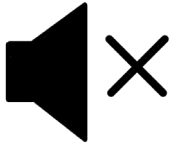


Aug 01, 2023

# What's New in Informatica Cloud Data Governance and Catalog (CDGC), Cloud Data Marketplace (CDMP) and Cloud Data Quality (CDQ)?

- Amrish Thakkar, Senior Director, Product Management
- Ian Stahl, Senior Director, Product Management
- Sean O'Shea, Senior Manager, Product Management

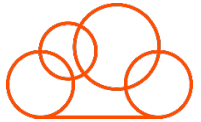
# Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our [Success Portal](#) - where you can also download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

# Feature Rich Success Portal

---



Bootstrap trial and  
POC Customers



Enriched Customer  
Onboarding  
experience



Product Learning  
Paths and Weekly  
Expert Sessions

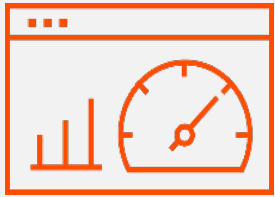


Informatica  
Concierge



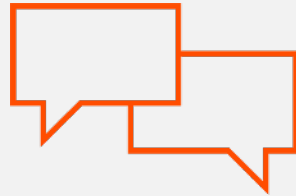
Tailored training and  
content  
recommendations

# More Information



## Success Portal

<https://success.informatica.com>



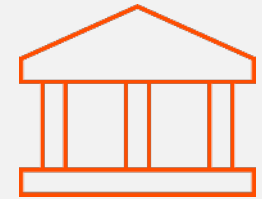
## Communities & Support

<https://network.informatica.com>



## Documentation

<https://docs.informatica.com>



## University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

# Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

# Cloud Data Governance and Catalog

# Features in CDGC 2023.07.M

- **Scanner Change policy**
  - Helps manage the catalog content to reflect changes on the Source system
- **Role based Default Dashboards**
- **Ticket Improvements**
  - New Ticket type - Data Quality Failure
  - Ticket Action - Resolve / Cancel
  - Commenting on Tickets
- **Dynamic Filter Support on Search**
  - Additional filters appear based on Asset type
  - All filterable fields including Custom Fields
  - Improves searchability and ease of use
- **Lineage Improvements**
  - **Progressive lineage loading**
  - Usability Improvements
- **Private Link Support for AWS**
- **Hierarchy Deletion\***
- **10+ New Scanners Added**
  - Also, enhanced Existing scanners
- **Profiling Updates**
  - SAP S/4 Hana, SAP ECC

\* - It will be delivered as EBF on July release

# Scanner Change Policy



# Scanner Change Policy

## Use Cases

- Allows customers to decide what happens the existing metadata when re-scanning

## Productivity

- No need to purge and rescan to sync metadata with the Source system
- Enrichment content will be saved when retained
- Customers can incrementally scan different source objects

## Low Cost

- It saves IPU burn
  - No purge required to keep Sync
  - Object storage count by only storing what is necessary
- It gives customers control

# Scanner Change Policy

## Feature description

- New behavior:
  - Catalog Source administrator can choose one of the following options
    - Delete – Objects removed from the source will be removed from the catalog
    - Retain – Objects removed from the source will be retained in the catalog
  - Catalog Source administrator can set the metadata change option in catalog source configuration and save for future scan execution
  - Catalog Source administrator can override the metadata change option at moment of execution
- SCP is applicable for all scanners

Metadata Change Option: \* ?

☒ Retain ☐ Delete

**i** Based on the objects deleted from the source and changes made to the filter, source changes are retained in the catalog. Objects deleted from the source are not deleted from the catalog.

# Scanner Change Policy

## Feature description

- Scanner behavior when metadata option is “delete”
  - When objects are deleted or filter criteria are changed to exclude previously loaded objects - Catalog objects will be deleted from the Catalog
  - When objects are deleted, any incoming or outgoing relationships are deleted
  - When objects are deleted, any enrichments are deleted and not recoverable
  - When objects are renamed in source, they are removed and recreated in catalog, all enrichments are lost
  - **Deleted relationships will be retained when source and target objects still exists in the source**

Metadata Change Option: \* ?

☐ Retain ☒ Delete

⚠ Based on the objects deleted from the source and changes made to the filter, metadata is deleted from the catalog. Enrichments made on deleted objects are also permanently lost.

