

Informatica Enterprise Data Catalog

Benefits

- Automatically catalog and classify all types of data across the enterprise using an AI-powered catalog
- Identify domains and entities with intelligent curation
- Enrich data assets with governed and crowdsourced annotations
- Find data assets through powerful Google-like semantic search
- Discover and understand your data assets with data profiling and quality stats, 360 relationship views and lineage
- Get a complete picture of your data environment
- Open APIs to integrate into your environment and expose intelligent metadata anywhere

Unleash the Power of Data with an Intelligent Data Catalog

Data is the lifeblood of our economy, and data-driven companies turn their data assets into revenue and profits. The first step in any data-driven digital transformation initiative is to manage your data as an enterprise asset: take inventory of it, assess its value, and maximize its use—just like you do with other significant capital and operational investments.

Data is diverse and distributed across many different departments, applications, data warehouses (some on-premises, others in the cloud), making it a challenge to know exactly what data you have and where. In the world of big data this becomes even more complex.

Informatica Enterprise Data Catalog is an AI-powered data catalog that provides a machine-learning-based discovery engine to scan and catalog data assets across the enterprise—across cloud and on-premises, and big data anywhere. The intelligence in Enterprise Data Catalog is provided by the CLAIRE™ engine, which provides intelligence in terms of leveraging metadata to deliver intelligent recommendations, suggestions and automation of data management tasks. This enables IT users to be more productive and business users to be able to be full partners in the management and use of data.

Informatica Enterprise Data Catalog provides business and IT users with powerful semantic search and dynamic facets to filter search results, data lineage, profiling statistics, 360-degree relationship views, data similarity recommendations, and an integrated business glossary. You can now easily and efficiently manage enterprise data assets to maximize their value throughout the company. Business users can quickly find data and easily manage the lifecycle of business terms, definitions, reference data, and more.

Informatica Enterprise Data Catalog is an AI-powered data catalog that provides a machine-learning-based discovery engine to scan and catalog data assets across the enterprise—across cloud and on-premises and big data anywhere.

Key Features

Metadata APIs to Integrate into Your Environment

Enterprise Data Catalog includes REST-based APIs that enable you to integrate it into your environment and consume catalog content anywhere. Organizations can share any intelligent metadata—applications, BI reports, and dashboards—with business users.

Plug-in for Tableau to Enable Governance and Trust in Data

Informatica Enterprise Data Catalog for Tableau delivers agile, self-service analytics with governed data. It enables Tableau users to access the full resources of Enterprise Data Catalog when creating and delivering data visualizations. Business users who are consumers of Tableau reports get a complete side-by-side view of the business and technical context of the worksheet, dashboard, or data source within Tableau.

Semantic Search with Intelligent Facets

Find and discover the most relevant data sets for your analysis using powerful semantic search with intelligent facets. Advanced keyword search with token matching finds the most relevant data assets in the catalog. Semantic search is even applied to inferred data domains so no data asset is left undiscovered. Intelligent facets, based on the search results, allow users to alter the search to the data sets of interest.

