Informatica Enterprise Data Catalog

Benefits

- Automatically catalog and classify all types of data across the enterprise using an AI-powered catalog

- Provide a metadata system of record for the enterprise with a catalog of catalogs

- Automatically extract the most granular metadata from a wide array of data sources, including complex enterprise systems

- Find data assets through powerful Google-like semantic search

- Discover and understand your data assets with a holistic view including lineage, relationship views, and data profiling stats and quality scorecards

- Identify domains and entities with intelligent curation

- Enrich data assets with governed and crowdsourced annotations, ratings, and reviews

- Automatically associate business glossary terms to technical data assets

- Open APIs to integrate into your environment and expose intelligent metadata anywhere

- Measure and optimize the value of your data assets with Data Asset Analytics

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, and data warehouses and data lakes (some on-premises, others in the cloud), making it a challenge to know exactly what data you have and where. As data sources proliferate, the data landscape 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 multi-cloud and on-premises. Enterprise Data Catalog is powered by the CLAIRE® engine, which provides intelligence by leveraging metadata to deliver recommendations, suggestions, and automation of data management tasks. This enables IT users to be more productive and business users to be full partners in the management and use of data.

Informatica Enterprise Data Catalog provides data analysts and IT users with powerful semantic search and dynamic facets to filter search results, detailed data lineage, profiling statistics, data quality scorecards, holistic relationship views, data similarity recommendations, and an integrated business glossary.

Collaboration capabilities leverage subject matter expertise and social curation combined with the power of AI to guide user experience and automate data curation. Users can quickly find data and easily manage the life cycle of business terms, definitions, reference data, and more.

With Data Asset Analytics in Enterprise Data Catalog, you get insights on the usage of data within your organization, enabling you to proactively manage and optimize the value of your data assets.

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.

Key Features

Semantic Search With Intelligent Facets

Find and discover the most relevant datasets for your analysis using powerful semantic search with intelligent facets. Advanced keyword search with token matching finds the most relevant data assets, and semantic search is even applied to inferred data domains. Intelligent facets, based on the search results, allow users to narrow the search to the datasets of interest.

Holistic Relationship Discovery

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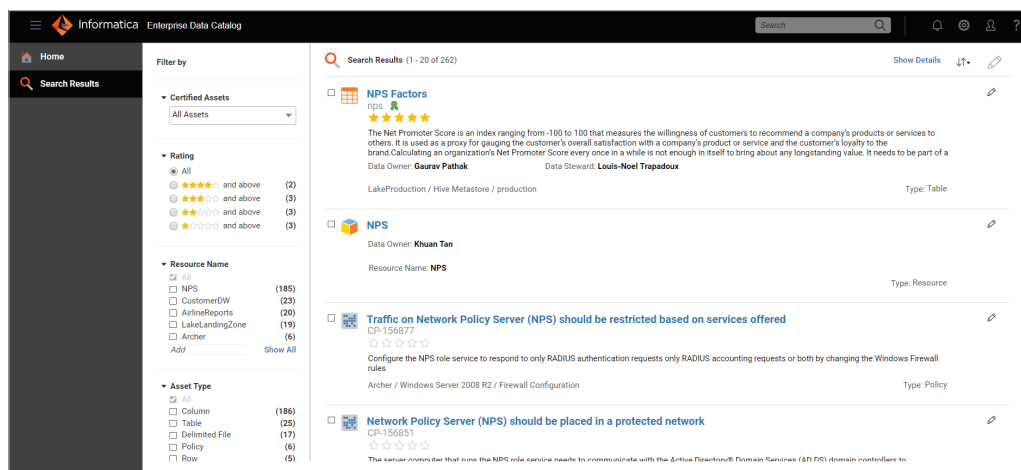

Figure 1: Quickly find datasets with smart semantic search and dynamic facets. View ratings and certified datasets.

Data Lineage and Impact Analysis

Interactively trace data origin through lineage views at any level—from business-friendly, system-level views that highlight the endpoints to granular views that include all the complex details in between. A drill-down lineage view expands any lineage path to show granular column- and metric-level lineage. Users can perform detailed impact analysis on upstream and downstream data assets.

Collaboration and Social Curation

Informatica Enterprise Data Catalog empowers data analysts and data scientists to easily find the most relevant and trusted data for analytics by harnessing the combined power of AI and human expertise and collaboration. Data owners and subject matter experts can certify datasets. Data consumers can provide ratings and reviews for datasets enabling social curation of data.

