

# Intelligent Data Cataloging for Cloud Data Warehouses, Data Lakes, and Lakehouses

## Key Benefits

- Maximize cloud data warehouse, data lake, and lakehouse value without disruption
- Accelerate even complex cloud migration and modernization initiatives
- Gain visibility into all your enterprise data assets with industry-leading metadata management
- Enable users to easily find relevant data in the cloud

## Accelerate AI-powered Data Discovery and Cloud Data Integration

Enterprises today are rapidly moving their data and analytics infrastructure to the cloud. While this move to capitalize on cloud's agility, scalability, and cost-efficiency affects all aspects of data analytics, it's particularly urgent for data warehousing and data lakes. For years, enterprise data warehouses were stable workhorses, powering enterprise analytics and reporting systems. Now, a massive shift is underway to modernize them in the cloud to achieve dramatic improvements in performance and competitive advantage. At the same time, organizations are adopting cloud data lakes to go hand-in-hand with their cloud data warehouses. And more recently, organizations are building cloud lakehouses, which merge the best of data warehouses and data lakes to provide analytics capabilities to power everything from BI dashboards to advanced AI and machine learning projects.

However, adopting cloud data warehouses, data lakes, and lakehouses can present new challenges. Although it may be initially straightforward to stand up a new cloud data warehouse or data lake, maximizing the value of your investment requires strategy and planning.

Whether you are building a new cloud data warehouse, data lake, or lakehouse, or modernizing data and workloads in the cloud over time, it's essential to understand and assess your current data landscape and make sure you have the tools and best practices in place to manage your data once it's in the cloud.

The Informatica® Intelligent Cloud Lakehouse Data Management Solution for Cloud Analytics is built on a foundation of industry-leading data integration, data quality and data governance, and metadata management. Informatica Enterprise Data Catalog provides the metadata management foundation so you can understand your data landscape with machine learning-based data asset discovery and end-to-end visibility. Informatica Intelligent Cloud Services<sup>SM</sup> (IICS) provides market-leading enterprise iPaaS services for cloud data integration, cloud mass ingestion, and cloud data quality.

With machine-learning capabilities built on top of comprehensive metadata management, Informatica Enterprise Data Catalog provides a common enterprise metadata foundation for intelligent, automated, end-to-end visibility and lineage across your environment. The solution enables you to bridge on-premises data stores with your modern cloud ecosystem to achieve maximum value without disrupting your data consumers.

## Key Features

### Broad Metadata Connectivity

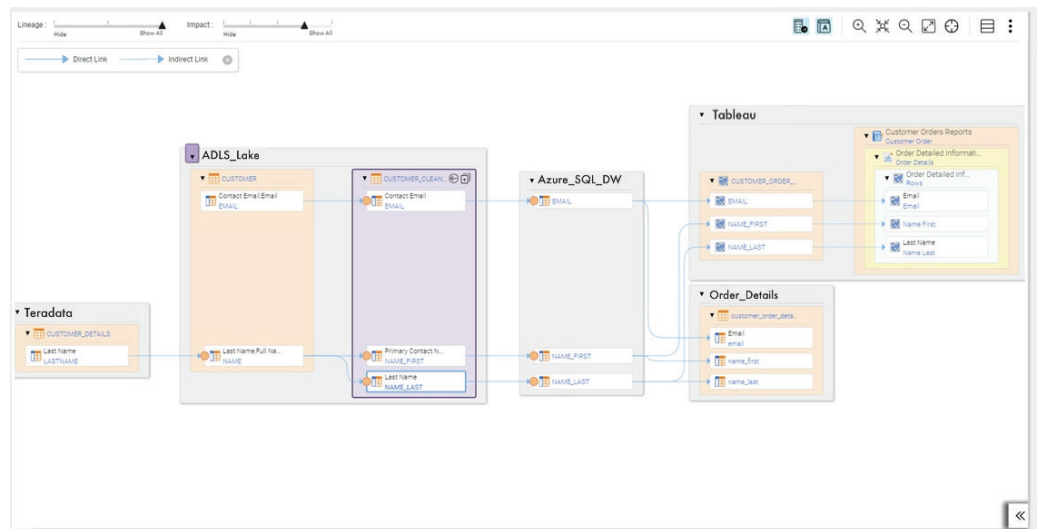
Informatica's intelligent data cataloging solution integrates with all the leading cloud ecosystems and modern data warehouses, data lakes, and lakehouses, including Amazon Redshift, Amazon S3, Microsoft Azure Synapse Analytics, Google BigQuery, Snowflake, and Databricks. It also includes out-of-box metadata connectivity to hundreds of cloud applications, on-premises systems, enterprise and middleware applications, data stores, and analytical BI tools.

