

Informatica Enterprise Data Catalog

Key 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

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 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

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).

The screenshot shows the Informatica Enterprise Data Catalog search interface. The search bar at the top right contains the text 'customer'. A dropdown menu below the search bar lists suggestions: 'customer', 'CUSTOMER', 'CUSTOMER_BEST_RECORDS', 'CUSTOMER_BEST_RECORDS_BKP', and 'CUSTOMER_BKP'. The main search results area displays a list of results for 'customer' (1 - 20 of 1369). The results are shown in a table with columns for Asset Type, Resource Type, Last Updated, and Size. The left sidebar contains filter options for Asset Type, Resource Type, Last Updated, and Size.

Asset Type	Resource Type	Last Updated	Size
customer	Hive	Aug 03, 2016 11:41am	
customer	Hive	Aug 03, 2016 11:41am	
CUSTOMER	Oracle	Aug 03, 2016 11:41am	3028
customer	Hive	Aug 03, 2016 11:41am	
CUSTOMER	Oracle	Sep 19, 2016 09:31pm	
customer_coup			
WEB_CUST_002	Oracle	Aug 03, 2016 11:41am	194
WEB_CUST_001	Oracle	Aug 03, 2016 11:41am	194
customers	Hive	Jun 16, 2016 03:40pm	10000
customers			

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

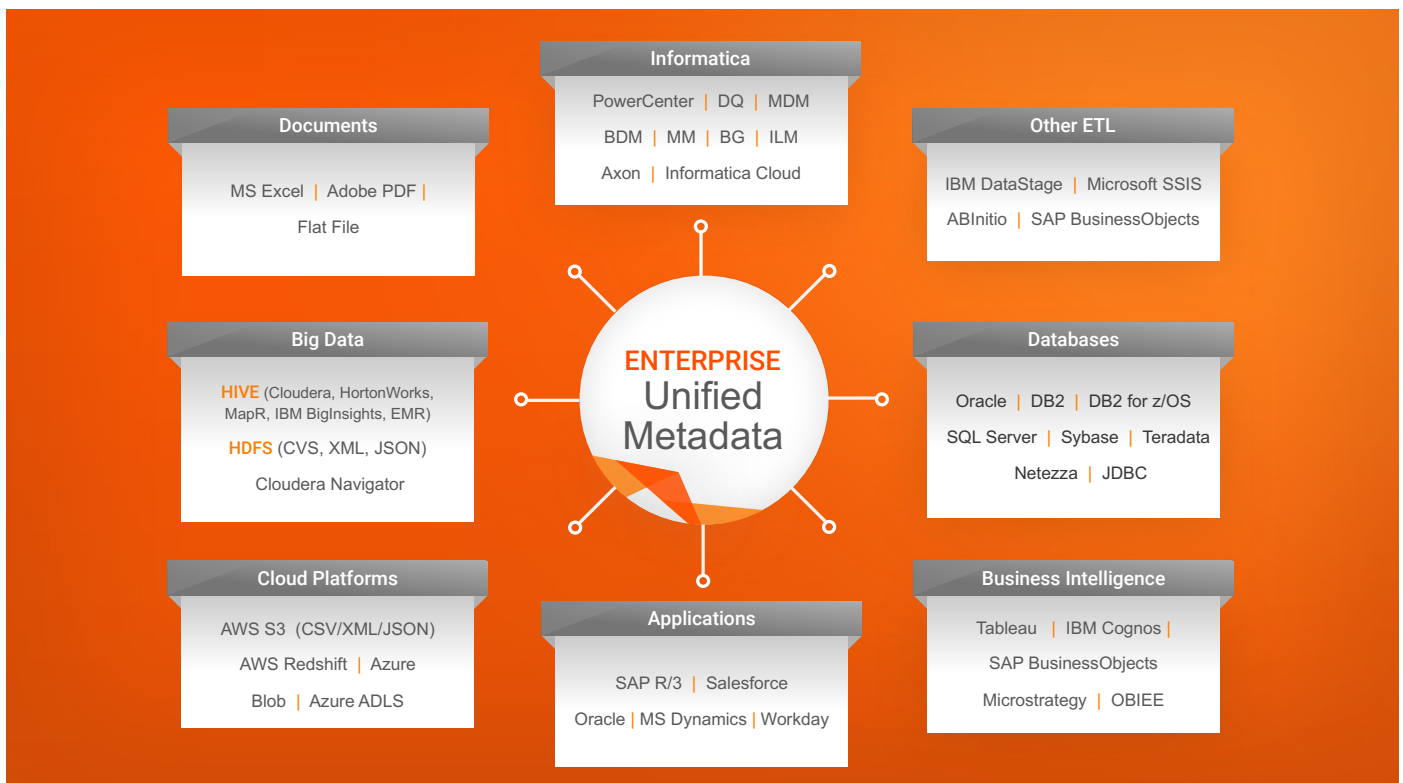
Advanced statistical and machine learning algorithms identify similar data and subsets of data. This powerful capability helps users find the most relevant and trusted data they need.

