Informatica Data Prep

Simplify Self-service Data Preparation Across Cloud Ecosystems

Multi-cloud repositories are rapidly becoming cornerstones in any data modernization initiative. Snowflake, Google BigQuery, Amazon Redshift and Azure Synapse offer flexible options to ingest, integrate, persist and process massive volumes across a wide variety at high velocity across cloud environments.

These cloud data warehouses offer a myriad of benefits. But the sheer complexity and diverse types of data ingested and stored presents a challenge for the modern enterprise. How can they build and operationalize agile data engineering pipelines at enterprise scale?

Industry research estimates data scientists are spending nearly 80% of their time on cumbersome data preparation tasks. Before the business can use data to support a plethora of use cases (including analytics and AI-enabled workloads), the data must be fit for purpose. Data consumers have to wait for data scientistsand data engineers so they can find, access, blend, standardize and transform the data they need into usable, governed datasets.

One approach is using a standalone solution coupled with a partially automated data preparation process — but this method often results in bottlenecks. These bottlenecks lead to greater inefficiencies, greater operational costs and delays in time to insight. Without scalable, repeatable and intelligent mechanisms for discovering, cleansing and curating data, you risk the opportunity cloud data warehouses promise.

Key Benefits

- Rapidly discover, enrich, standardize and govern data pipelines for faster insights

- Prep data across multi-cloud and on premise at scale

- Reduce data preparation efforts with codeless integration

- Easily operationalize data preparation for reusability at enterprise scale

Powered by the Informatica **Intelligent Data Management Cloud** (IDMC), the Informatica® Data Prep service allows you to systematically discover, blend, standardize and transform large volumes of data. With Data Prep, you can empower data consumers to turn datasets into trusted and governed information for use and analysis at enterprise scale.

Key Features

Integrated Data Governance, Cataloging and Preparation

With Informatica Data Preparation, data consumers can rapidly discover the data they have across multi-cloud environments. A Google-like semantic search includes certified datasets along with key attributes about the data such as data domains, users and usage as well as other data assets. This allows users to easily visualize data sources, track datasets from source to destination and enable effective, data-driven business transformations with end-to-end data lineage and impact analysis capabilities.
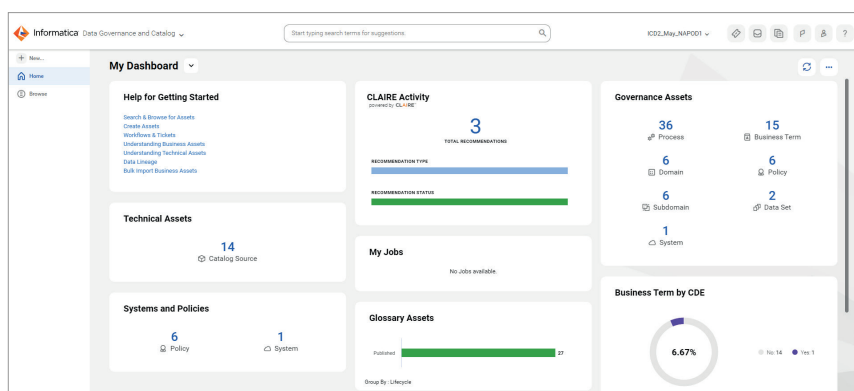

Figure 1. Rapid data discovery at enterprise scale across cloud ecosystems.

Simplified Data Compilation and Curation

With easy-to-use and intuitive data compilation capabilities in an Excel-like interface, users can harness the combined power of IDMC to easily build the recipes that will simplify and accelerate the data preparation process.
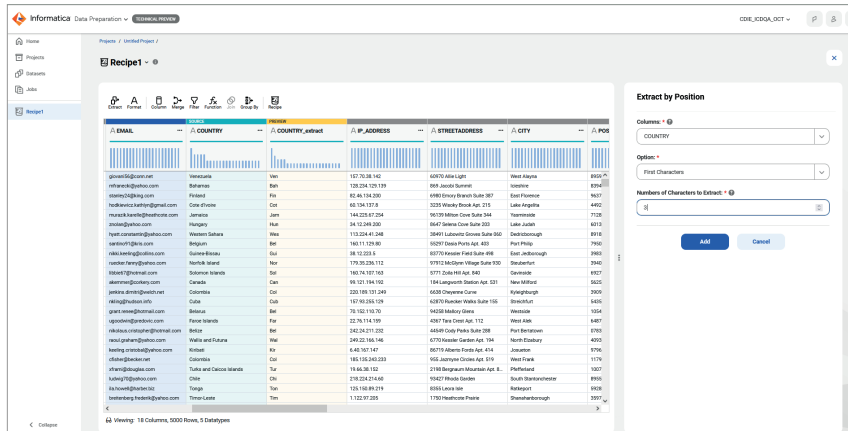

Figure 2. Easily build recipes to simplify and accelerate the data preparation process.