### Machine Learning-Based Data Discovery

Informatica's metadata-driven artificial intelligence engine, CLAIRE®, applies AI and machine learning algorithms to technical, business, operational and usage data, automatically understanding and parsing complex files and metadata, enabling you to discover data regardless of where it resides across a complex, and multi-cloud environment.

### Automated End-to-End Data Lineage

Lineage gives you summary and detail views that enable you to understand where your data originated, how it moved through your organization, and how it was transformed along the way. End-to-end data lineage across on-premises and multi-cloud environments enables you to trace data from source to target, so that you can tell how data is being used and what is dependent upon it. Informatica Enterprise Data Catalog enables you to perform detailed impact analysis on upstream and downstream data assets. As a result, you can easily understand the impact of migration across data assets, resources, and users. In addition, impact analysis enabled by detailed data lineage can help you demonstrate the cost benefits of moving certain data assets and workloads to the cloud.



## Collaboration and Social Curation

Migrating data to a cloud data warehouse, data lake, or lakehouse typically involves multiple stakeholders and users. Informatica Enterprise Data Catalog provides collaboration and social curation capabilities that enable data owners and subject matter experts to certify data assets and help identify those that should be migrated. In addition, data consumers can rate and review data assets, adding custom attributes and annotations that make it easier to find trusted data during and after migration.

## Integration With Cloud Data Management Services

Informatica Enterprise Data Catalog integrates with IICS data management services that provide critical capabilities for cloud data warehouses, data lakes, and lakehouses. IICS Cloud Data Integration provides high-performance ETL, ELT, data ingestion, synchronization, and replication for multi-cloud environments. IICS Cloud Mass Ingestion supports streaming, file, and database ingestion to cloud data warehouses, data lakes, and message queues. IICS Cloud Data Quality enables you to profile, cleanse, and standardize data so you can quickly fix and monitor data quality issues. From within IICS, you can use Informatica Enterprise Data Catalog to discover and select data, view data sources and mappings, view data quality rules and scorecards, and gain comprehensive visibility into end-to-end data lineage.

