Informatica

Data Sheet

Informatica Enterprise Data Lake

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

Discover Raw Data and Prepare It for Advanced Analytics, Again and Again

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 give 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 the IT organization. Business analysts wait for quality data from Hadoop, while IT staff 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 risks 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 Enterprise Data Lake 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.

You can open data sets within the project workspace at any time in the easy-to-use Excel-like data preparation tool in Enterprise Data Lake. The metadata-driven approach to data preparation is the intelligent way to turn big data into trusted information assets that deliver sustainable business value.

Key Features

Intelligent Search and Intelligent Visualizations

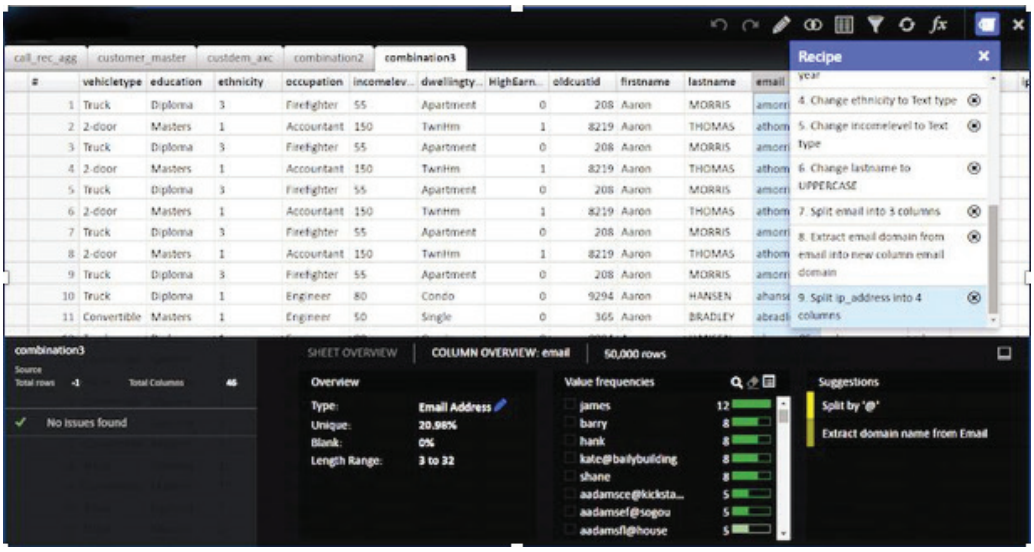
Find data in the data lake as well as in other enterprise systems using intelligent semantic search and inference-based results. You can filter data assets based on dynamic facets using system attributes and visualize with Apache Zeppelin–based charts.

Comprehensive Data Exploration

Get an overview of data assets, including custom attributes, profiling statistics for data quality, data domains for business content, and usage information. Crowdsource 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. Understand 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.

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 you can use to automatically generate data flows that you can schedule on a repeatable basis to operationalize analytical insights.



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

Operationalization of Data Preparation

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

Enterprise Collaboration

Manage data publications on a self-service basis while organizing work by adding data assets to project 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.

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. Easily find trusted data assets using intelligent semantic search and dynamic facets to filter results. An automated machine-learning-based discovery process transforms related data assets into intelligent 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.

Collaborate with Governance

To increase the efficiency of big data analytics projects, business analysts collaborate on data sets using project workspaces. As they add data sets to project workspaces, analysts can view profile statistics, end-to-end lineage of data sets, and 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. Business analysts can manage data publications on a self-service basis as well. 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.

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.

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 squanders any gains from ongoing scale and reusability. Informatica Enterprise Data Lake records data preparation steps that you can schedule as repeatable and automated data pipelines. This transforms data preparation from a manual process into a reusable and sustainable system.

For more information, visit the Informatica Enterprise Data Lake product page.

Informatica

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_03237