Accelerate Data Platform Modernization With AI-Powered Data Flow Analytics

Fuel Reuse and Efficiency With Intelligent, Automated Mapping Discovery

As enterprises move to the cloud, data platform modernization becomes critical. Data architectures based on traditional data warehouses must be rearchitected for the cloud. An essential part of modernization involves discovering and understanding your data landscape with the help of an enterprise-class data catalog, which provides intelligent and automated data asset discovery. But you also need to discover your data mappings—that is, the movement of data from source to target destination. Until now, there was no simple and efficient way to discover, understand, and manage your data mappings at enterprise scale.

The sheer number of mappings across systems poses a significant challenge in cloud migration. In most enterprises, ETL mappings have been built over years, even decades. Over time, thousands of mappings get created by developers in an enterprise, making it nearly impossible to make sense out of them. Developers often copy or modify mappings instead of reusing mapping components. It is also very common to have old, unused, and duplicate mappings that need to be removed. This creates a mapping "hairball" that can be a challenge to untangle. But reading through and categorizing mappings manually requires enormous effort. Lack of insights into mappings can slow down digital modernization initiatives, impact the ability to govern sensitive data flows, increase cost, and affect data quality.

Informatica® Enterprise Data Catalog with AI-powered Data Flow Analytics addresses the data mapping challenge by helping enterprises discover similar mappings, duplicates, and reusable mapping components. What was a near-impossible manual task can now be accomplished quickly and easily.
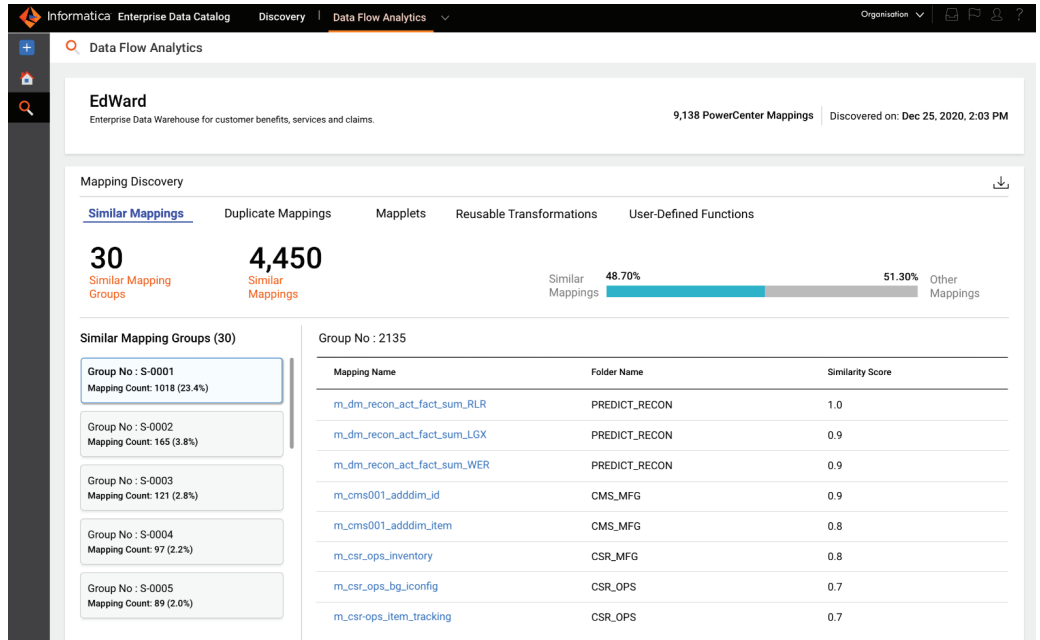
Figure 1. Informatica Enterprise Data Catalog Data Flow Analytics shows key metrics on mapping resources with detailed drill downs.

Key Features

Similar Mappings Discovery

Accurately estimating conversion effort for thousands of mappings is a considerable challenge and it is imprecise at best. Enterprise Data Catalog Data Flow Analytics discovers groups of similar mappings and identifies a representative mapping from each group. With a representative mapping identified, it's easier to more accurately estimate the effort needed for on-premises to cloud mapping conversion and optimization. Similar mappings are mappings with similar data integration logic, including identical source and target types, such as DB2 and Oracle, as well as transformation types.

The solution first partitions mappings into clusters and then selects representative mappings within the selected cluster. Each mapping in a similar group is then rated with a similarity score indicating how similar the mapping is to other mappings in the group. The mapping with the highest similarity score indicates the best representative mapping for the group. The Informatica CLAIRE® AI and ML engine automates the complex task of grouping mappings into clusters. For enterprises with thousands of mappings, this complex task is often simply not possible without automation.

Duplicate Mappings Discovery

Duplicate mappings are groups of mappings with identical source, target, transformations, and mapping links. These duplicates take time to identify and can add to the time and effort involved in data migration, hampering the effectiveness of data migration initiatives. The Duplicate Mapping Discovery feature identifies duplicate and redundant mappings that can then be removed by the administrator.

Mapplets and Reusable Transformations Discovery

Data Flow Analytics discovers repeated usages of mapping transformations and sequences of mapping transformations. This discovery enables you to refactor mappings with reusable transformations and mapplets to increase reuse and improve efficiency.

User-Defined Functions Discovery

The Data Flow Analytics solution can discover syntactically identical complex expressions inside expression transformation fields to assist in creating reusable user-defined functions. This minimizes effort and improves quality by enabling changes of complex logic in user-defined functions instead of making changes in multiple expression transformation fields. The CLAIRE AI and ML engine helps select complex expressions and parse them using ANTLR grammar, which enables the clustering of identical parse trees.

Key Benefits

Accelerate Cloud Migration and Modernization

Data mapping is essential to integration. But this critical task is often a complex and time-consuming bottleneck in cloud migration. Manual efforts to reconcile the difference in schemas between source and destination and consolidate information from across the organization can delay modernization initiatives. Data Flow Analytics provides three key benefits to accelerate cloud analytics modernization. It helps accurately estimate the effort needed for conversion with an in-depth Data Flow Analytics similarity mapping. It keeps technical debt in check with user-defined functions and mapplets and AI-powered recommendations. And it prevents post-migration mapping quality decay with a mapping quality dashboard.

Reduce Risk of Mapping Error and Improve Data Mapping Efficiency

Data mapping can be a tedious process made more so by the explosive growth in data volume. In addition, duplicate, redundant code can create risks in the form of data errors and inconsistencies. With Data Flow Analytics, enterprises can leverage the benefits of AI and ML to gain operational efficiency by identifying similar and duplicate mappings and reusable mapplets. The solution shows the number of groups and the percentage of mappings. It explains each discovery and even suggests how best to use it. For example, the Similar Mapping Discovery feature enables you to understand mapping complexities and estimate the effort involved in migrating similar mappings.

Reduce Operational Cost and Technical Debt

Any IT operation will inevitably have technical debt, which occurs when IT puts off updating or perfecting code in order to meet delivery deadlines. Over time, technical debt accumulates, and the time and cost involved in bringing code up-to-date balloons. Where mappings are managed manually, data professionals face accumulating technical debt and inefficiency. With automation provided by Data Flow Analytics, enterprises can move from unintelligible mapping complexity to an easy-to-understand graph that saves time and improves cost-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.

Next Steps

Learn more about Informatica Enterprise Data Catalog and intelligent, automated metadata management solutions for data cataloging.