your data. It brings together data cataloging, governance, quality and democratization capabilities into a singular, unified solution for data intelligence.

This multi-tenant solution is built for organizations with vast cloud data assets on platforms like Amazon S3 and Redshift. Combining data discovery, data lineage, profiling, business glossary creation and stakeholder and policy management, Cloud Data Governance and Catalog enables self-service analytics and data governance for cloud data lakes and data warehouses. It also allows the ability to document and manage AI models and their implementations.

¹ New Vantage Partners, Data and AI Leadership Executive Survey 2022

² Capgemini Research Institute, "The AI-powered Enterprise: Unlocking the potential of AI at scale," July 2020

Cloud Data Governance and Catalog integrates into your existing data landscape and scans sources on AWS or hybrid cloud environments, analytics/BI systems, ETL tools and more. As an AWS-native solution, infrastructure is available immediately and at exactly the scale needed.

Key Capabilities for AWS Users

AI/ML-powered Engine to Drive Insights from Metadata

Automation is critical to manage and govern large data ecosystems. Cloud Data Governance and Catalog automates metadata extraction from heterogeneous sources (including S3, Redshift and Athena), classification of data assets and association of glossary terms to data. It can also infer relationships like joins and lineage among datasets using AI/ML capabilities including schema matching.

Browsable Hierarchies and Powerful Search Capabilities

Users can use natural language to search for critical assets across business and technical domains. Filtering and preview capabilities allow them to quickly review and identify desired assets. The solution provides a powerful browsing experience across business and technical assets to help data analysts and scientists understand different data elements in context.

Data Lineage and Impact Analysis

Users can interactively trace data origin through lineage views at any level, from business-friendly views that highlight the endpoints to granular views that include all the intricate details in between. This includes the ability to parse SQL scripts and stored procedures in Redshift and other data sources. A drill-down lineage view expands any lineage path to show granular column- and metric-level lineage. In addition, users can perform detailed impact analysis on upstream and downstream data assets.

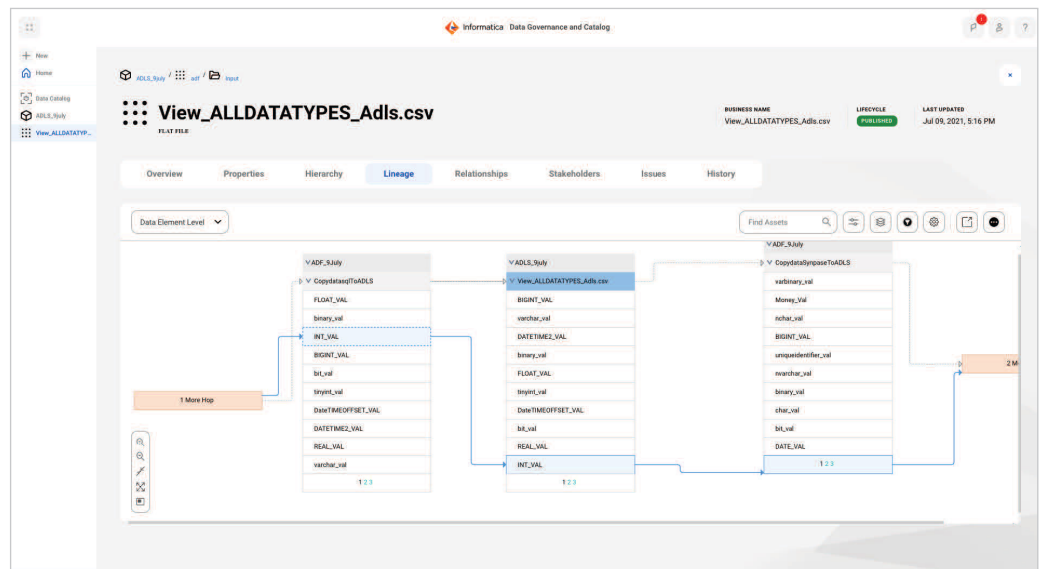


Figure 1. Explore data lineage from source to target.

AI Model Governance

AI model governance provides visibility into AI models, including their underlying algorithms, the training data used for creating the models and the relevant policies. It helps track and monitor model performance and key metrics like data drift that may lead to model performance degradation.

Workflows and Dashboards

The interactive and graphical dashboards put the user in command, providing summarized information in a visual form, including stakeholder/owner assignments and glossary metrics. Users can also monitor automated pre-defined workflows, check task completions and view notifications. With a variety of visualizations and drill-down capabilities, users can quickly ascertain summary status and see the next level of detail.

Key Benefits — Informatica on AWS

Quickly Meet Demands and Scale

With Cloud Data Governance and Catalog, data teams can quickly adopt and scale the solution based on organizational needs, allowing them to govern data and AI models efficiently and cost-effectively where they reside. The cloud-native solution enables increased access to trusted data, allowing them to make more informed decisions and rapidly start achieving value upon implementation.

Drive Business Decisions with a Comprehensive Understanding of Data

Cloud Data Governance and Catalog is designed to deliver a multi-dimensional view of your data. The powerful semantic search helps discover the most relevant data assets. Data lineage views help you understand the full context of your data flow including source, how data sets are built, the quality of data sets and its transformations along the data journey. This capability ensures that business decisions are driven by accurate, complete and trustworthy data. It also gives them the ability to automatically associate business glossary terms and gain insight into stakeholders and relationships, policies and classifications, making rich business context available to all data consumers.

Enable Trusted AI With Integrated Data and AI Model Governance

In this age of data science, AI models are often opaque, built with poor quality datasets and potentially non-compliant with organizational principles. The AI Model Governance capability provides insights into AI models and the data used to train the models. Insights are also provided for outputs produced, potential impacts of related policies and which models are available for reuse. This approach ensures models consumed are relevant, their lineage understood and policies applied are checked. It also provides visibility into any data drift to check the impact on the model's prediction capability. Informatica offers the only holistic solution for integrated governance of AI models as well as the data utilized by the models.

About Informatica

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. Informatica. Cloud First. Data Always.™

Establish an Analytics Governance Framework

A key benefit of Cloud Data Governance and Catalog is its ability to facilitate the establishment of an Analytics Governance Framework. Its interactive dashboard helps view, track and report the metrics required for monitoring governance. This framework enables you to define KPIs for analytics, create glossary hierarchies for context, define policy hierarchies and terms of use for data consumption and allows you to set up automated workflows that can be triggered by events and changes.

Discover Sensitive Data Sets for Assessment of Privacy Risks

Enable stewards to quickly identify datasets and data sharing lineage that may indicate privacy risks to gain more insight into policy compliance for sensitive information. With increased transparency, your data protection planning can help limit customer information exposure and avert risks from abuses and data loss.

Break Down Silos and Improve Data Intelligence

Facilitate data sharing and empower data consumers with trusted data. Every major decision made by your organization should be built on a foundation of trusted data and insights. It is paramount to ensure that your entire team across the globe can access the information that they need. With Cloud Data Governance and Catalog on AWS, your organization will be able to find, understand, govern and trust the data they need.

Next Steps

To learn more, please visit Informatica's [Cloud Data Governance and Catalog](#) solution page and [Informatica on AWS](#).



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

IN17_0922_04301

© Copyright Informatica LLC 2022. Informatica and the Informatica logo 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.