Data Lineage and Impact Analysis

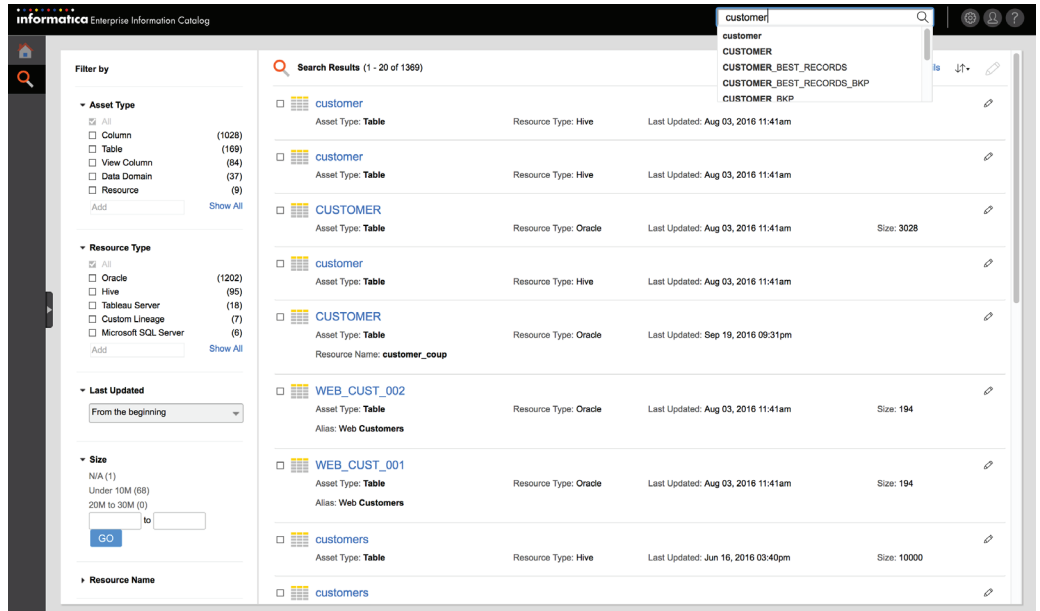
Interactively trace data origin through business-friendly summarized lineage views that highlight the end points and not all the complex details in between. A drill-down lineage view expands any lineage path to show columns and lineage diagram metrics. Users can perform detailed impact analysis on upstream and downstream data assets.

360-degree Relationship Discovery

Get a 360-degree view of data in a knowledge graph that lets you quickly search, discover, and understand enterprise data and meaningful data relationships. Automatically discover related data sets, technical, business, semantic and usage-based relationships. The 360-degree data view shows related datasets, tables, views, data domains, reports, and users. This aids in progressive discovery of other data sets of interest.

Automated Classifications with Intelligent Domain and Entity Recognition

Automatically classify and identify domains and entities such as customer, product, order etc. across all structured and unstructured data assets at the field, column and table level. This is a crucial step in the ability for companies to catalog, govern, and extract value from their data assets. This classified data enables better search, filtering of search results and business glossary recommendations. Informatica provides over 60 packaged data domains such as email, credit card number, social security number, country, city, URL, and company name. Users can add their own custom domains too. Data assets can be classified using data rules (i.e., columns with data that matches specific logic defined in the rule) or column name rules (i.e., Finds columns that match column name logic defined in the rule).



Quickly find data sets with smart semantic search and dynamic facets.

Integrated Data Quality Statistics

View data profiling statistics alongside technical metadata to understand the quality of data assets before using data for analysis. Profiling statistics include value distributions, patterns, and data type and data domain inference.

Business Glossary

Informatica Enterprise Data Catalog includes an integrated Business Glossary that provides a central place to define and manage the lifecycle of business terms, definitions, associated reference data, related terms, links, ad hoc documentation, and notes. Business Glossary allows business and IT stewards to collaboratively manage business metadata that includes efficient human workflow automation. Associate business terms with the right technical metadata and Informatica Enterprise Data Catalog will even recommend term associations. Business glossary assets such as terms, policies, and classifications can be easily imported from Informatica Business Glossary and third party tools.