# Dashboard updates

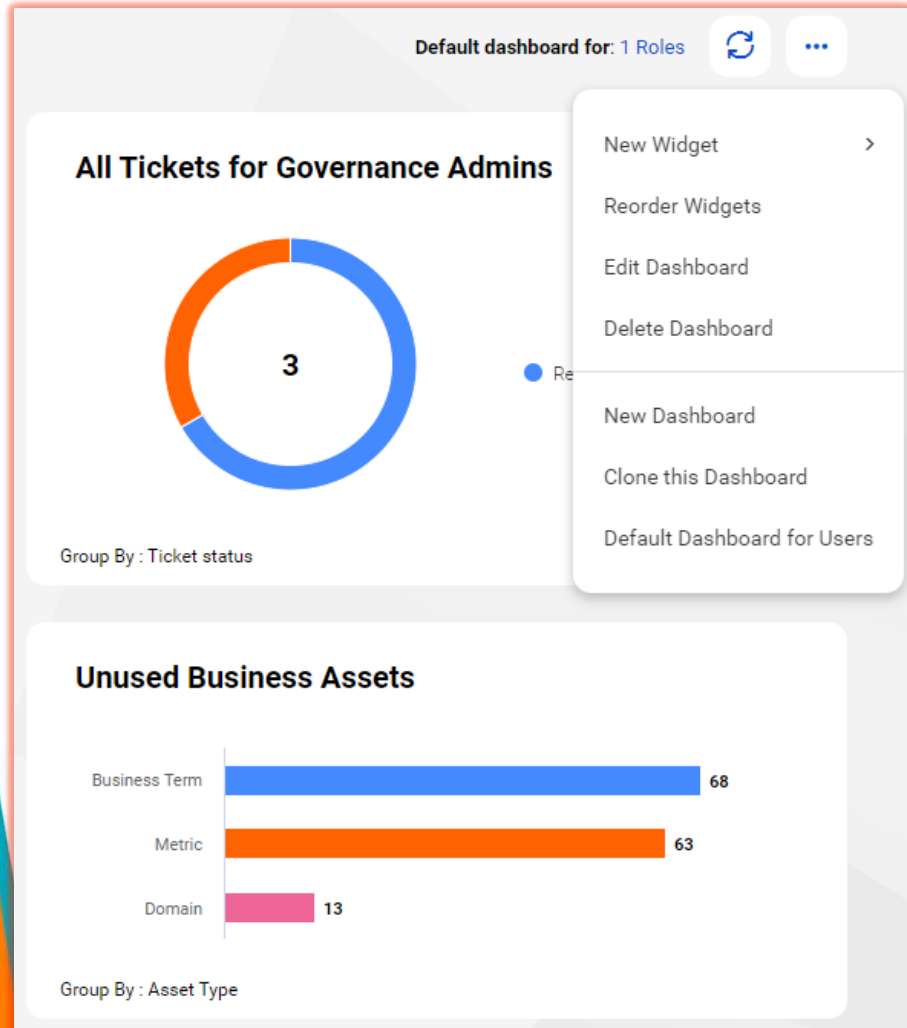
# Role Specific Dashboards

## Setting up your own Dashboard per Role

- Until now, there was a single dashboard for all users, regardless of their specific roles or responsibilities.
- With this feature, Role Specific Dashboards can be setup to drive persona based experience
- This will enable role based shared understanding with customized, shared dashboards representing most critical indicators
- Everyone with the Super Admin privilege can configure it.

# Role Specific Dashboards

## Super Admin: Dashboard Menu



- As part of July release, options are added to set the dashboards for specific roles or user personas.
  - Users having multiple roles will see dashboards for all the roles assigned to him/her.
- Super Admins have authority to update and replace the shared dashboards.
- All Users still retain the ability to create their own dashboards and make it default for themselves.

### Set Default Dashboard

☐ For all users ☒ For specific roles ☐ Remove dashboard for users

**i** Do you want to set 'Governance Admin Dashboard' as the default dashboard for specific roles in your organization? They can view this dashboard, you have earlier provided another default dashboard, 'Governance Admin Dashboard' will replace that dashboard.

If users have set their own dashboard as default, they will remain as default.

