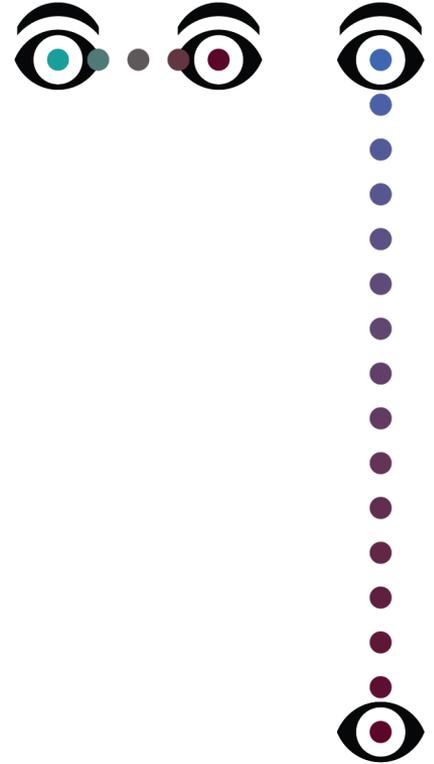


Exploring Data Governance

Pavel Maltsev
Technical Sales Specialist
pmaltsev@cz.ibm.com



A design concept and architecture for data management
that addresses the challenges of data complexity

Data Fabric

The background features a complex network of glowing blue nodes and lines, creating a fabric-like structure that represents data connectivity. The nodes are small white dots, and the lines are thin, glowing blue lines that connect the nodes in a mesh pattern. The overall effect is a sense of dynamic, interconnected data.

Making your data ready for high value analytics and AI is widely time consuming today

Prepare

Data operations
Discover, understand, ingest, integrate, assess quality, cleanse

The diagram shows a vertical flow starting with a circle, followed by a downward arrow, and then a row of three shapes: a circle, a square, and a triangle. Below this is another row of four shapes: a circle, a square, a circle, and a circle.

Build

Run
Productionalize

Manage
Expandability

The diagram consists of three adjacent dark teal rectangular boxes. The first box contains the word 'Build'. The second box contains the word 'Run' with the word 'Productionalize' written in italics below it. The third box contains the word 'Manage' with the word 'Expandability' written in italics below it.

← Where analytics and data science teams focus →

← Where 80% of data science time is spent →

← Where business impact is created →

← Lineage (the ability to retrace steps used to build data products):
Supports *auditability / compliance* and *customization / reuse* →

Business benefits of a data fabric

Simplifies data complexity through automating data integration, data governance, and data consumption

Intelligent integration

Unified governance

Knowledge insights

Automates data engineering tasks and augments data integration across hybrid cloud resources

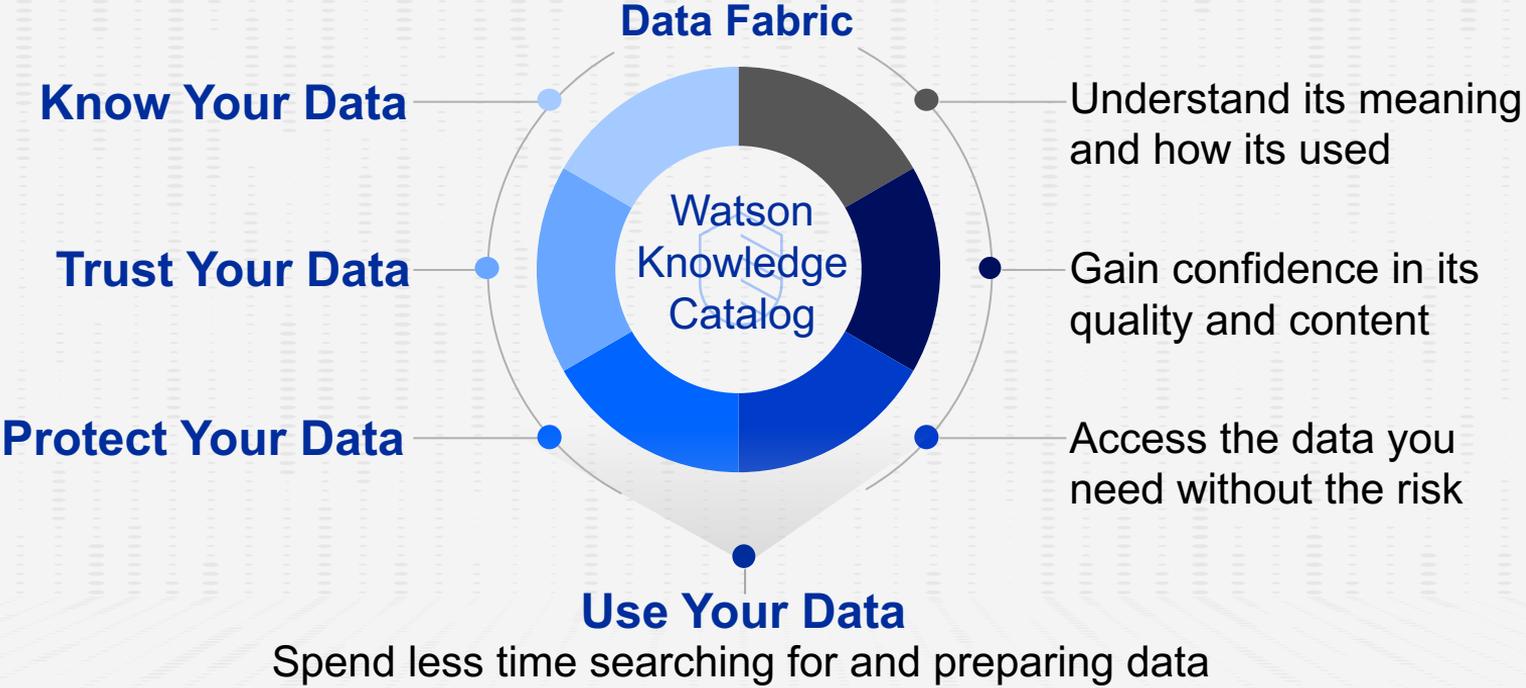
Automates data governance, protection and security tasks; enabled by active metadata

