Maximize the Value of Your Microsoft Investments With an AI-Driven Data Catalog

Key Benefits

- Empower technical and non-technical users with rapid data discovery and collaboration on data at scale

- Gain insights into your data inventory and usage across Microsoft and non-Microsoft data sources

- Easily visualize, trace, and understand your data from source to target with end-to-end data lineage

- Extract deep metadata and data lineage from your Microsoft and non-Microsoft data sources

- Enable robust enterprise-wide data governance, privacy, and regulatory compliance programs

Rapid Data Discovery and End-to-End Data Lineage Across Hybrid and Multi-Cloud

Enterprises of all types and sizes are on an accelerated path to digital transformation with a new sense of urgency. At the heart of virtually every digital transformation initiative is the need to derive value from data in order to drive innovation, improve customer experience, and increase overall operational agility and speed.

With data proliferating at an exponential rate, IDC predicts that by 2022 over 90% of enterprises worldwide will be running their businesses across on-premises, dedicated private cloud, and multiple public cloud environments.[1] This will further increase the complexity of the data landscape in which enterprises operate with petabyte-scale data spread across hundreds of disparate data sources.

One of the biggest hurdles for enterprises when operating in a complex data landscape is the lack of end-to-end visibility and understanding of their data. Most enterprises don't know what data they have, where it resides, who owns it, what all the data dependencies are, and whether or not the data is certified for use. This challenge is further amplified when the metadata that resides in various data sources is trapped and buried. It is often difficult to extract and even harder to understand.

Moreover, regulatory compliance, enterprise data governance, as well as a myriad of other data-driven digital transformation initiatives—including modernizing data warehouses and data lakes and migration to a multi-cloud environment—require you to have end-to-end visibility and an in-depth understanding of your data across your data landscape with comprehensive audit trails to mitigate all types of risks. This means you need the ability to trace data and its many transformations throughout the data lifecycle—from source to target—and across data pipelines at a granular level.

[1] https://www.idc.com/getdoc.jsp?containerId=prMETA46165020

For instance, it is estimated that nearly 95% of enterprises are leveraging Microsoft technologies —from software and middleware to hardware. For many, their data resides across Microsoft SQL Server, Azure SQL Data Warehouse, Azure Data Lake Storage, Azure Synapse Analytics, and Power BI, as well as in stored procedures for databases, to name a few. In addition, virtually every enterprise has a plethora of non-Microsoft data sources in their ecosystem. Across these disparate vendor systems—each with its own unique environment—extracting metadata can very quickly become a cumbersome undertaking as many of these data sources do not provide easily shareable descriptions of internal storages, processes, and relationships.

According to IDC, enterprises need two foundational pillars to drive successful digital transformation initiatives at scale. First, a fully integrated and agile technology architecture. And second, comprehensive and unified data management, data cataloging and governance capabilities. One that allows enterprises to obtain a deep understanding of all their data, including technical metadata, its definition, meaning, provenance, lineage and relationships.

Informatica® Enterprise Data Catalog enables you to build a comprehensive and complete repository of all your data regardless of where it resides—inclusive of Microsoft and non-Microsoft data sources. Powered by the metadata-driven intelligence in the Informatica CLAIRE® AI engine, Enterprise Data Catalog delivers advanced capabilities designed for rapid data discovery, curation and sharing of data, and metadata at scale. With end-to-end data lineage and impact analysis, you can visualize, trace, and understand your data across data sources from source to destination, enabling you to drive successful data-driven business transformations. Enterprise Data Catalog is a foundational pillar for enforcing a holistic data governance strategy across on-premises, hybrid, and multi-cloud environments.

Key Capabilities

Rapid Data Discovery Powered by Advanced Machine Learning

Enterprise Data Catalog enables rapid discovery of data with powerful Google-like semantic search empowering non-technical and IT users to easily find the data they need across Microsoft and non-Microsoft data sources. Users can quickly discover and profile data, identify its location, and obtain other key attributes about the datasets at scale. Semantic search is also applied to inferred data domains including synonyms and concept matching, so that no data asset is left undiscovered across on-premises and multi-cloud environments.

Using advanced statistical and metadata-driven machine learning algorithms, Enterprise Data Catalog tackles computational complexity inherent in data when discovering, tagging, clustering, and identifying similarities and patterns in data, enabling you to intelligently catalog all types of data at scale.

Enterprise Data Catalog also allows for easy import of business glossary assets such as terms, policies, and classifications from Informatica AxonTM Data Governance as well as third-party tools. You can add rich business context to the data by automatically associating business terms with the right technical metadata.