Select roles: \*

Governance Administrator Governance Data Steward

- ☐ GA\_CustomRole
- ☒ Governance Administrator
- ☒ Governance Data Steward
- ☐ Governance User

# Lineage Enhancements

# Lineage Enhancements - Summary

- **Usability Improvements**

- Opening assets from lineage diagram tooltip **opens lineage page** of that asset.
- Lineage context is maintained as well for business / technical lineage
- **Retain full screen** mode – when an asset is open from full screen
- Restrict **tooltip** to user hover over **asset name only** (not full rectangular node).
- **Highlight page numbers** to easily identify corresponding assets on different pages.
- Single and double click on hop to expand lineage diagram (Right click behaviour is removed).
- **Default filter popup** on data element level lineage for more specific lineage views.
- **Find** functionality works **across pages** and highlight the page number of searched asset.
- **Calculation Expression Overlay** : Display the calculation icon for view columns within the lineage diagram.
- Lineage performance improvements for large lineage diagrams.
- Loop Visualization when data flow is going back



# Lineage Enhancements

- **Loop is now shown in lineage.**
  - Loop option is now available in lineage “settings”.
  - Direct Sibling Data Flow : Improved lineage representation for data flows between direct siblings.
  - Indirect Sibling Data Flow : Addressed lineage rendering challenges for data flows between indirect siblings.
  - Parent-Child Data Flows :Enhanced visualization of data flows between parent and child assets.
  - Single or Multiple Hop loop is also handled.