Interactive Data Profiling

Interactive profiling of the datasets that visualize sheet-level and column-level descriptive statistic overviews make it easy to facilitate recipe creation. You can add value distributions, numeric and data distributions. You can extrapolate data analysis through Column Profiling, which allows you to find out descriptive stats, interact with profiling histogram to quickly apply filters or remove outliers.
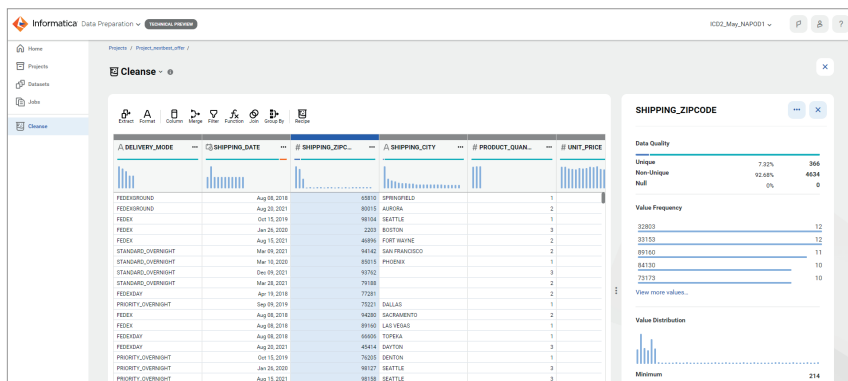

Figure 3. Easily identify data outliers and data orphans across vast data sets.

Data consumers can easily and iteratively prepare data for analysis by blending and transforming with prebuilt filter, aggregate, merge, lookup, shape and join functions. You can also easily combine data from multiple sources, allowing you to slice and dice data assets with code.
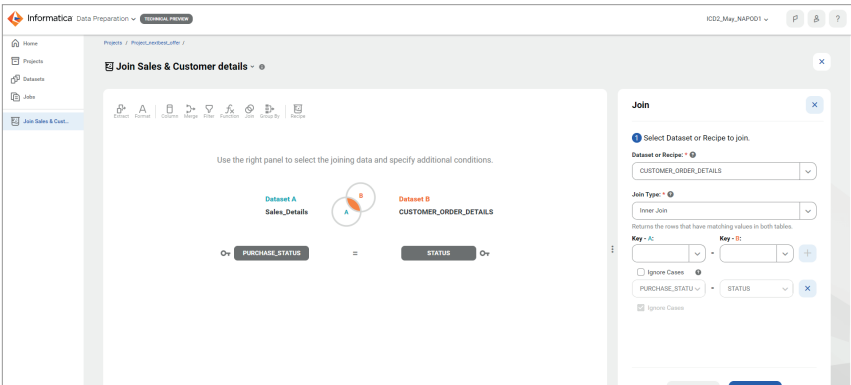

Figure 4. Self-service data preparation: data blending with Join.

Operationalized Data Preparation with Reusable Workflows

Data consumers often must repeat data preparation activities on new sets of data, which squanders any gains from ongoing scale and reusability. With Informatica's cloud-native Data Prep service, all steps are recorded in recipes enabling users to automatically generate data flows that can be scheduled on a repeatable basis to operationalize analytical insights. These recipes, along with the newly prepared datasets, are automatically pushed to the catalog where data consumers can search for it.
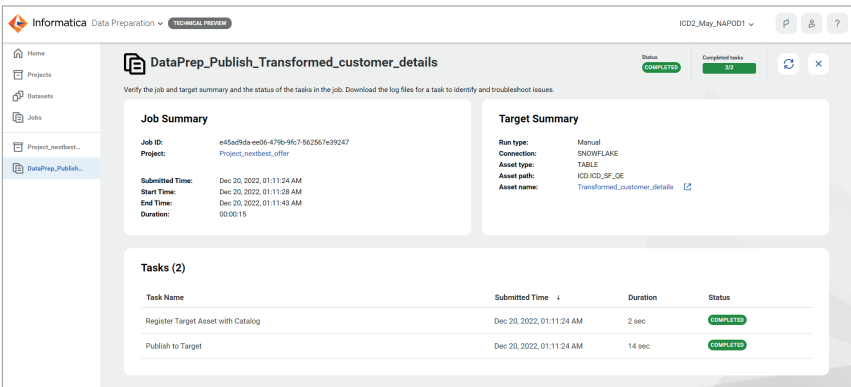