Broad and Deep Metadata Connectivity With End-to-End Data Lineage and Impact Analysis
Enterprise Data Catalog is the catalog of catalogs with both broad and deep metadata connectivity. It offers the most comprehensive set of scanners that are purpose-built to extract deep metadata and data lineage from a myriad of widely adopted data sources across on-premises, hybrid, and multi-cloud environments.

End-to-end data lineage and impact analysis capabilities allow you to easily visualize, trace, and understand the flow of data from enterprise applications, data warehouses, data lakes, and databases to ETL and BI tools. You can perform detailed impact analysis of transformations within your Microsoft data sources as well as on third-party upstream and downstream data assets and linked systems. Moreover, you can easily analyze the data dependencies from source to target, understand the impact of proposed changes, and perform root-cause analysis of data issues.

You can 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. Additionally, a drill-down lineage view expands any lineage path to show granular column- and metric-level lineage.

Advanced Scanners for Microsoft SQL Server, SSIS, SSAS, SSRS, and More
Dynamic SQL generation, parametrized procedures, ETL code controlling data transformations, file copying scripts, and many others are typical examples where lineage usually cannot be extracted or is not detailed enough to support various data-driven digital transformation initiatives.

Enterprise Data Catalog Advanced Scanners are purpose-built for extracting deep metadata and derive detailed data lineage and data relationships at scale from SQL dialects and stored procedures for Microsoft SQL Server, SSIS, SSAS, and SSRS. With the Advanced Scanners, you can scan both static and dynamic code as well as perform language parsing to obtain automated data lineage. Extracted data lineage provides full visibility into the procedure calls with parameter tracking, dynamic SQL generation from values based on parameters, database queries, and more. Supported objects include views, procedures, functions, triggers, macros, external tables, and so on. Analysis can be performed online on a live database or offline using metadata-only extracts. For instance, the Advanced Scanner for SSIS allows you to easily visualize every component in an ETL job and gain quick access to every expression being applied to the data from source to target.
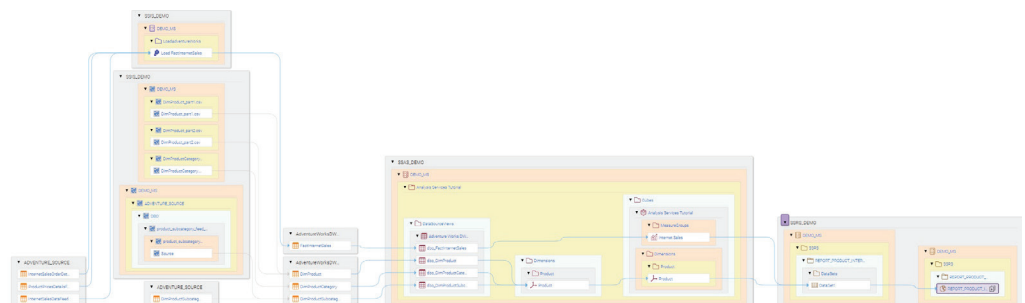

Figure 1: Eliminate the black box effect and obtain comprehensive data lineage.

Advanced Scanners allow you to 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. This enables you to understand every transformation that has occurred to the data at a granular level.