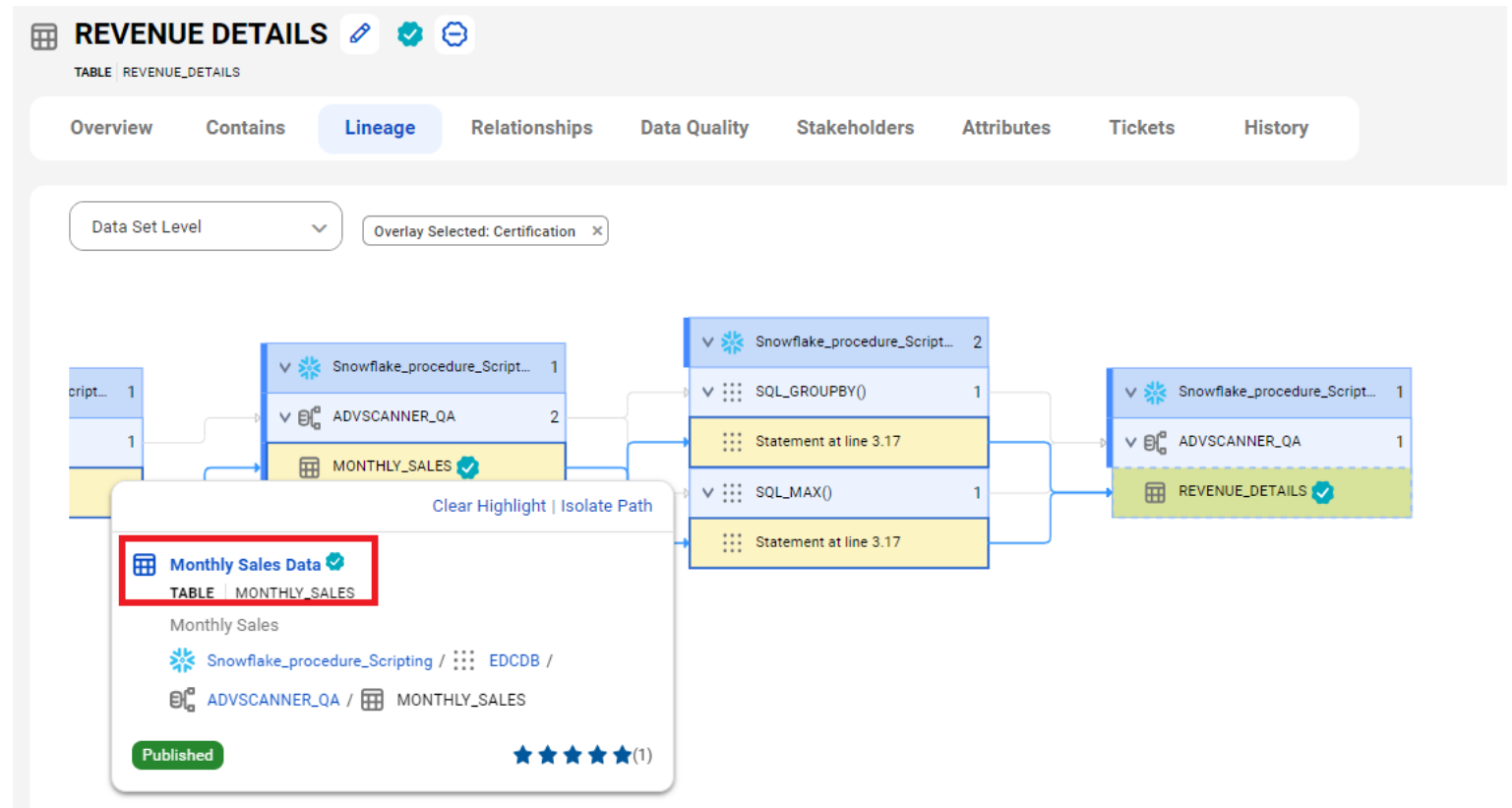
# Lineage Enhancements

2023.07 : Open Asset lineage from lineage diagram

When you hover over an asset, the asset tooltip appears which displays details of the asset.

- Now, you can open lineage of any asset directly from this panel.
- full screen lineage view will be retained for user - when user opens assets from lineage diagram.