Enables self-service for easy end user data consumption and collaboration

Data governance and privacy



Data fabric for data governance and privacy



Create a trusted and protected business-ready analytics foundation

Know your data

Data governance capabilities of IBM's data fabric



Establish a business-ready data governance foundation



Define standards and policies and set boundaries and rules



Comply with government and industry data privacy regulations



Know your data

Data lineage is a critical capability of data governance and privacy



Comply with industry and government regulations like GDPR and CCPA, etc.



Capture business processes across the organization



Analyze the impact of changes to data tools and diagnosis issues and discrepancies in data and reports



Know your data

Data lineage capabilities of data governance and privacy



Connectivity to and automatic data lineage capture from all data products being used; data modeling, data integration, reports etc.



Visualization of how data moves from one source to another through a **business view**



Ability to dive into the details of each transformation at different points in time for a **technical view**



Trust your data

Data quality capabilities of IBM's data fabric



Automated data discovery for a vast array of hybrid cloud data sources



Automatic data content analysis and association to business metadata



Automated quality analysis with rule definition and enforcement



Protect your data

Data privacy capabilities of IBM's data fabric



Automate and enforce industry and government data privacy regulations



Automatic detection and autonomous protection of sensitive information



Advanced data privacy protection methods



Use your data

Data consumption capabilities of IBM's data fabric



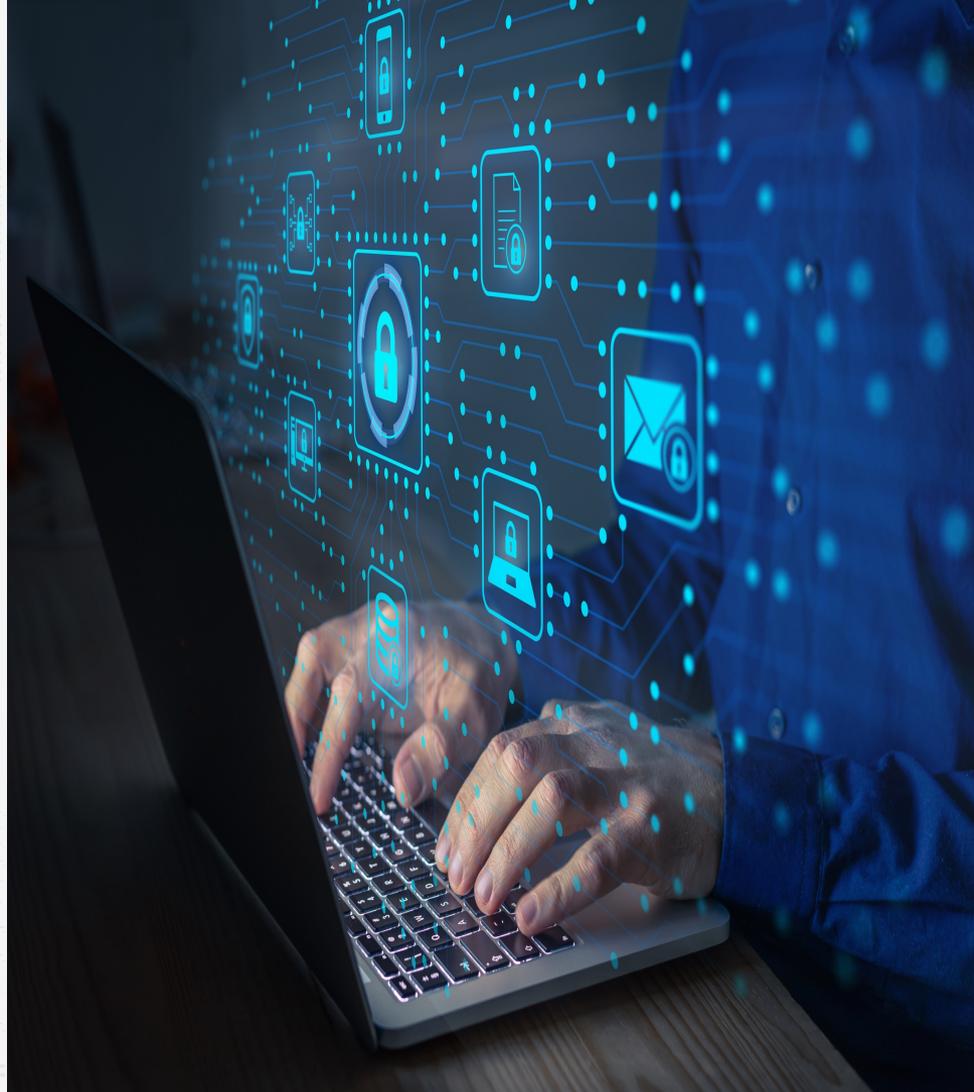
Augmented catalog with curated and governed information assets