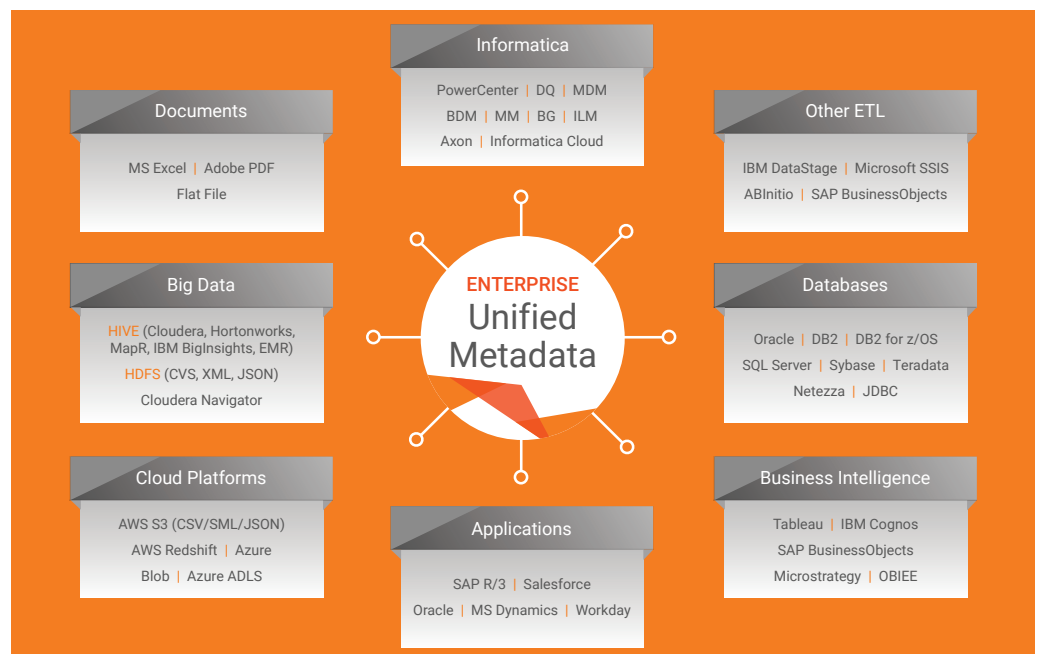
Intelligent Data Similarity

Advanced statistical and machine learning algorithms identify similar data and subsets of data. This powerful capability helps users find the most relevant and trusted data they need. For example, a telecom analyst interested in customer churn analysis might query data containing pre-paid customer activity for the current quarter. Informatica Enterprise Data Catalog can recommend a cleaner version of the data (substitute data), data containing customer activity for the previous quarter (unionable data), and a customer detail table to enrich the data set (joinable data).

Universal Metadata Connectivity

Extract metadata from any type of data sources across the enterprise such as databases, data warehouses, applications, cloud data stores, BI tools, Hadoop and NoSQL, and more. Below are some examples of data sources supported for metadata extraction:

- **Databases:** Oracle, IBM DB2 LUW, SQL Server, Sybase ASE, Netezza, Teradata, JDBC, MySQL, Amazon Redshift, Azure SQL DB, Azure SQL DW
- **Hadoop:** Cloudera Navigator, Hive(Cloudera/HW/ MapR/HDInsights/EMR), HDFS, Hortonworks Atlas
- **Mainframes:** DB2 z/OS, DB2 i5/OS
- **BI:** SAP BusinessObjects, Tableau, Cognos, Microstrategy, OBIEE
- **File systems:** HDFS, Amazon S3
- **Applications:** Salesforce, SAP



Custom Attributes with Business Classifications

Enrich data sets by crowdsourced or expert classifications, comments, and other attributes available to anyone with appropriate security permissions. Assigning custom attributes and annotations to data sets including business glossary terms enhances business-IT collaboration and search results.

Resource-Level Security

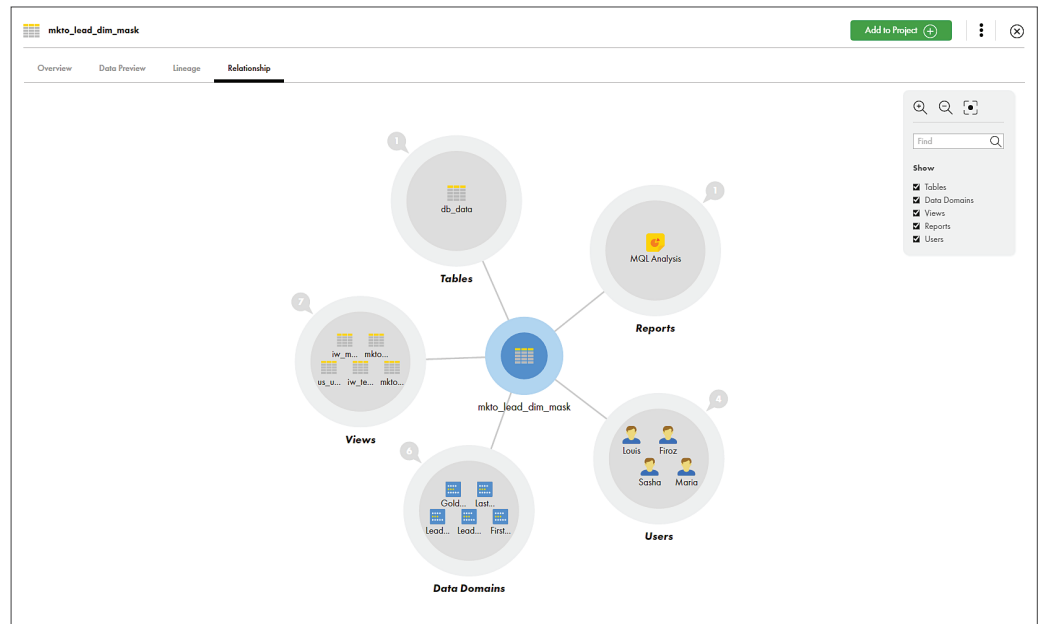
Grant user and group read/write permissions at the resource level to allow users to view or edit custom attributes, perform domain curation, and associate business glossary terms.