*\* If there is no lineage for the asset default asset overview page is opened.*

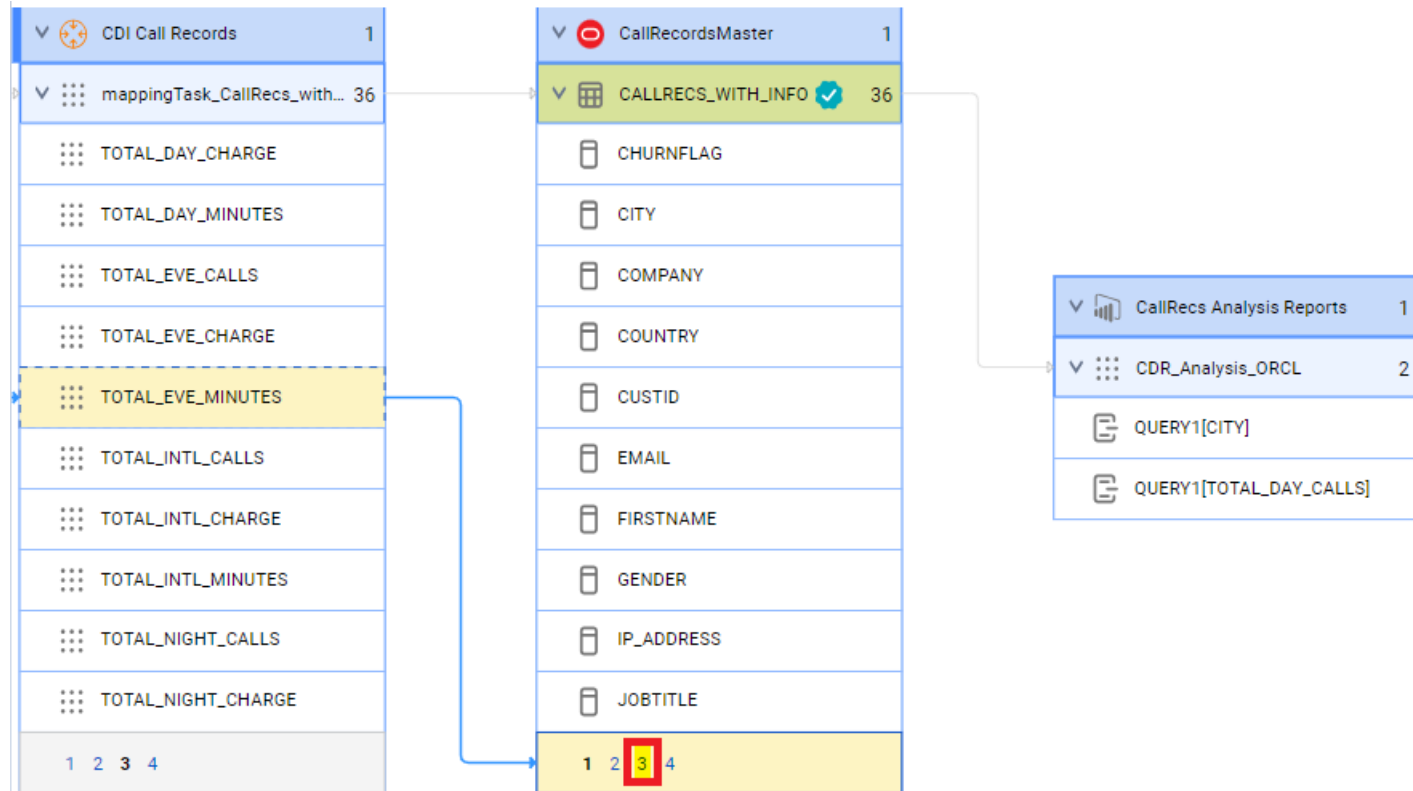


# Lineage Enhancements

2023.07 : Highlight page number in lineage diagram

When you click on asset in multi page objects list in lineage diagram – page number is highlighted.

\* The page number is highlighted if asset is not one the same page.

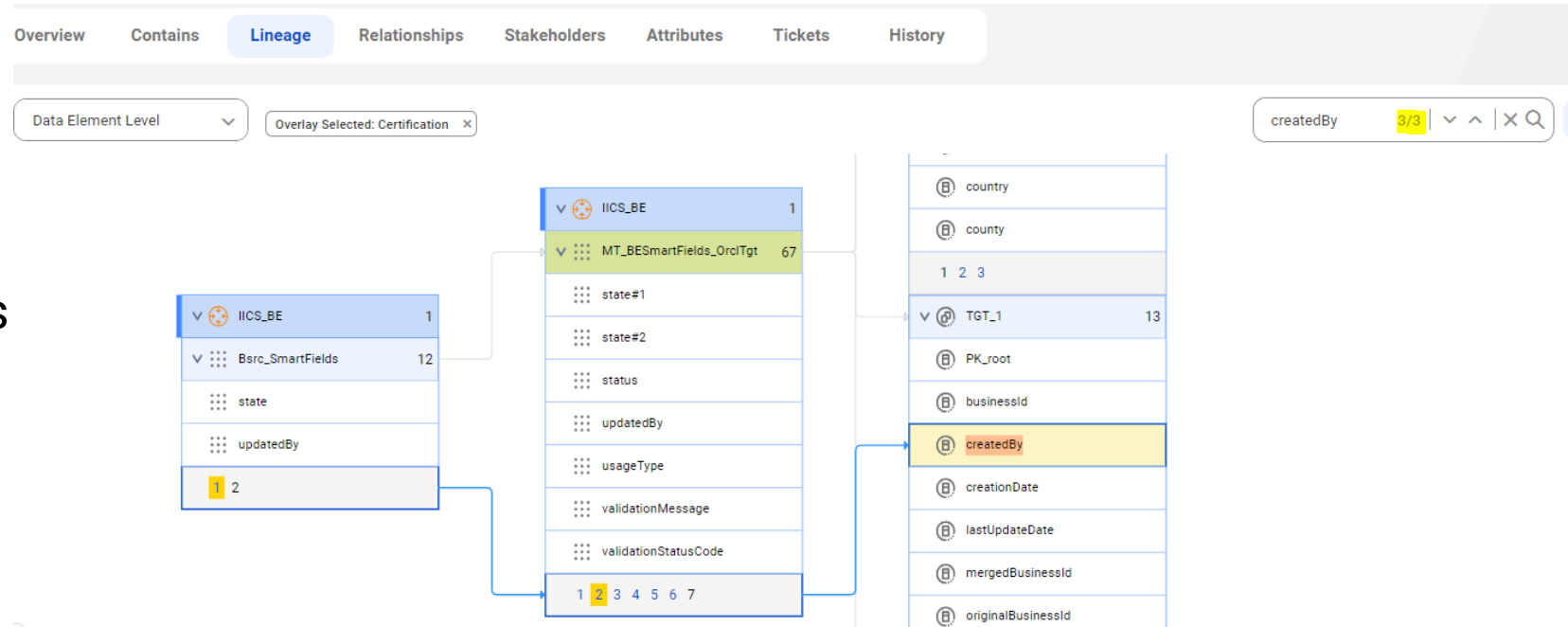


# Lineage Enhancements

2023.07 : Find across pages - Highlight page number of search assets

When you find asset in lineage the search works across all pages and not only visible asset.