Self-service access to trusted data to the right people at the right time



Easy, no-code, data preparation for analytics and AI use cases



Automated data governance and privacy



AI-powered data discovery



Augmented metadata enrichment



AI-powered asset recommendations



Assessment and scoring of data quality



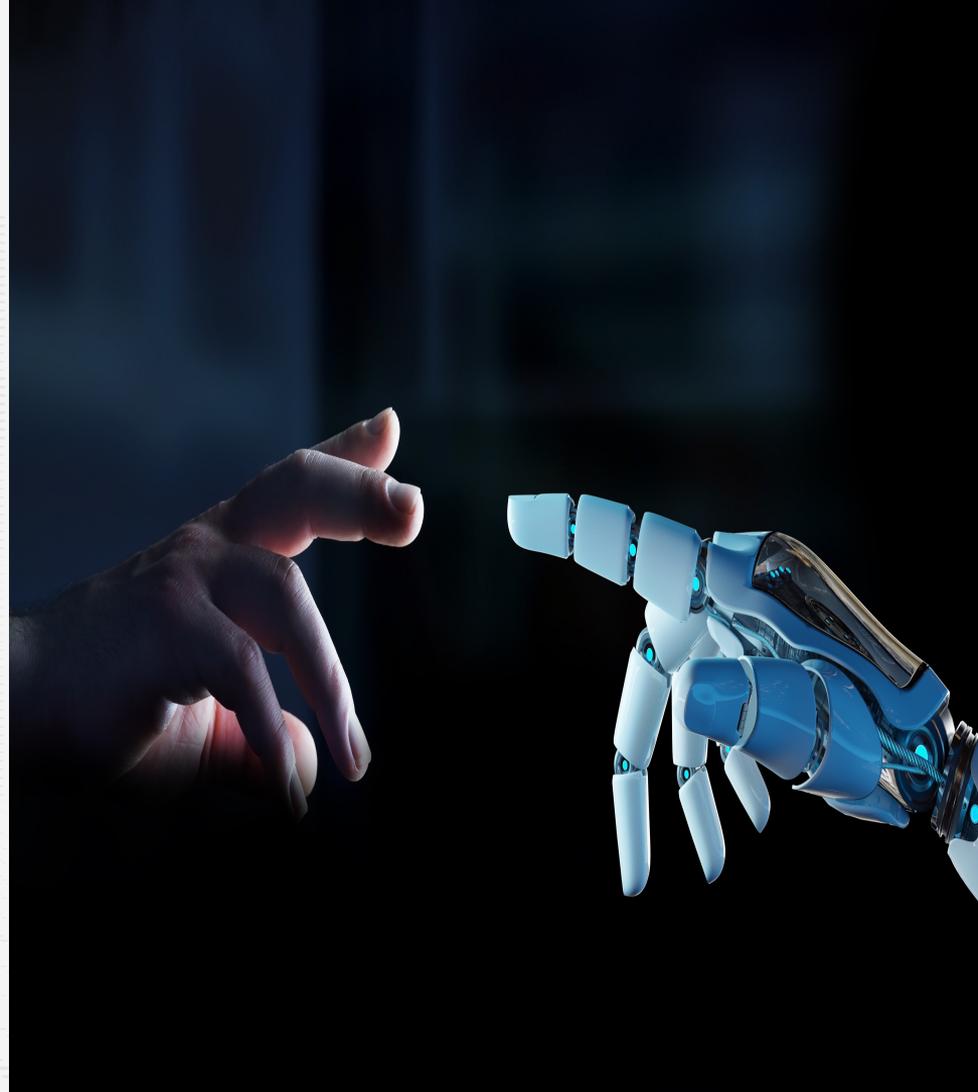
Assignment of business terms to assets



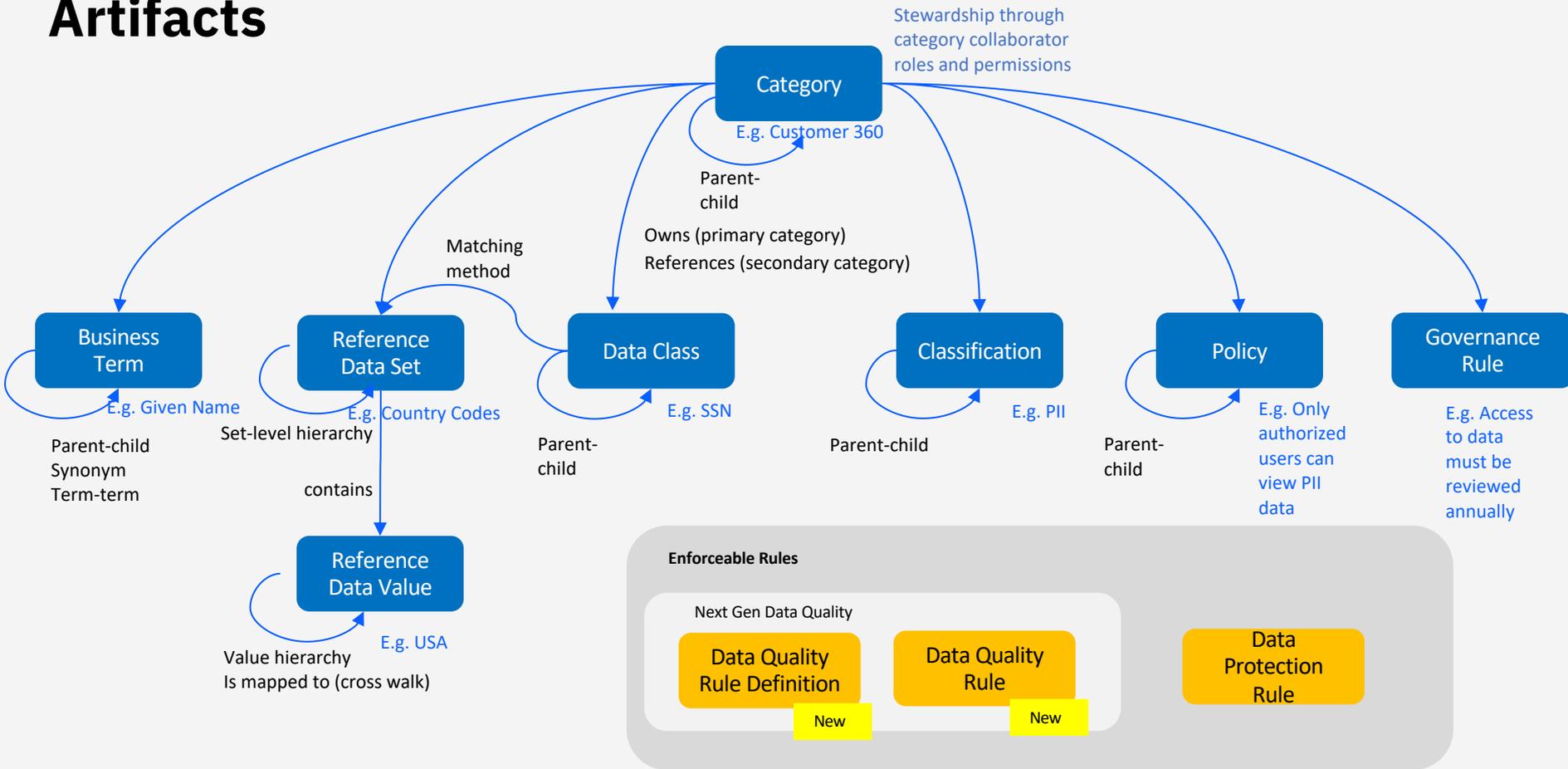
Auto enforcement of data protection rules



