

Architecting a Governed Data Lake With Informatica on AWS

Key Benefits

- Accelerate data migration and digital transformation with enterprise-wide data discovery, integration, and quality
- Unleash the power of your governed data for greater business agility
- Enable analytics, insights, and self-service for all users

Discover, Govern, Catalog, Protect, and Scale Your Data Lake on AWS

Today's organizations rely on trusted, governed data to make quick, accurate decisions that are critical to the business and its digital transformation.

As demand for data grows, Chief Data Officers (CDOs) and other data professionals must provide just-in-time data to their internal data consumers, while ensuring quality, controlling risk, and adhering to security and compliance requirements.

While a data lakehouse can store petabytes of raw data to support digital transformation, the data must be safe and used appropriately, so that it does not violate data owner and consumer trust. Without the right governance, data lake initiatives can become a barrier to successful digital transformation.

The Informatica® cloud-native Governed Data Lake Management Solution for Amazon Web Services (AWS) enables you to govern, protect, and scale a data lake on Amazon S3.

Simplify Data Lake Governance With a Single, Comprehensive Platform

Informatica offers the first and only end-to-end, AI-powered, metadata-driven, cloud-native Intelligent Data Management Cloud (IDMC) for AWS data lakes, providing unified mass data ingestion (streaming, file, and database), cloud ETL and ELT, advanced serverless data integration, application integration, cataloging, data quality and standardization, data governance and privacy, master data management, and Business 360 capabilities.

Informatica Intelligent Cloud ServicesSM (IICS), Informatica's cloud-native integration platform as a service (iPaaS), is the product instantiation of IDMC. IICS enables you to discover, catalog, ingest, cleanse, integrate, optimize, govern, protect, prepare, and master data for state-of-the-art business analytics.

The centralized integration platform:

- Directly ingests discovered assets into AWS through a low-code to no-code visual canvas
- Comes with more than 200 out-of-the-box high-performance connectors and accelerates data migrations to AWS (low-code to no-code implementation)

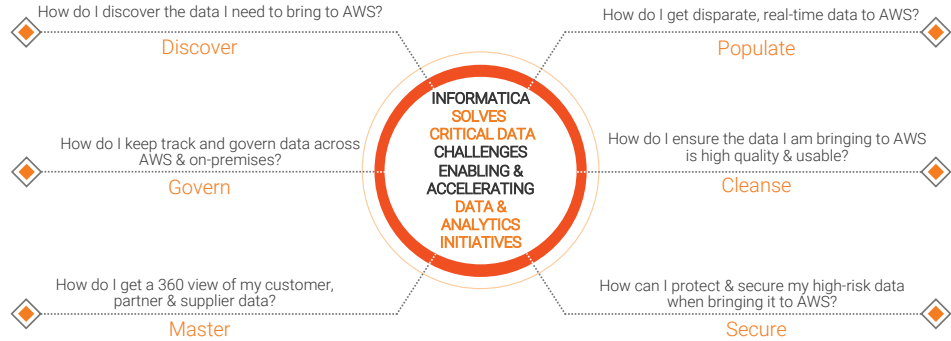


Figure 1. Informatica Addresses Critical Data Challenges for AWS Customers.

Key Features

Discover Data and Ensure Data Lineage

Gain visibility into data, location, lineage, and business context.

Informatica Enterprise Data Catalog enables data consumers to discover data automatically, leveraging AI and machine learning algorithms, and identify the right data for integration across cloud and on-premises systems. The solution automatically scans and collects metadata from all your enterprise systems, curating and augmenting the metadata with business context and inferring relationships and lineage.

With end-to-end data lineage and visibility, you can trace data from source to target, so that you can tell how data is being used and what is dependent upon it. Detailed impact analysis helps you understand the impact and cost benefits of migration and modernization, and collaboration and data certification by data owners and experts establishes what should be migrated.

With seamless integration between Informatica Enterprise Data Catalog and IICS, you can easily find, manage, and move data to your cloud data warehouse and data lake using a single codeless interface. Preconfigured mapping templates, easy-to-use task wizards for data replication and synchronization, and mass data ingestion for high-volume workloads enable moving more data to the cloud in less time.

Integrate and Optimize Data

Integrate and scale data and applications in minutes. Support new and complex integration patterns easily. Process data without moving it to an endpoint for analysis.

Informatica's unified cloud-native mass ingestion service leverages a wizard-based approach for ingesting large volumes of data into cloud repositories, and messaging hubs speed database synchronization and processing for real-time analytics.

Informatica Advanced Serverless Cloud Data Integration and Cloud Data Integration Elastic offer fully serverless ETL and ELT data processing, simplifying capacity management while enabling dynamic scaling, resulting in significant cost-savings and cost optimization. As a result, you can quickly and efficiently build data pipelines without hand coding and rapidly migrate on-premises data workloads to Amazon S3 data lakes.