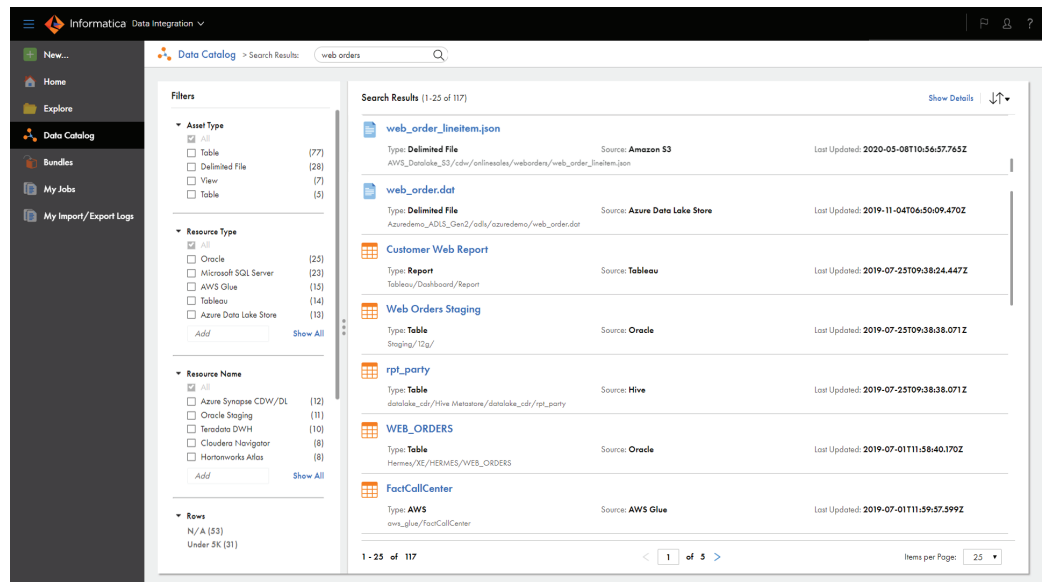


Figure 2. Discover and select data using Informatica Enterprise Data Catalog within IICS.

### **Preconfigured Mappings**

Preconfigured mapping templates, easy-to-use task wizards for data replication and synchronization, and mass data ingestion for high-volume workloads empower you to move more data to the cloud in less time. Seamless integration between Informatica Enterprise Data Catalog and IICS enables you to find the data you want to migrate and easily move it to your cloud data warehouse, data lake, or lakehouse using a single codeless interface. “Shop” your data catalog with IICS Cloud Data Integration’s visual designer, and then build simple to complex data loads in just a few clicks to connect diverse endpoints across cloud and on-premises, transform data, and schedule analytical workloads.

### **Key Benefits**

#### **Maximize Cloud Data Warehouse, Data Lake, and Lakehouse Value Without Disruption**

Comprehensively assess your data landscape to understand what to migrate or consolidate and where you’ll generate the most value from a cloud data warehouse, data lake, or lakehouse. With end-to-end data lineage, gain visibility into data relationships and dependencies. Empower data consumers with self-service access to easily find data for analytics in the cloud. Informatica Enterprise Data Catalog enables you to ensure that structures and schemas from your on-premises data warehouse are transferred to the cloud, thus avoiding potential disruption during and after migration. After migration, end-to-end lineage views help you validate that data has been mapped correctly. In addition, the metadata-driven automated data discovery and curation supports scalable data governance and enables delivery of trusted data in the cloud.

#### **Accelerate Even Complex Cloud Migration and Modernization Initiatives**

Significantly decrease the time needed for cloud migration and modernization. Discover the data you want to migrate or consolidate, then build simple to complex data integration loads using a visual mapping designer that connects diverse endpoints across cloud and on-premises with pre-built transformations to accelerate development.

#### **Gain Visibility Into All Your Enterprise Data Assets With Industry-Leading Metadata Management**

You need to discover and understand your data before you can derive value from it. The Informatica solution leverages industry-leading metadata management capabilities, so you’re able to leverage hundreds of connectors for cloud-based and on-premises systems. With metadata-driven intelligence, the solution helps you identify all the places across all of your organization’s systems where data is stored and used, then you can move all of that data seamlessly to your enterprise cloud data warehouse, data lake, or lakehouse. In addition, it captures and stores critical metadata about contingencies and relationships between datasets and subsets to maintain the connections among them during and after data migration and modernization in the cloud.

## 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.

## Enable Users to Easily Find Relevant and Trusted Data in the Cloud

When data is moved during a migration, data consumers may have trouble locating the data they depend on to do their jobs. Bookmarked links and recurring reports can break. And data consumers might continue to use data from a familiar on-premises source, even though cloud now provides a better alternative. Data analysts and data scientists also need to know where data is located and understand which data assets are best suited for their needs. By consulting Informatica Enterprise Data Catalog, users can easily find the data they need without having to ask IT for help, and therefore avoid bottlenecks and inefficiencies.

## Next Steps

Learn more by reading the [Informatica Enterprise Data Catalog data sheet](#) and the [Modernize Your Cloud Data Warehouse and Data Lake in the Cloud](#) solution brief.



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

IN17\_1120\_03625

© Copyright Informatica LLC 2020. Informatica, the Informatica logo, Informatica Intelligent Cloud Services, 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.