AI governance of machine learning models

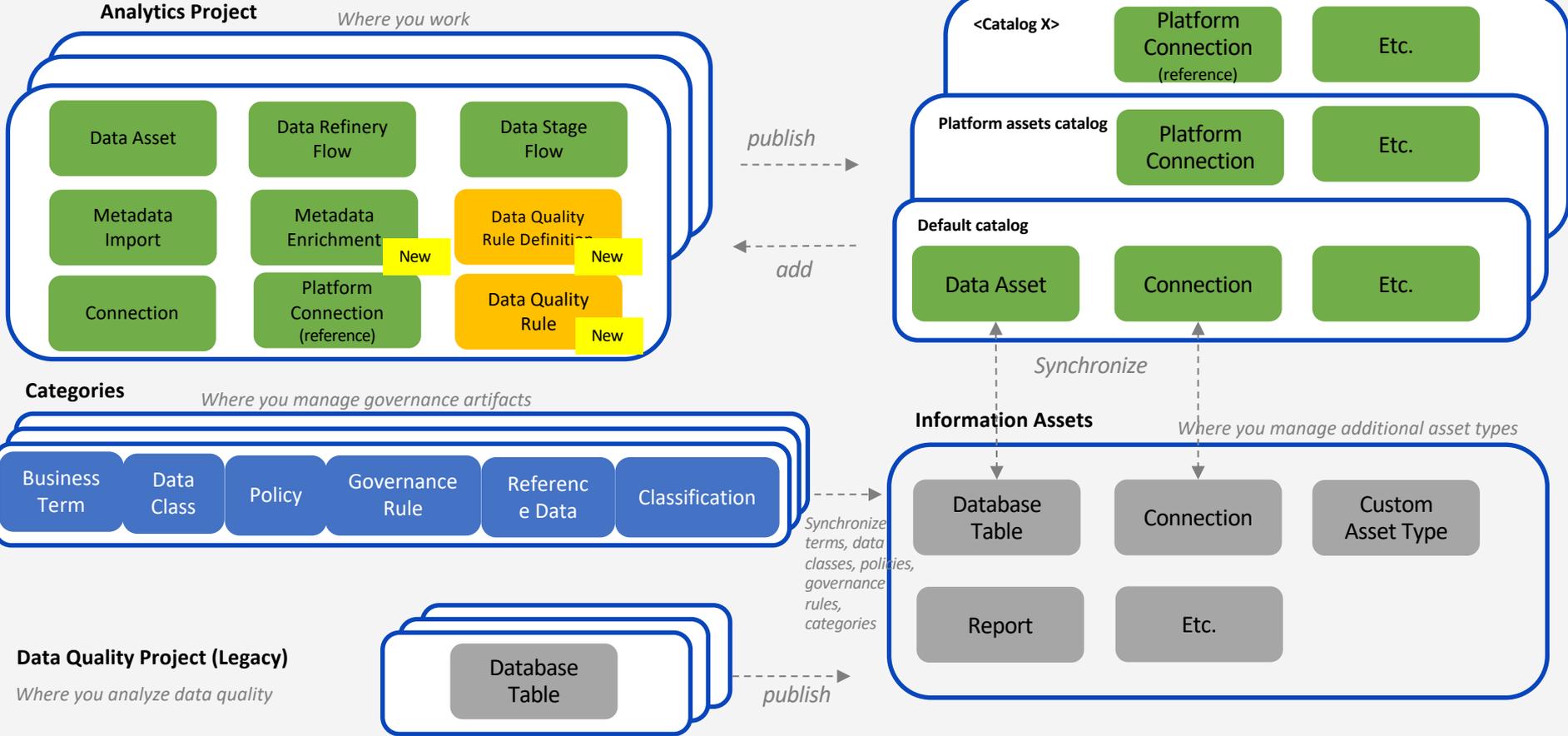