Informatica Advanced Serverless enables customers to leverage serverless computing to process data integration pipelines. Users can run their data integration jobs (whether Spark-based or not) in serverless mode. Using a serverless mode eliminates the need to manage hardware or software and simplifies DevOps and DataOps, allowing developers to focus on business logic and deploy new data pipelines quickly while reducing costs.

You also can drive cloud analytics insights by moving the data on Amazon S3 and other enterprise systems to Amazon Redshift.

Informatica's Advanced Pushdown Optimization provides the data you need to power your business insights with a simple drag-and-drop GUI and the ability to convert mappings of source to target into native instructions and SQL queries, thus processing millions of records in just a few seconds and powering your business insights.

With enhanced pushdown performance, heavy transformations such as aggregate functions in a mapping can be pushed down to the Amazon Redshift endpoint using Amazon Redshift Connection. Enhanced pushdown to Amazon Redshift improves performance by using Amazon Redshift cluster processing power. Like other Informatica connectors, the IICS Amazon Redshift connector supports IAM, AWS Key Management Service (KMS) for server-side or client-side encryption.

Ensure Cloud-Native Data Quality

Automatically identify and fix data quality issues, ensuring access to clean, trusted data for analytics.

Informatica Cloud Data Quality can help you quickly profile your data and identify, fix, and monitor data quality problems—without coding—through prebuilt reference tables, wizards, and algorithms to deduplicate data assets before moving them into your Amazon Redshift data warehouse or Amazon S3 data lake. Using Informatica Cloud Data Quality, you can read data from Amazon S3, create data profiling rules, and perform data quality checks. The service provides profiling and cleansing of data to ensure that the data delivered is not only accurate, but certifiably accurate.

Enable Data Governance and Compliance

Document your governance program's policies, processes, stakeholders, and more to ensure consistent trust of your enterprise assets.

With Informatica Data Governance, operationalizing your data governance program for your Amazon data lake is easy. Collaborate with business and technical counterparts so that policies and processes are flexible enough to facilitate data-driven insights from your data lake and enable accuracy in analytics.

With the solution's built-in shopping experience, Informatica's Data Marketplace, internal data consumers can access governed data on AWS and have it delivered to them, where they need it for democratized, self-service analytics.

Informatica Data Governance provides data with deeper business meaning and relevance, automatically understands the lineage of data, and accelerates compliance and governance with risk assessment and protection/ownership policies. With automated workflow collaboration, users can identify changes from business or technical origins and define collaboration with drag-and-drop capabilities.

Informatica Data Marketplace enables organizations to create data sets, documents the appropriate quality and privacy standards, and publishes them within the appropriate categories to make them shoppable once users are authorized to consume and use the data.

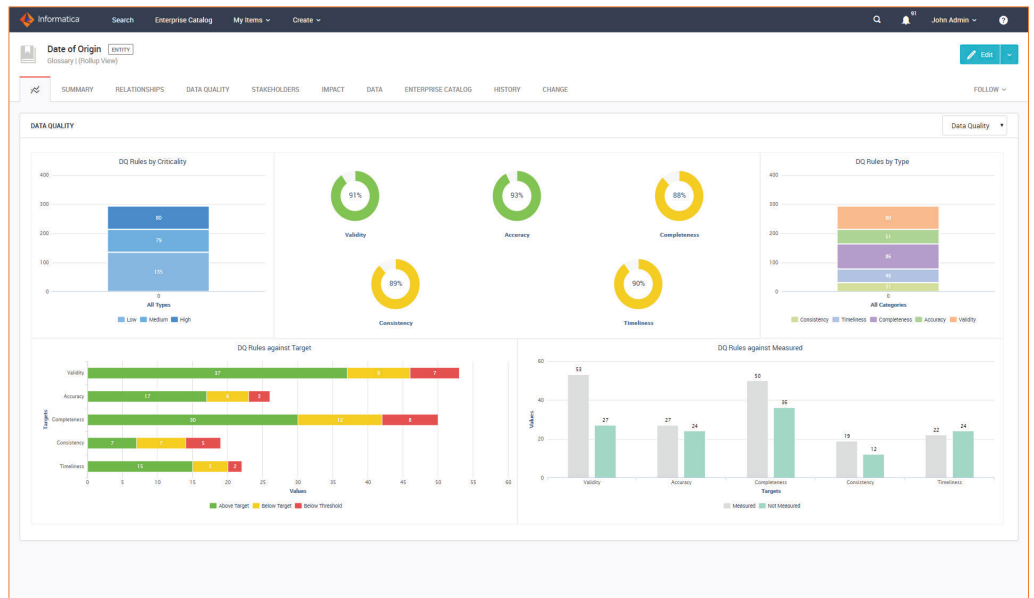


Figure 2. "Date of Origin" Data View.

