

Automated Data Lineage with IBM Watson Knowledge Catalog

Gain deeper visibility into your data
and its journey from source to end-use



Highlights

Enable regulatory compliance, conduct impact analysis and build trust in your data

Deliver deeper data lineage and faster time to value

Simplify data lineage with automated scanning of data flows in third-party tools

Data proliferation has made it harder than ever to trust your data. It's important to know where your data comes from, how it has changed and who is using it. A lack of visibility into the data journey from source to end-user can have far-reaching consequences. If you don't know what sensitive data your organization has or who can access it, you run the risk of failing to meet data privacy regulations. This can lead to great financial and reputational damage. When analysts can't fully explain calculations, the origins of the underlying data, and the quality or privacy attributes of the data, it can lead to a decreased level of trust in their reports. Furthermore, data engineers can spend a disproportionate amount of time on impact analysis due to a lack of understanding of the environment. This slows down the delivery of trusted data for the generation of insights.

To overcome these challenges, organizations need to simplify their understanding of a data set's journey from its origin to its end-use. They need specific details on how data was transformed and by whom along the way. Data lineage is a visual representation of a data set's journey from its origin to end use. It has evolved into the main enterprise tool to understand the flow of data and the contribution of each person and program along data's lifecycle. To help organizations overcome these challenges, IBM has partnered with MANTA to boost native data lineage capabilities available within [IBM Watson® Knowledge Catalog](#) on IBM Cloud Pak® for Data.





Enable regulatory compliance, conduct impact analysis and build trust in your data

Data lineage is essential for modern data management and has a wide range of use cases. Data lineage is a required aspect of regulatory compliance and helps identify the origins of sensitive data, the various locations where it's stored, who can access it, and what data should be anonymized.

Enterprises are constantly implementing changes to their data architecture and pipelines. Without data lineage, it can be difficult or impossible to assess the impact of planned changes. Data lineage gives teams insight into the downstream impacts of these changes before potentially costly bugs are introduced. Data lineage can also enable analysts and data consumers to conduct root cause analysis by diagnosing issues and discrepancies in data and reports.

Organizations can also use data lineage to speed up migration processes while undergoing digital transformation. Data lineage gives engineers visibility into which architectural components must be migrated at once and which need not be migrated at all. Only when analysts and data scientists have a complete understanding of data can they rely on it for confident decision-making. Data lineage is a critical capability of modern data governance to deliver trust in the data used for analytics and AI by providing visibility into a data's provenance and end-to-end journey.

Deliver deeper data lineage and faster time to value

Together with business-friendly native data lineage delivered by Watson Knowledge Catalog, MANTA Automated Data Lineage for IBM Cloud Pak for Data provides the deep technical lineage that data engineers need, historical versioning to view lineage changes over time, and indirect lineage to record even the most specific data operations within MANTA's Lineage Flow UI.

Collectively, this means that the addition of MANTA will provide quicker time to value not only through the automation of previously manual processes, but also through the ability to more rapidly answer questions about whether certain data is trustworthy.

IBM's partnership with MANTA enables organizations to create complete end-to-end data lineage for full understanding, observability and control of their data.

Simplify data lineage with automated scanners

The partnership between IBM and MANTA helps organizations ease the amount of manual effort necessary for robust data lineage by providing scanners for the automated discovery of data flows in third-party tools such as Power BI, Tableau, and Snowflake. This information is then automatically scanned into Watson Knowledge Catalog's Data Lineage UI and becomes available to view alongside the data quality, business terms and other metadata previously available to Watson Knowledge Catalog users.

Automated data lineage helps avoid the manual creation of data lineage, which can be tedious, time intensive, can contain contradictory or missing information, and leads to teams relying on unsound lineage to make critical decisions.

Conclusion

The partnership with MANTA allows IBM to provide quick time to value not only through the automation of manual processes, but also through the ability to answer questions more rapidly to build trust in data.

MANTA Automated Data Lineage for IBM Cloud Pak for Data is available as an add-on to Watson Knowledge Catalog, further improving the ability of [IBM's data fabric solution](#) to satisfy [governance and privacy use cases](#). Surrounded by existing capabilities like consistent cataloging, automated metadata generation, automated governance, reporting, and auditing assistance, MANTA Automated Data Lineage for IBM Cloud Pak for Data helps to bolster IBM's data governance capability.

Why IBM?

IBM has been named a Leader in multiple analyst assessments including the [Gartner 2022 Magic Quadrant for Data Quality solutions](#), the [Forrester Wave: Enterprise Data Fabric Q2 2022](#) and the [Gartner 2022 Magic Quadrant for Data Integration Tools](#).

IBM offers integrated data governance, data integration, automated data lineage, data quality, data privacy and entity resolution as part of [IBM Cloud Pak for Data](#), which is an extensible data and AI platform that can be deployed on almost any cloud. The integration of these capabilities within a unified environment helps streamline data governance tasks to accelerate understanding of what data means, where it comes from and how it relates to other assets.

For more information

To learn more about data lineage solutions from IBM, please contact your IBM representative or IBM Business Partner.

Access the [free trial](#) today to try our data quality and governance capabilities.

To explore how IBM enables the creation of a governed, compliance-ready data foundation visit the [IBM data governance page](#).

© Copyright IBM Corporation 2022

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
December 2022

IBM, the IBM logo, IBM Cloud Pak, and IBM Watson are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on [ibm.com/trademark](https://www.ibm.com/trademark).

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

