

Informatica Data Lake Management

Benefits

- Find the right data for your analytics projects on a self-service basis
- Quickly prepare and share the data you need
- Easily operationalize data preparation for reusability

Collaboratively find, govern, and prepare data for analysis.

Data is undoubtedly the foundation of digital transformation. Organizations use new data processing platforms such as Apache Hadoop to derive previously unattainable—if not inconceivable—insights. The emergence of Apache Hadoop and the data lake approach now gives organizations the luxury of pooling all data so that it is accessible for users at any time for any type of analysis.

The sheer volume of data being ingested into Hadoop systems is overwhelming IT. Business analysts wait for quality data from Hadoop, while IT is burdened with manual, time-intensive processes to curate raw data into fit-for-purpose data assets. Without scalable, repeatable, and intelligent mechanisms for curating data, all the opportunity that data lakes promise risk stagnation. The key to solving the crisis of so-called data swamps is Informatica's metadata-driven artificial intelligence technology, known as the CLAIRE™ engine, that automatically discovers, profiles, and infers relationships about data assets.

Informatica Data Lake Management enables raw big data to be systematically discovered so that business analysts are empowered to turn data sets into trusted information on a self-service basis. Data scientists and business analysts can quickly find the data they're looking for using semantic and faceted search, while automatically understanding data lineage and data relationships. Data analyst teams can also easily collaborate with one another and share results in project workspaces. As they add data sets to their project workspace, machine-learning algorithms work in the background to recommend alternative data sets they might be interested in using.

Data sets within their project workspace can at any time be opened in Intelligent Data Lake's easy-to-use Excel-like data preparation tool. The metadata-driven approach to data preparation is the intelligent way to ensure quickly and repeatably turn big data into trusted information assets that deliver sustainable business value.

Key Features

Smart search and smart visualizations

Data scientists and analysts can quickly find data in the lake as well as in other enterprise systems using smart semantic search and inference-based results. Data assets can be filtered based on dynamic facets using system attributes and visualized using Apache Zeppelin-based smart charts.

Comprehensive, holistic data exploration

Get an overview of data assets, including custom attributes, profiling statistics for data quality, data domains for business content, and usage information. Crowd source information about data sets through tagging and enrich metadata to add business context. Quickly understand your data by previewing sample data based on user credentials. View its lineage to know where data has come from and where it is going, thereby building trust in the data. Know how the data asset is related to other assets in the enterprise based on associations with other tables or views, users, reports, and data domains. Progressively discover additional assets with lineage and relationship views.

Guided data preparation

Use an intuitive Excel-like interface to interactively prepare data for analysis with built-in transformations to filter, aggregate, merge, and combine data. Perform column-level data cleansing and data transformation using string, math, date, and logical operations. Guided intelligence helps prepare data sets such as recommendations of join keys when blending data sets. See sheet-level and column-level descriptive statistic

overviews, including value distributions and numeric and date distributions. All steps are recorded in recipes that can be used to automatically generate data flows that can be scheduled on a repeatable basis to operationalize analytical insights.

Operationalization of data preparation

Using thousands of pre-built business rules, business analysts can systematically improve data quality for products, while pushing data preparation steps to IT for repeatable execution.

Project collaboration

Organize work by adding data assets to projects workspaces. Collaborate with other analysts by adding team members to projects with different roles, such as co-owner, editor, or viewer, each with personalized privileges.

Data asset recommendations

Improve productivity and increase the reuse of trusted assets with automated recommendations based on machine-learning algorithms applied to the behavior and shared knowledge of other users. Alternate and additional assets are recommended for a project based on the data assets added to that project.

Wizard-based data uploads

Upload personal delimited files to the data lake using a wizard-based interface. Hive tables are automatically created for the uploads in the most optimal format. You can create, append to, or overwrite assets for uploaded data.

#	vehicletype	education	ethnicity	occupation	incomelev	dwellingty	HighEarn	oldcustid	firstname	lastname	email	year
1	Truck	Diploma	3	Firefighter	55	Apartment	0	208	Aaron	MORRIS	amorr	
2	2-door	Masters	1	Accountant	150	TwnHm	1	8219	Aaron	THOMAS	athom	
3	Truck	Diploma	3	Firefighter	55	Apartment	0	208	Aaron	MORRIS	amorr	
4	2-door	Masters	1	Accountant	150	TwnHm	1	8219	Aaron	THOMAS	athom	
5	Truck	Diploma	3	Firefighter	55	Apartment	0	208	Aaron	MORRIS	amorr	
6	2-door	Masters	1	Accountant	150	TwnHm	1	8219	Aaron	THOMAS	athom	
7	Truck	Diploma	3	Firefighter	55	Apartment	0	208	Aaron	MORRIS	amorr	
8	2-door	Masters	1	Accountant	150	TwnHm	1	8219	Aaron	THOMAS	athom	
9	Truck	Diploma	3	Firefighter	55	Apartment	0	208	Aaron	MORRIS	amorr	
10	Truck	Diploma	1	Engineer	80	Condo	0	9294	Aaron	HANSEN	ahans	
11	Convertible	Masters	1	Engineer	50	Single	0	365	Aaron	BRADLEY	abrady	

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

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

Find and access any data

Business analysts yearn for an efficient way to manage the ever-growing “volume, variety, and velocity” typically associated with big data. Trusted data assets are easily found using smart semantic search and dynamic facets to filter results. An automated machine-learning-based discovery process transforms related data assets into smart recommendations of new data assets that may be of interest to the analyst. This greatly increases confidence and reduces duplicate data sets being created for similar projects.

Efficiently collaborate with governance

To increase the efficiency of big data analytic projects, business analysts collaborate on data sets using project workspaces. As data sets are added to project workspaces, analysts can view profile statistics, end-to-end lineage of data sets, and a comprehensive, holistic view of all related data assets, data domains, users, and more. This aids in assessing quality of the data, sharing trusted data, and progressively discovering other data sets useful to the project. Role-based security ensures that only those analysts added to a project have access to the data.

Quickly prepare and share the data you need

As business cycles continue to shrink, speed is one of the few competitive advantages that data analysts can rely on in the race to add business value. Quickly prepare and share data instrumental in delivering competitive analytics. Informatica's self-service data preparation capabilities provide a familiar and easy-to-use Excel-like interface for business analysts, allowing them to quickly combine, filter, and blend data into the insights they need. Crowdsourced data asset tagging and sharing gives business analysts control over the data curation process enhancing operational efficiency.

Operationalize data preparation into reusable workflows

Regardless of automation and the self-service tools at their disposal, analysts often have to repeat data preparation activities on new sets of data. This simply squanders any gains from ongoing scale and re-usability. Informatica Data Lake Management records data preparation steps that can be scheduled as repeatable and automated data pipelines. This transforms data preparation from a manual process into a reusable, sustainable, and operationalized system.