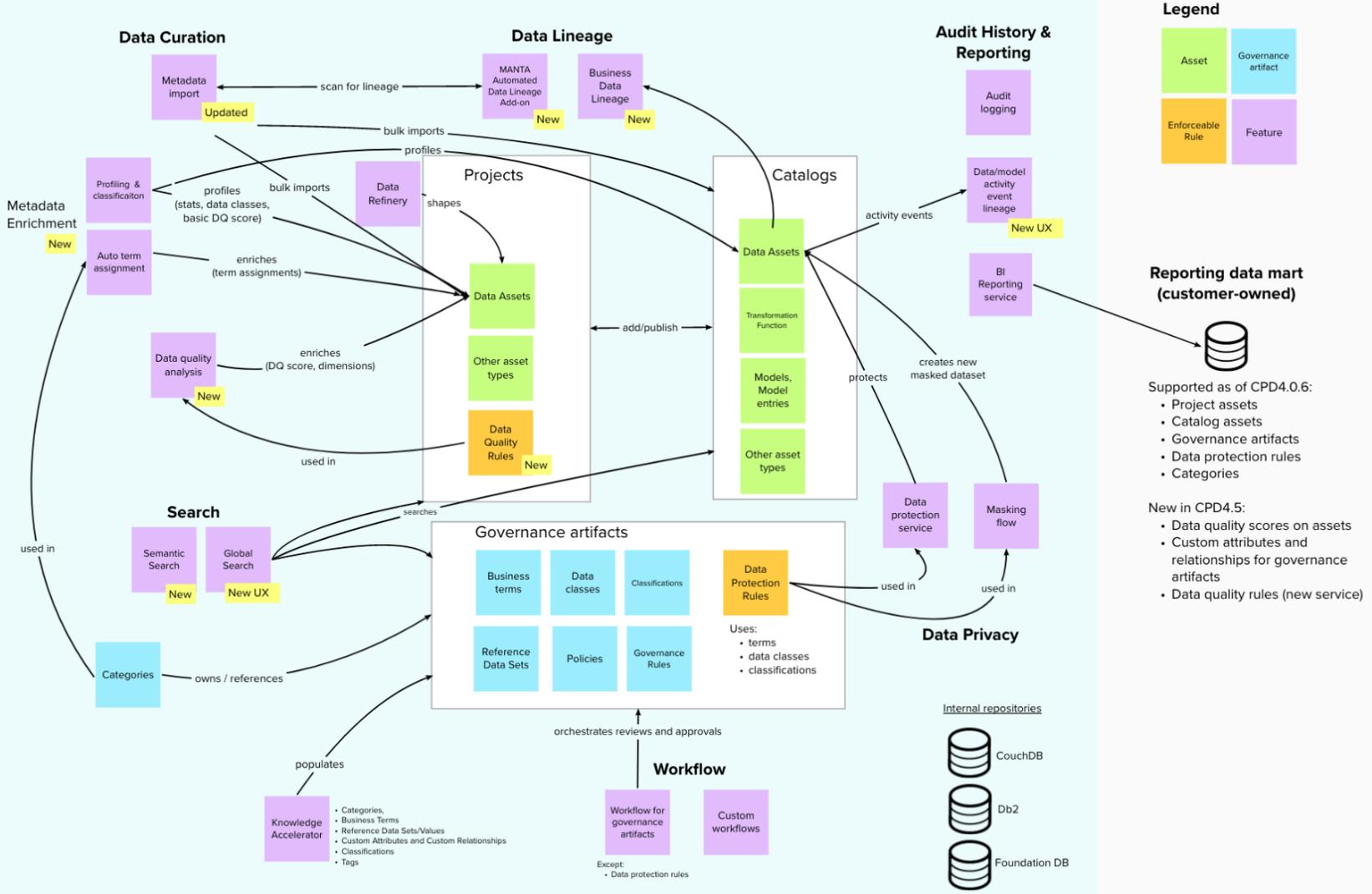
Data Governance with Watson Knowledge Catalog Artifacts