Figure 5. Publish prepared data assets to catalog.

Visualize Data Preparation for Transparency

The data flow diagram makes it easy for business users to understand how each recipe block is connected and to get insights into modular recipes for specific operations. The data flow clearly displays how data is cleaned, standardized and transformed from source to desalination. This also allows data consumers to reverse engineer and look for any issues in their data prep projects.
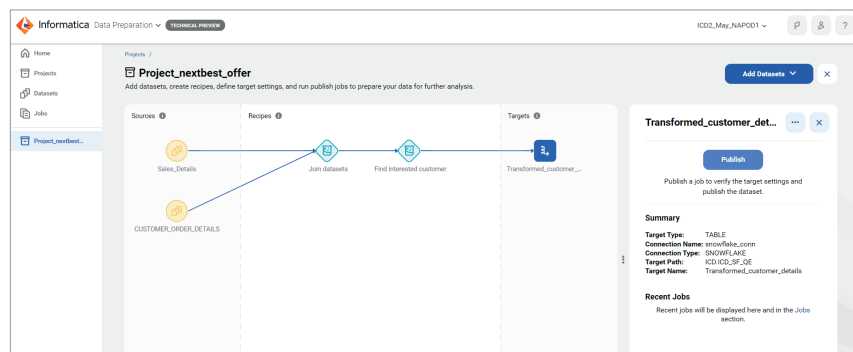

Figure 6. Visual representation of data flow.

Fully Governed User Privilege Control

Governance is critical to any data preparation initiative, especially in self-service environments. Informatica Data Prep provides comprehensive IT-governed user activity control for import, upload, publish, export or download activities across the cloud.

Key Benefits

Easily Discover and Prepare Your Data Across Clouds

Data cataloging is the foundational first step for any modern data preparation initiative. With petabytes of data residing across multi-cloud environments, data consumers can use the AI-powered Informatica Cloud Data Governance and Catalog to easily find the data they have with Google-like semantic search. The built-in data prep capability allows consumers to prepare their assets for various analytics, reporting and data science use cases.

informatica.com

Accelerate Time to Value for Data Consumers

The sheer complexity of the data that resides across cloud repositories requires simplified automation for modern data preparation initiatives at enterprise scale. Informatica Data Prep leverages recipe wizards to automate various tasks in the data preparation pipeline. And its intuitive, easy-to-use Excel-like user interface lets users iteratively transform raw data into curated, ready-to-use datasets for selfservice data integration and BI or analytics-driven use cases, along with building robust and accurate machine learning models.

Operationalize Data Preparation at Enterprise Scale

To truly derive value from their data, data consumers must be able to operationalize curated and governed datasets at enterprise scale. With Informatica Data Prep, all steps in the data preparation pipeline are recorded in recipes, allowing users to automatically generate data flows. Users can then schedule these data flows on a repeatable basis for machine learning models and analytical insights. Users can build, manage and deploy the lifecycle of the data preparation pipeline at scale across cloud ecosystems. End-to-end data preparation capabilities empower the business with comprehensive support for governance, performance and scalability.

Learn More

Preview access is now available for a limited number of customers and prospects. Contact dataprep.preview@informatica.com for more information.

At Informatica (NYSE: INFA), we believe data is the soul of business transformation. That's why we help you transform it from simply binary information to extraordinary innovation with our Informatica Intelligent Data Management Cloud™. Powered by AI, it's the only cloud dedicated to managing data of any type, pattern, complexity, or workload across any location—all on a single platform. Whether you're driving next-gen analytics, delivering perfectly timed customer experiences, or ensuring governance and privacy, you can always know your data is accurate, your insights are actionable, and your possibilities are limitless.