Users can follow datasets of interest and get notified of changes, and a Q&A platform allows subject matter experts to answer common questions from users. In addition, users can add custom attributes and annotations to datasets, further enhancing business-IT collaboration and search results.
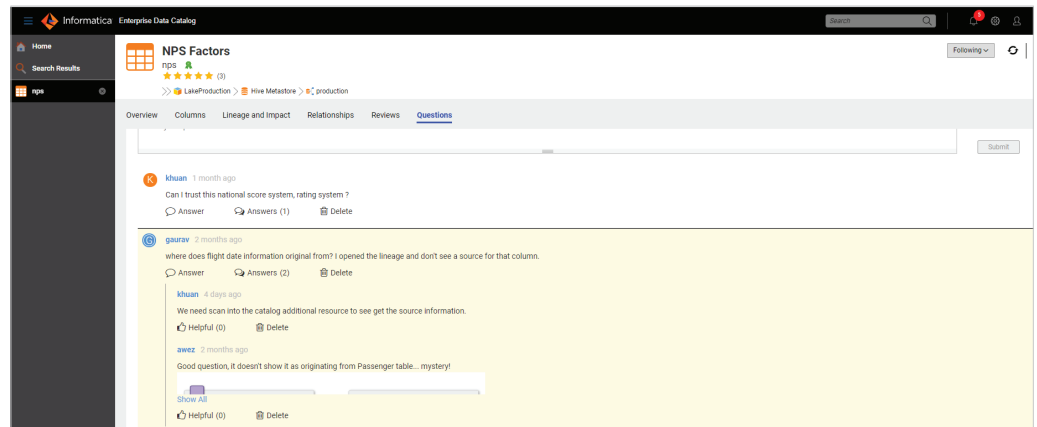

Figure 2: Enable collaboration with Q&A capabilities.

Integrated Data Quality

View data profiling statistics, data quality rules, scorecards, and metric groups 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.

Automatic Association of Business Glossary Terms

Informatica Enterprise Data Catalog allows for easy import of business glossary assets such as terms, policies, and classifications from Informatica Axon™ as well as third-party tools. Add rich business context to the data by automatically associating business terms with the right technical metadata, eliminating a tedious manual process and allowing business and IT stewards to collaboratively manage business metadata that includes efficient human workflow automation.

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 (union-able data), and a customer detail table to enrich the dataset (joinable data).

Data Asset Analytics for Data Value

Data Asset Analytics provides prepackaged reports and dashboards on data asset inventory, usage, enrichment, level of collaboration, and more. Reports are extensible and can be exported, enabling data leaders to share business adoption and value metrics with stakeholders. Automated Data Value Calculator, a first-of-its-kind capability, allows an enterprise to measure and optimize the value of its data assets based on key factors that impact data value.

Universal Metadata Connectivity

Enterprise Data Catalog offers deep and broad metadata connectivity that spans on-premises, hybrid, and multi-cloud environments. Extract metadata from any type of data source such as databases, data warehouses, cloud data lakes, BI tools, Hadoop clusters, NoSQL, and complex enterprise systems including legacy and mainframe systems, multi-vendor ETL tools, SQL dialects, various enterprise applications, and stored procedures.

With Enterprise Data Catalog Advanced Scanners, you can visually inspect every script, procedure, or process to fully understand its logic and internal data flow. You can obtain a complete column-level data lineage, including a full inventory of all the potential lineage sources with rich details. The Advanced Scanners allow you to scan both static and dynamic code, as well as perform language parsing to obtain automated data lineage.

Data sources supported include:

- **Databases/Data warehouses:** Oracle, MS SQL Server, SQL Scripts, Sybase ASE, IBM Netezza, Teradata, JDBC, SAP HANA, SAP BW, SAP BW/4HANA, Snowflake, Stored Procedures
- **Big Data:** Cloudera Navigator, Hive (Cloudera/Hortonworks/MapR/IBM BigInsights/EMR), HDFS, Hortonworks Atlas, Cassandra, MongoDB, Kafka, Greenplum
- **Mainframes:** DB2 z/OS, DB2 i5/OS, COBOL, JCL
- **BI and Analytics:** SAP BusinessObjects, Tableau, Microsoft Power BI, Cognos, MicroStrategy, OBIEE, QlikView, Qlik Sense, Microsoft SSRS and SSAS, SAS
- **ETL:** Informatica PowerCenter®, Informatica Data Engineering Integration, Informatica Intelligent Cloud Servicessm, Informatica Data Integration Hub, Microsoft SSIS, IBM InfoSphere DataStage, Oracle Data Integrator, Talend Data Integration, AWS Glue
- **Business Glossary:** Informatica Axon Data Governance, Informatica Business Glossary
- **Data Modeling:** Erwin Data Modeler, SAP PowerDesigner
- **Enterprise Applications:** Salesforce, Oracle, Workday, Informatica MDM, SAP ECC, SAP S/4 HANA
- **File Systems:** Microsoft SharePoint, Microsoft OneDrive, Windows/Linux Filesystems
- **File Formats:** MS Excel, MS Word, MS PowerPoint, Adobe PDF, Flat Files, CSV, Delimited, XML, JSON, Avro, Parquet
- **Cloud Platforms:** AWS S3, AWS Redshift, Azure SQL DB, Azure Synapse Analytics, Azure ADLS, Azure ADLS Gen 2, Azure Blob, Google Cloud Storage, Snowflake, Google BigQuery
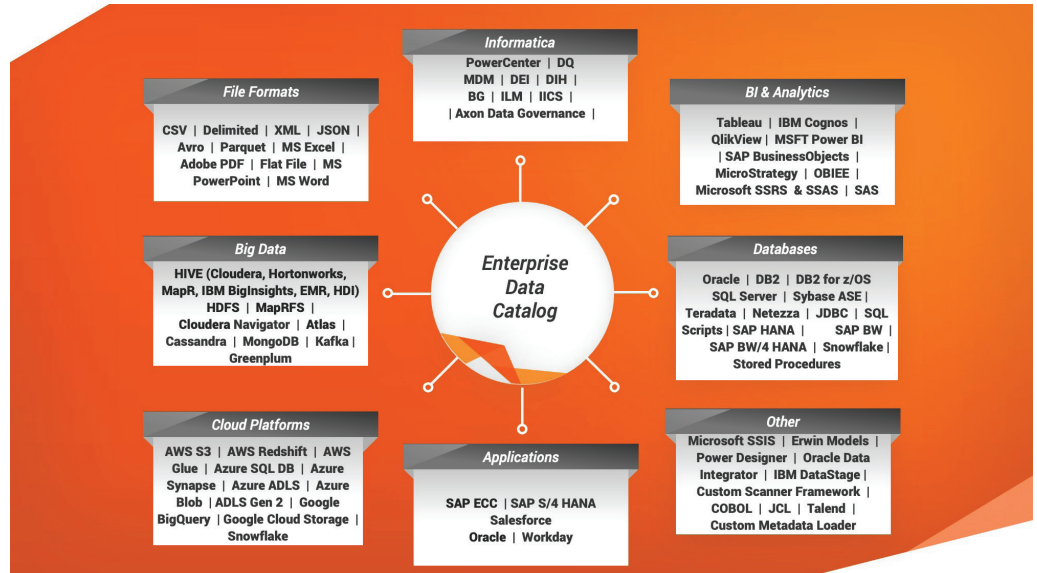
Informatica
PowerCenter | DQ
MDM | DEI | DIH |
BG | ILM | IICS |
| Axon Data Governance |

File Formats

CSV | Delimited | XML | JSON |
Avro | Parquet | MS Excel |
Adobe PDF | Flat File | MS
PowerPoint | MS Word

BI & Analytics

Tableau | IBM Cognos |
QlikView | MSFT Power BI
| SAP BusinessObjects |
MicroStrategy | OBIEE |
Microsoft SSRS & SSAS | SAS

Big Data

HIVE (Cloudera, Hortonworks,
MapR, IBM BigInsights, EMR, HDI)
HDFS | MapRFS |
Cloudera Navigator | Atlas |
Cassandra | MongoDB | Kafka |
Greenplum

Enterprise Data Catalog

Databases

Oracle | DB2 | DB2 for z/OS
SQL Server | Sybase ASE |
Teradata | Netezza | JDBC | SQL
Scripts | SAP HANA | SAP BW |
SAP BW/4 HANA | Snowflake |
Stored Procedures

Cloud Platforms

AWS S3 | AWS Redshift | AWS
Glue | Azure SQL DB | Azure
Synapse | Azure ADLS | Azure
Blob | ADLS Gen 2 | Google
BigQuery | Google Cloud Storage |
Snowflake

Applications

SAP ECC | SAP S/4 HANA
Salesforce
Oracle | Workday

Other
Microsoft SSIS | Erwin Models |
Power Designer | Oracle Data
Integrator | IBM DataStage |
Custom Scanner Framework |
COBOL | JCL | Talend |
Custom Metadata Loader

Figure 3: Informatica Enterprise Data Catalog supports universal metadata connectivity.

Self-Service Data Provisioning

After you find the relevant datasets for your analysis, easily move your dataset to the target
of your choice with simple click-through provisioning from within Informatica Enterprise
Data Catalog. You can choose from a broad choice of sources and targets including Amazon
Redshift, Azure Synapse Analytics, Google BigQuery, Snowflake, and BI tools like Tableau.
This capability leverages the integration of Informatica Enterprise Data Catalog with Informatica
Cloud Data Integration.