Data Governance with Watson Knowledge Catalog

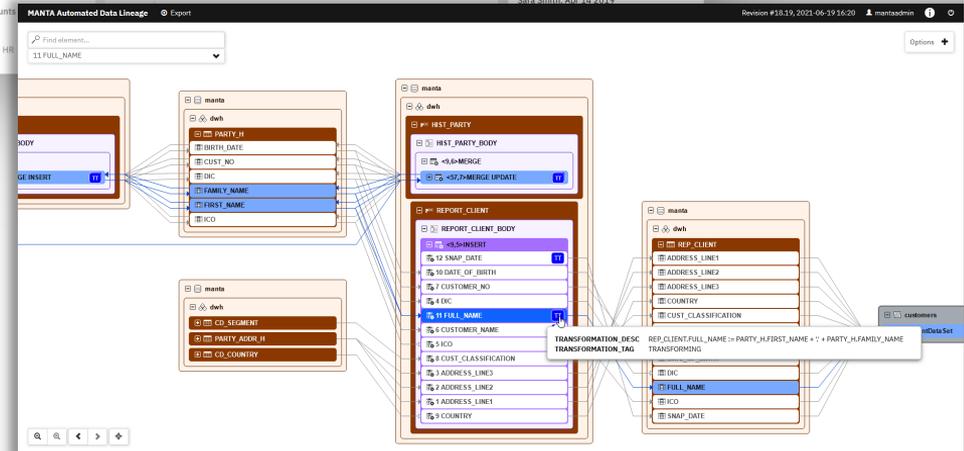
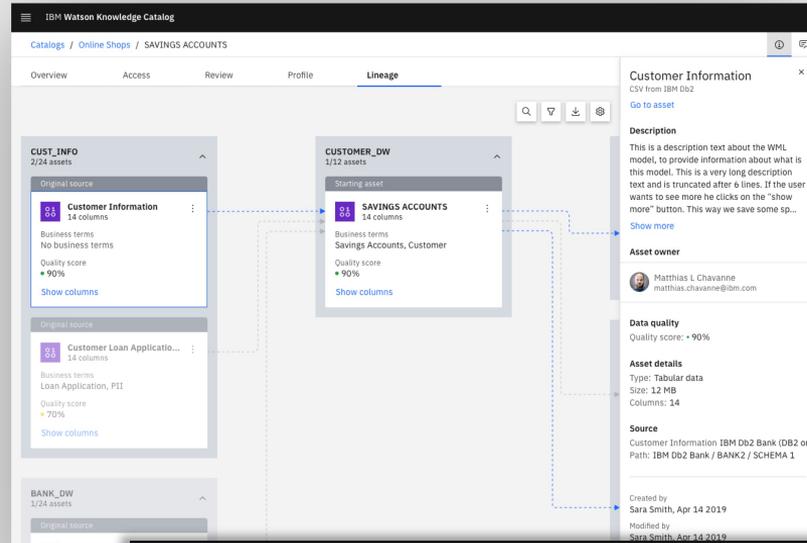
Artifacts





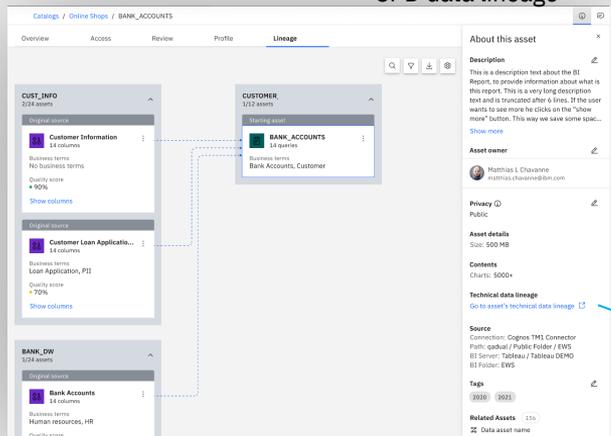
Data Lineage

- ✓ Provides business-user-friendly summary view
- ✓ Drill down to view technical details
- ✓ Automated discovery of data flows from 3rd party tools
- ✓ Code analysis of stored procedures & ETL jobs capture transformations
- ✓ Manual lineage through APIs
- ✓ Technical, Indirect, and Historical data lineage GUI
- ✓ Rules to filter for things needed in data lineage



MANTA Automated Data Lineage automates the discovery and analysis of lineage flows

CPD data lineage



1

MANTA Automated Data Lineage Add-On

- Automated discovery of data flows from 3rd party tools
- Code analysis of stored procedures and ETL jobs to capture transformations
- Advanced manual lineage
- Technical data lineage GUI
- Indirect data lineage
- Historical data lineage
- Rules to filter for things needed in data lineage

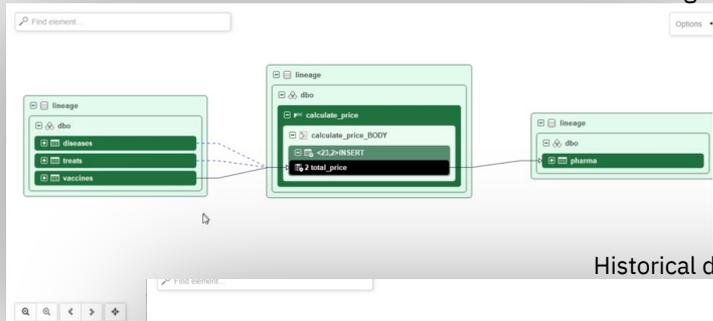
2

Scanned assets are stored in WKC catalogs and visible through native UI

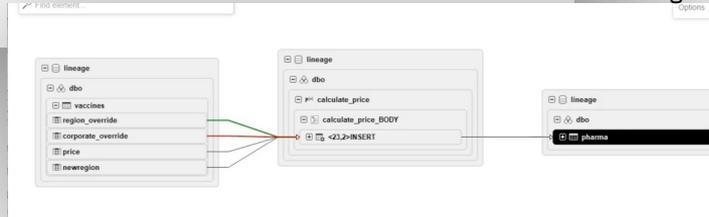
3

Option to view additional lineage details

Technical data lineage



Historical data lineage



CPD Data Lineage

- View lineage graph via WKC UI
- REST APIs to capture lineage info
- Import/export lineage through CSV
- Automatically captures data flows from the CP4D platform
 - Models from Auto AI
 - Data Virtualization
 - Next Gen DataStage Jobs
 - Data Refinery
 - Advanced Data Privacy