Advanced Custom Metadata Loader for Microsoft Excel

The Enterprise Data Catalog Advanced Custom Metadata Loader is specifically designed to enable you to easily surface and extract metadata, custom code and custom content from Microsoft Excel spreadsheets. For instance, you can use the loader to easily surface custom code and custom content from your Excel spreadsheets that you want to extract data lineage from and load it into the Enterprise Data Catalog. Once your data catalog is populated with this extracted data lineage, you can easily combine it with the rest of the data lineage information that's already in your catalog. This allows users to leave no metadata behind and meet stringent data governance and regulatory compliance reporting stipulations.
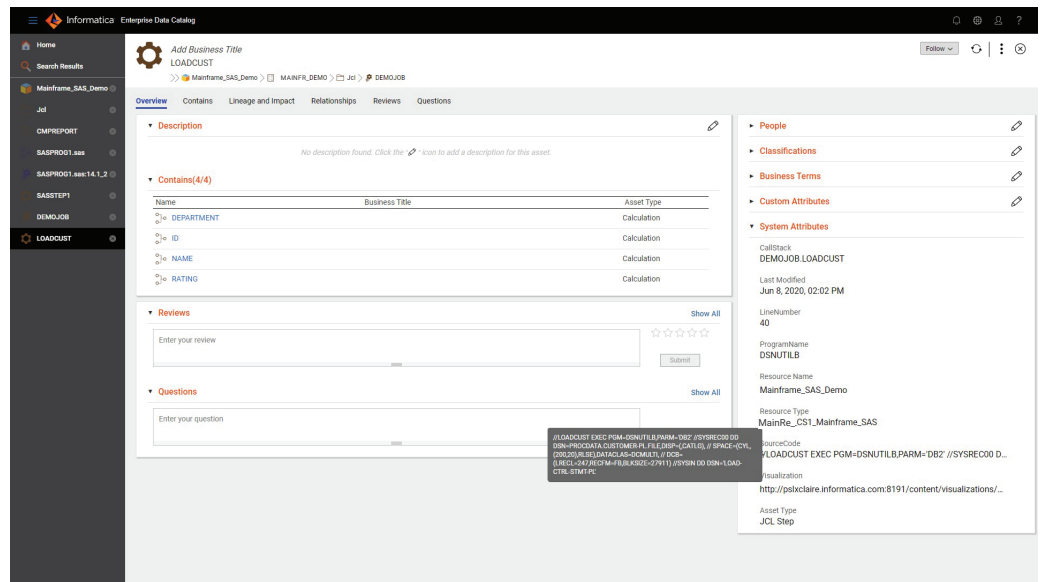

Figure 2: Enterprise Data Catalog Advanced Custom Metadata Loader enables users to extract custom metadata at scale.

Data Collaboration and Social Curation With Intelligent Crowdsourcing and Annotations

Enterprise Data Catalog empowers non-technical and IT users to easily find the most relevant and trusted data for analysis by harnessing the combined power of sophisticated machine learning algorithms, human expertise, and collaboration. Data owners and subject matter experts can certify datasets and provide ratings and reviews, enabling social curation of data. A Q&A platform enables subject matter experts to answer common questions from users.

Dataset Provisioning at Scale

Finding all your enterprise data is the first step; seamless provisioning enables you to maximize its usage and value through self-service. Enterprise Data Catalog integrates seamlessly with Informatica Cloud Data Integration to deliver the data a user has found to a place where they can consume it (assuming they have sufficient privileges) through simple click-through provisioning for self-service analytics. The integration supports a number of Microsoft data sources including Microsoft Azure SQL Data Warehouse, Azure Synapse Analytics, and Azure Data Lake Storage as well as a broad array of third-party data sources and targets.

Rich Data Quality Statistics

Users can view data profiling statistics alongside technical metadata to understand the quality of data assets before using data for analysis. Profiling statistics include value distributions, patterns, data types, and data domain inferences.

Advanced Data Asset Analytics

With Data Asset Analytics in Enterprise Data Catalog, you can obtain deep insights on the usage of data within your organization across your Microsoft and non-Microsoft data sources, enabling you to proactively manage and optimize the value of your data assets. For instance, you can obtain information into what percentage of your data inventory resides in Microsoft data sources as well as the types of data your users are accessing from these data sources. This will enable you to better prioritize your Microsoft data warehouse modernization and multi-cloud migration strategy.

Key Benefits

Gain End-to-End Visibility Into Your Data Across Microsoft and Non-Microsoft Data Sources

Comprehensive data lineage and impact analysis capabilities coupled with deep metadata extraction enable you to take a strategic approach to driving data warehouse modernization, multi-cloud migration, and enterprise data governance and regulatory compliance initiatives. Enterprise Data Catalog lets you trace even the most complex data lineage from source to destination across Microsoft and third-party enterprise applications, databases, data warehouses, ETL, and BI tools regardless of where your data resides—from on-premises, Microsoft Azure to multi-cloud environments. With Enterprise Data Catalog, you can gain an in-depth understanding of your data and all its transformations throughout the data lifecycle and across the data pipeline in order to drive successful data-driven business transformations.

Speed Enterprise-Wide Data Governance and Regulatory Compliance Reporting

Enterprise Data Catalog seamlessly integrates with Axon Data Governance as well as third-party tools that are in your environment, enabling you to streamline the process of configuring, deploying and maintaining a holistic data governance program.

Complex audit and reporting tasks are simplified with Enterprise Data Catalog enabling you to create an end-to-end data flow at the summary level for faster, easier data governance and compliance reporting. You can trace critical data elements and track transformations in detail for any necessary documentation.

Next Steps

To learn more, please visit the web pages for [Informatica Enterprise Data Catalog](#) and [Enterprise Data Catalog Advanced Scanners](#).