The solution:

- Collects the knowledge already in your business
- Provides searchable, reportable, actionable, enterprise views
- Enables monitoring and reporting over time

Protect Data and Ensure Privacy

Define and automatically enforce data management and data privacy policies to ensure delivery of trusted data across the enterprise, while achieving compliance with regulations such as CCPA and the GDPR.

Informatica Data Privacy Management provides protection and monitoring of personal and sensitive data for privacy and compliance efforts. Its centralized visibility across data platforms and types provides the support needed for today's complex environments in the cloud. Continuous risk analysis of personal and sensitive data, AI-driven detection to uncover high-risk and anomalous data usage, and automated orchestration and protection of sensitive data intelligence remediate privacy and security risks.

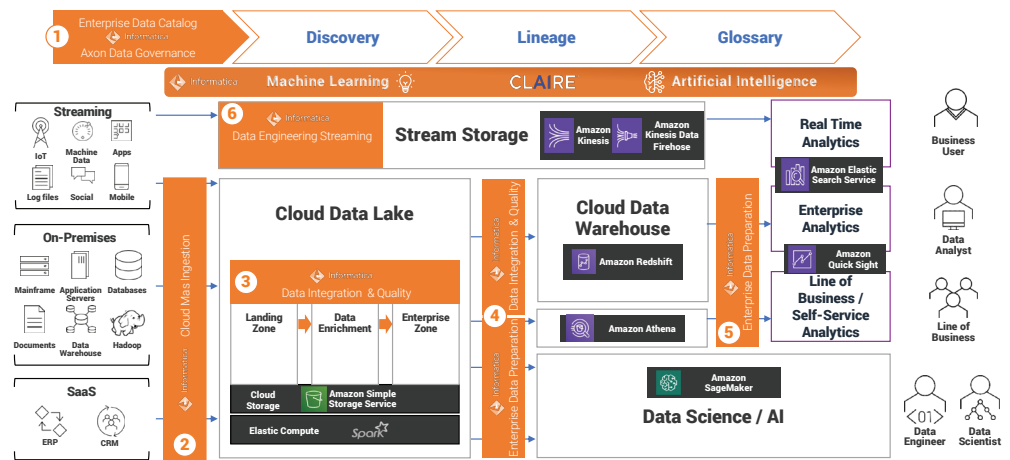


Figure 3. End-to-End Governed Data Lakehouse Architecture on AWS.

Master Your Data and Analytics

Grow the business value of your data with 360-degree insight into connections, interactions, and relationships.

Informatica cloud-native Master Data Management and 360 solutions on AWS leverage AI and machine learning that make it easier to fix, master, and manage your data for business and customer 360 views. Informatica manages all master domains and connects them with relevant business data such as transactions and interactions. The solution improves analytic accuracy from the shared foundation of consistent, connected, and trusted master data throughout the business. It also increases operational efficiency by replacing manual processes such as data onboarding, synchronization, and reporting with intelligence and automation.

With a single view of data to fuel your business and analytical applications with clean, consistent, and complete information, you benefit from a single, actionable, modern 360-degree view of relationships between customers, products, and suppliers and a centralized trusted application to onboard, manage, and share business-critical trusted data.

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.

Key Benefits

Accelerate Data Migration to Amazon S3 and Amazon Redshift and Digital Transformation With Enterprise-wide Data Discovery, Integration, and Quality

Informatica's AI-powered data catalog provides a machine learning-based discovery engine to scan and catalog data assets across the enterprise—across multi-cloud and on-premises.

Informatica provides users with powerful semantic search and dynamic facets to filter search results, data lineage, profiling statistics, data quality scorecards, holistic relationship views, data similarity recommendations, data asset analytics, and an integrated business glossary.

Unleash the Power of Your Governed Data for Greater Business Agility

Informatica is the industry leader in data governance, data quality, and metadata management, with an automated, end-to-end, AI-powered, metadata-driven, cloud-native platform for data warehouses and data lakes on AWS.

Informatica's data governance solution provides technology metadata management and comprehensive business glossary including end-to-end detailed lineage, automated data quality, data privacy, and compliance (for example, CCPA, GDPR, and CISO).

Enable AI/ML Analytics, Insights, and Self-Service for All Users

The solution solves critical data challenges, enabling and accelerating data and AI/ML analytics initiatives for every persona and user type.

Informatica's data governance solution is a persona-based offering that addresses customers' needs based on the personas of consumer citizen, data steward, ETL engineer, or data scientist.

Next Steps

Learn more about Informatica solutions for AWS at www.Informatica.com/AWS.

Listen to [data governance and privacy customer leaders across industries share real-world advice and lessons learned](#).



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

IN17_1021_04050

© Copyright Informatica LLC 2021. 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.