MANTA Automated Data Lineage – Supported Scanners

Databases

Yellowbrick
Amazon RDS for PostgreSQL
Amazon Redshift
Google Big Query
Snowflake
Greenplum
Hive
IBM Db2
Microsoft Azure SQL Data Warehouse
Microsoft Azure SQL Managed Instance
Microsoft SQL Server Database
Netezza
PostgreSQL
SAP ASE (Sybase)
SAP Hana
Teradata Database
Oracle Warehouse

Modeling

ER/Studio
Erwin Data Modeler
SAP Power Designer

Data Integration

Apache Kafka
Infosphere DataStage
Informatica Power Center
Microsoft SQL Server Integration Services
Oracle Data Integrator
Pig
Sqoop
StreamSets
Talend
Azure Data Factory (MANTA Roadmap 2022)
Next Gen Data Stage (MANTA Roadmap 2022)

Reporting & Analysis

Qlik Sense
Cognos
Microsoft Excel
Microsoft SQL Analytics Platform System
Microsoft SQL Server Analysis Services
Microsoft SQL Server Reporting Services
Oracle Business Intelligence (OBIEE)
Power BI
SAP Business Objects
SAS
Tableau

Programming languages

COBOL
C#
Java
Python (experimental)

Note:

Scanners not on this list can be supported through either MANTA's Advanced Manual Lineage or through IBM Business Data Lineage's native Manual Lineage Capabilities

Data Lineage User Flow (Demo)

1. Initiate lineage scans through platform connections and metadata import
2. Select data sources to scan (or provide extract)
3. Discovered asset and lineage metadata are added to selected catalog and knowledge graph
4. Visualize data lineage from catalog asset using CPD data lineage or view data lineage in Manta

Create MANTA connections (scanners) through existing Connections UI

The screenshot shows the 'Create a connection' interface. On the left, there are filters for 'Provider' (IBM, User-defined, Third party), 'Compatibility' (Cognos Analytics, Cognos Dashboards, DataStage, Data Virtualization, Metadata Import: discovery, **Metadata Import: lineage**, Watson Query), and 'Tunneling protocol' (Satellite, Secure Gateway). The main area displays a grid of connection types, with a search bar and a filter for 'Metadata Import: lineage'. A yellow callout box points to the 'Metadata Import: lineage' filter, containing the text: 'Annotation This box is not part of the UI MANTA supported connection types (includes overlapping SCAPAPI types)'. At the bottom, there are 'Cancel', 'Back', and 'Next' buttons.

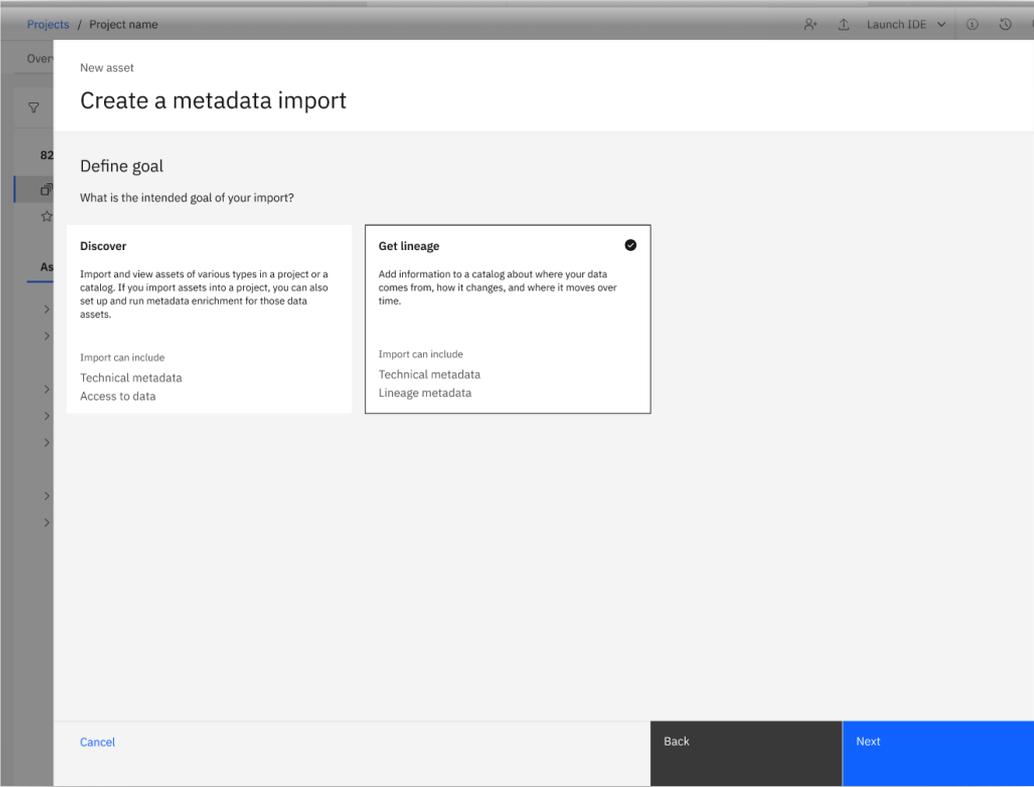
Under Compatibility, user can filter on:

- Metadata import: discovery
- Metadata import: lineage

“Metadata import: discovery” will include all existing supported data connection types, plus the new MANTA scanners for metadata-only import.

“Metadata import: lineage” will list only MANTA scanners.