Big Data Scale Deployments

Enterprise Data Catalog is built for big data scale deployments that can be deployed on Hadoop clusters. Supports parallel metadata ingestion and high-speed distributed indexing to quickly update catalog content and deliver unmatched search performance. Provides fault tolerant high availability for 24x7 implementations.

Unified Administration

Manage and monitor the catalog resources, metadata extract schedules, profiling runs and more from one unified admin console. A job control dashboard provides widgets for task monitoring and resource views. Email alerts assist administrators in proactively responding to catalog issues.



Understand your data with complete 360-degree data relationship views.

Key Benefits

Intelligently Catalog All Types of Data Across the Enterprise

Informatica Enterprise Data Catalog intelligently discovers many types of data and their relationships across the enterprise. Pre-built scanners collect metadata from databases, data warehouses, applications, cloud data stores, BI tools, Hadoop and NoSQL, and more. All the metadata is indexed and cataloged in a highly scalable graph database architected for fast updates, smart search, and fast queries. As more and more data is created and propagated throughout the enterprise, similar and duplicate data sets inevitably arise. Informatica Enterprise Data Catalog leverages advanced statistical and machine learning algorithms to discover similar data and subsets of data, helping users find the most relevant and trusted data they need.

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.

Find Data Assets Quickly Through Powerful, Google-Like Semantic Search

Trying to find the data you need across hundreds of enterprise systems may sometimes seem futile. Only through powerful semantic search built on comprehensive metadata services and a scalable infrastructure can one even hope to find relevant data. Informatica Enterprise Data Catalog delivers semantic search with intelligent facets to further refine search results. Because Informatica uniquely associates business, technical, and operational metadata, business users can search on business terms to find their data and then browse 360-degree relationship views to find related data assets.

Discover and Understand Your Data Assets with 360-degree Relationship Views and Lineage

The classic saying, "You can't manage what you can't measure" is true when it comes to managing data assets. To get the most value from data, you need to understand what you have, where it came from, how it has changed, and what level of trust you have in the data. Informatica Enterprise Data Catalog answers all these questions and more with complete end-to-end summary and detail lineage, profiling statistics, and 360-degree relationship views, providing a clear picture of your data.

Enrich Data Assets with Business Context Through Governed and Crowdsourced Annotations

Informatica Enterprise Data Catalog (EDC) maximizes the reuse and value of data by automatically classifying enterprise data assets down to the field/column level. To further increase the value of data, EDC captures the context of who is using the data and for what purpose along with crowdsourcing tags and annotations. This "wisdom of crowds" helps to enrich and curate data, making it even more valuable throughout the enterprise. Informatica Enterprise Data Catalog includes an intuitive business-friendly Business Glossary providing a central place to define and manage the lifecycle of business terms, definitions, associated reference data, and more. This business metadata is associated with technical metadata and operational metadata so that business analysts, data stewards, and other users can quickly find, understand, and collaborate on data assets.



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

IN06_0418_03238

© Copyright Informatica LLC 2018. Informatica, the Informatica logo and the CLAIRE engine are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. 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.