Metadata APIs to Integrate Into Your Environment

Informatica 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. Users can
export and share selected catalog content and associated enrichment metadata.

Tableau Integration for Governed Self-Service Analytics

The Chrome browser plug-in and Tableau extension for Informatica Enterprise Data Catalog
provide two different options for Tableau users to access the full resources of Informatica
Enterprise Data Catalog from within the native Tableau user interface. Without leaving the
Tableau interface, users can leverage an intelligent search bar to find trusted data assets, access
business and technical context, and collaborate with their peers.

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.

Enterprise-Scale Deployments

Informatica Enterprise Data Catalog is built for true enterprise-scale deployments with the ability to scan tens of millions of datasets across hundreds of data sources. It supports parallel metadata ingestion and high-speed distributed indexing to quickly update catalog content and deliver unmatched search performance and fault tolerant high availability for 24x7 implementations. With Spark-based data profiling, you can profile massive amounts of data at scale to get a deeper understanding of enterprise data.

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.
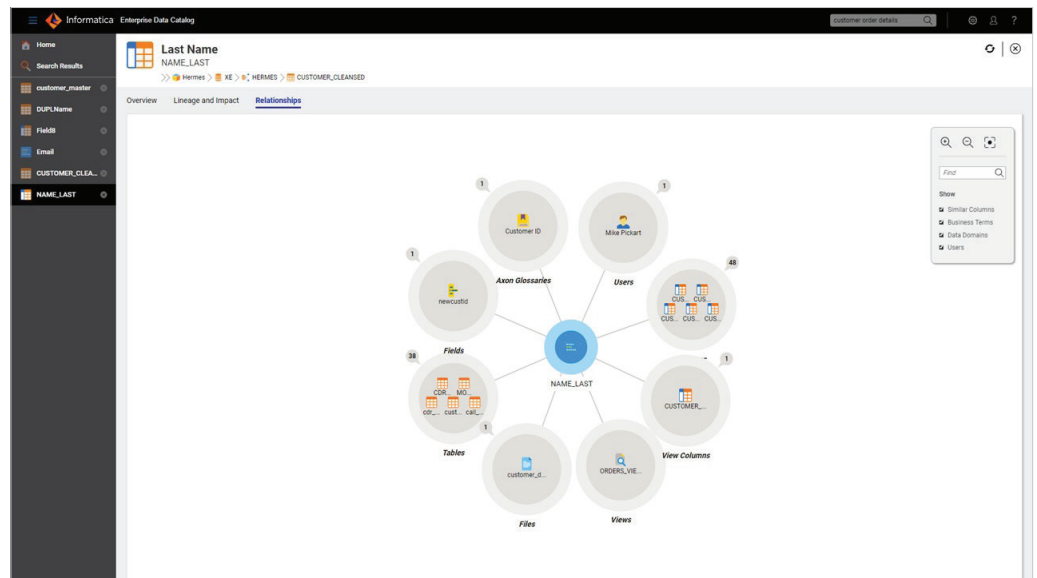
Figure 4: Understand your data with holistic data relationship views.

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, data lakes, cloud data stores, applications, BI tools, ETL tools, third-party metadata catalogs, 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 datasets 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-driven intelligence 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 with business terms to find their data and then browse holistic relationship views to find related data assets.

Discover and Understand Your Data Assets With Holistic 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 detailed lineage, profiling statistics, data quality scorecards, and holistic relationship views, providing a clear picture of your data.

Enrich Data Assets With Business Context Through Governed and Crowdsourced Annotations

Informatica Enterprise Data Catalog 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, Informatica Enterprise Data Catalog captures the context of who is using the data and for what purpose, along with crowdsourced tags, annotations, ratings, and reviews. This "wisdom of crowds" helps to enrich and curate data, making it even more valuable throughout the enterprise. Informatica Enterprise Data Catalog integrates with Informatica Axon for easy import of business glossary assets such as business terms, definitions, and policies from Axon. This business metadata is automatically 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.

Gain Insight Into Data Usage, Share Best Practices, and Estimate Asset Value

With Data Asset Analytics in Enterprise Data Catalog, you gain insights into data usage and users, with visibility into what data assets are in demand, who is using them, and more, enabling you to discover the most valuable data assets within your enterprise. Visual dashboards and exportable reports empower data leaders to share best practices, socialize data catalog adoption, and drive data-driven decision-making. By calculating data asset value—according to parameters you provide—the Automated Data Value Calculator helps you proactively manage and optimize your most important data assets.

Learn More

To learn more about Informatica Enterprise Data Catalog, please visit https://www.informatica.com/products/data-catalog.html.