Initiate lineage scan through metadata import in projects



MANTA discovered assets will be added directly to a catalog

New asset | Goal: Get lineage

Create a metadata import

- Select target
- Select sources
- Set schedule
Optional
- Define details
- Review import

Select target catalog

This project <project_name> Catalog

Which catalog would you like to import the metadata into?

Find catalogs

Catalog name	Role	Date created	Last modified	↑
<input checked="" type="radio"/> Online Shops	Admin	Jan 1, 2021, 4:32 pm	Dec 6, 2021, 5:18 pm	
<input type="radio"/> WRM Catalog	Admin	May 21, 2021, 10:30 pm	Dec 6, 2021, 2:45 pm	
<input type="radio"/> Test Catalog	Editor	Feb 9, 2021, 2:24 pm	Dec 2, 2021, 10:30 am	
<input type="radio"/> Finance team	Admin	Sept 12, 2021, 8:12 am	Nov 22, 2021, 4:21 am	
<input type="radio"/> Weather data	Editor	Aug 19, 2021, 9:36 pm	Nov 18, 2021, 9:36 pm	
<input type="radio"/> HR Catalog	Admin	Nov 2, 2021, 2:36 pm	Nov 12, 2021, 2:36 pm	
<input type="radio"/> Catalog Z	Editor	Jan 9, 2021, 1:11 pm	Nov 7, 2021, 1:11 pm	
<input type="radio"/> Marketing catalog	Admin	May 26, 2021, 8:40 am	Nov 1, 2021, 8:40 am	

[Cancel](#) [Back](#) [Next](#)

Imported assets and their lineage info will be added to the designated catalog

Metadata import for lineage : [Edit metadata import](#) [Reimport assets](#)

Metadata import

Overview

- 100% Completed
- ↑ 24 Assets imported
- 3 Errors detected
- ⌚ 01:14:08 Hours lapsed

Import for capturing lineage Navigate to each asset's lineage from the overflow menu or the target catalog. [View catalog](#)

Imported assets

24 assets

Search assets

Asset name	Asset type	Source	Last imported	Status
BANK_ACCOUNTS	BI Report	qadual / ... / EWS	Feb 10, 2021 12:18 pm	Imported
BANK_CUSTOMERS	BI Report Query	qadual / ... / EWS	Feb 10, 2021 12:18 pm	Imported
BANK_CLIENTS	BI Report Query	qadual / ... / EWS	Feb 10, 2021 12:18 pm	Imported
PerformanceMetrics	BI Report Query Item	qadual / ... / EWS	Feb 10, 2021 12:18 pm	Imported
MeasurementFacts	BI Report	qadual / ... / EWS	Feb 10, 2021 12:18 pm	Imported
DebiasedFairnessMetrics	BI Report Query	qadual / ... / EWS	Feb 10, 2021 12:18 pm	Imported

About this metadata import

Description
This metadata import brings assets from multiple connections for lineage.

Import details
Goal
Get lineage
Data connections
PowerBI for WKC
Metadata connections
PowerBI for WKC
Import target
[Online Shops](#)

Job details
Job name: [Custom import job name](#)
Last run: Run 1, Feb 10, 2021 12:18 pm

Schedule
Next run
Dec 8, 2020, 11 AM
Repeats
MonthlyFirst and last Monday at 11:00 AM

Related Assets 1
[Custom import job name](#)

Tags
Gray tag Longer name tag Another tag

Data consumers can view the data lineage in the catalog or launch the MANTA data lineage view

The screenshot shows the 'Lineage' tab for the 'BANK_ACCOUNTS' asset. On the left, a tree view shows the lineage from 'CUST_INFO' (2/24 assets) to 'CUSTOMER' (2/22 assets), which then leads to 'BANK_ACCOUNTS' (14 queries). Below this, 'BANK_DW' (1/24 assets) is also shown. The main area displays the 'About this asset' details for 'BANK_ACCOUNTS', including a description, asset owner (Matthias L. Chavanne), privacy (Public), asset details (Size: 500 MB), and contents (Charts: 5000+). A blue box highlights the 'Technical data lineage' link with the text 'Go to asset's technical data lineage'. A blue arrow points from this link to the right-hand screenshot.

Click on the URL to open MANTA data lineage in new tab

The screenshot shows the 'IBM Automatic Data Lineage' (MANTA) view. The main area displays a lineage diagram with nodes for 'BMO_CRR_EXT_RPT', 'dbo', 'V_FX_RATE', 'test', and 'Customer Information'. A line connects 'V_FX_RATE' to 'Customer Information'. The right-hand side shows an 'Options' panel with various filters and details, including 'Resources', 'Filters', 'Detail', and 'Colors'. The 'Filters' section is expanded, showing a list of selected filters such as 'Amazon S3', 'Business Glossary', 'Business Transformation', 'Cognos', 'Compliance Perspectives', 'Data Lake Perspectives', 'DataStage', 'DB2', 'DB2 DDL', 'DB2 PL/SQL', 'Erwin Data Modeler', 'Filesystem', and 'Hive'. An 'Apply' button is at the bottom right.

Thank you!

Pavel Maltsev
Technical Sales Specialist
—
pmaltsev@cz.ibm.com

© Copyright IBM Corporation 2022. All rights reserved. The information contained in these materials is provided for informational purposes only, and is provided AS IS without warranty of any kind, express or implied. Any statement of direction represents IBM's current intent, is subject to change or withdrawal, and represent only goals and objectives. IBM, the IBM logo, and ibm.com are trademarks of IBM Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available at [Copyright and trademark information](#).