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 Business Objects, 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

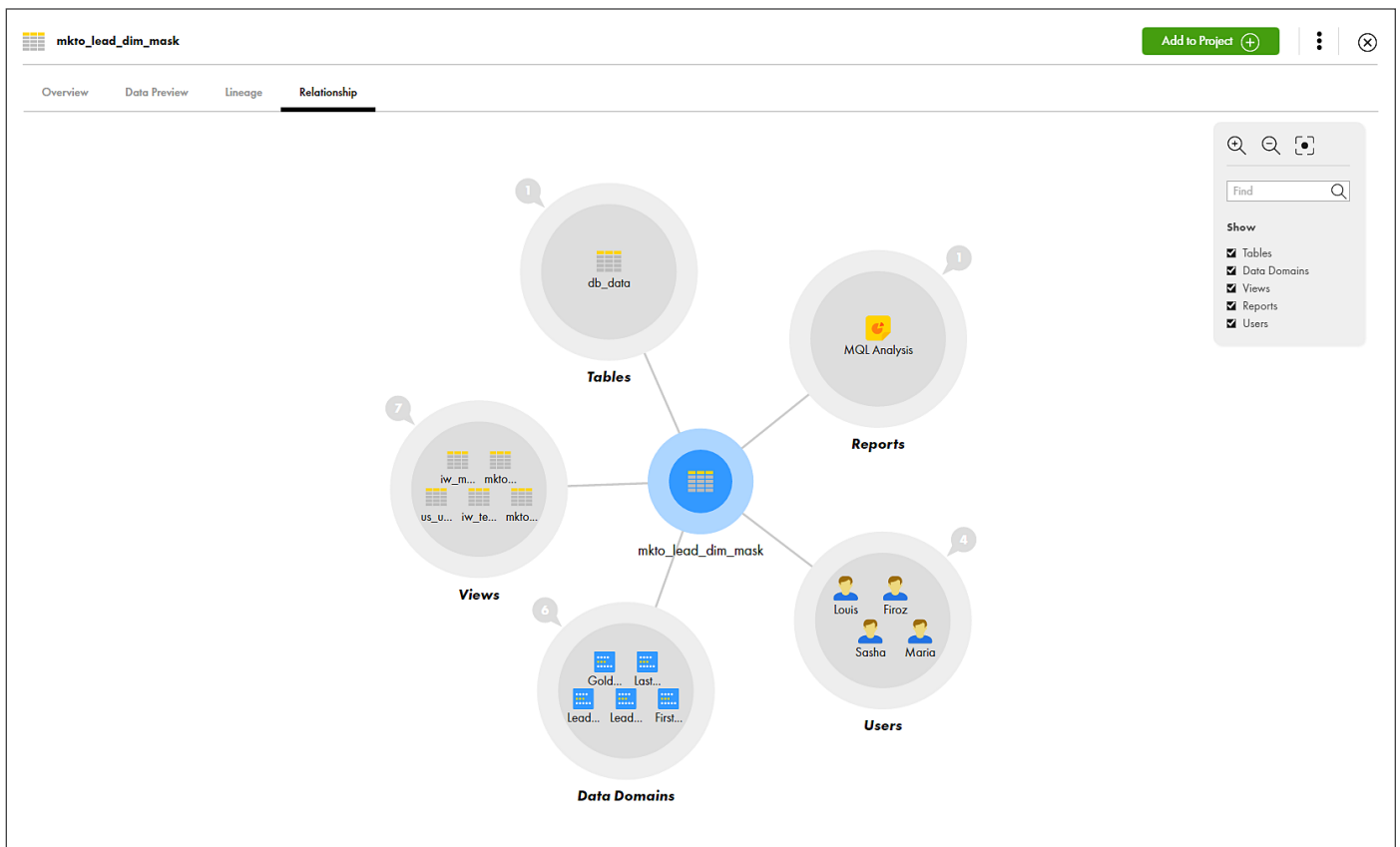
Grant user and group read/write permissions at the resource level to allow users to view or edit custom attributes, perform domain curation, and associate business glossary terms.

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.

About Informatica

Digital transformation is changing our world. As the leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead the way. To provide you with the foresight to become more agile, realize new growth opportunities or even invent new things. We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption. Not just once, but again and again.

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.

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 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 (EIC) 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, EIC 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.