Page number of the asset is highlighted as well.



# Lineage Enhancements

2023.07 : Default filter popup on data element lineage

When you switch to data element lineage default filter popup is shown to easily select only necessary column in lineage.

- You can filter data elements from any dataset in lineage.
- You can select all columns as well.
- You can remove the applied filters from lineage diagram

## Filter

Select Data Set and then select Data elements from Snowflake\_procedure\_Scripting / ... / TABLE\_FOR\_BIND / to filter the view

Select Data Set:

TABLE\_FOR\_BIND

Data Element

TABLE\_FOR\_BIND  
Statement at line 15.13  
Statement at line 18.13

Find

<input type="checkbox"/>	Name	Type	Description
<input type="checkbox"/>	DATE1	Column	-
<input type="checkbox"/>	FLOAT1	Column	-
<input type="checkbox"/>	INT1	Column	-
<input type="checkbox"/>	NUMERIC1	Column	-
<input type="checkbox"/>	TIME1	Column	-
<input type="checkbox"/>	TS1	Column	-
<input type="checkbox"/>	TS_NTZ1	Column	-
<input type="checkbox"/>	V	Column	-

TABLE TABLE\_FOR\_BIND

Overview Contains Lineage Relationships Data Quality Stakeholders Attributes Tickets History

Data Element Level

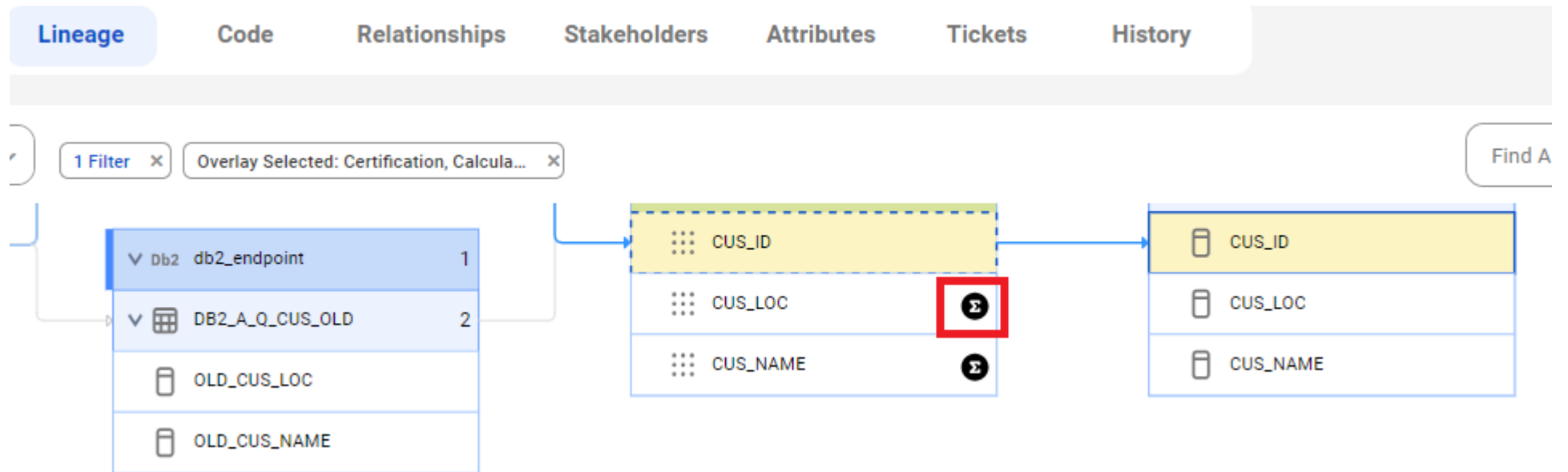
1 Filter

Overlay Selected: Certification

# Lineage Enhancements

2023.07 : Lineage Calculation Overlay on Views

Lineage Calculation Expression overlay is extended to views as well.



# Lineage Enhancements

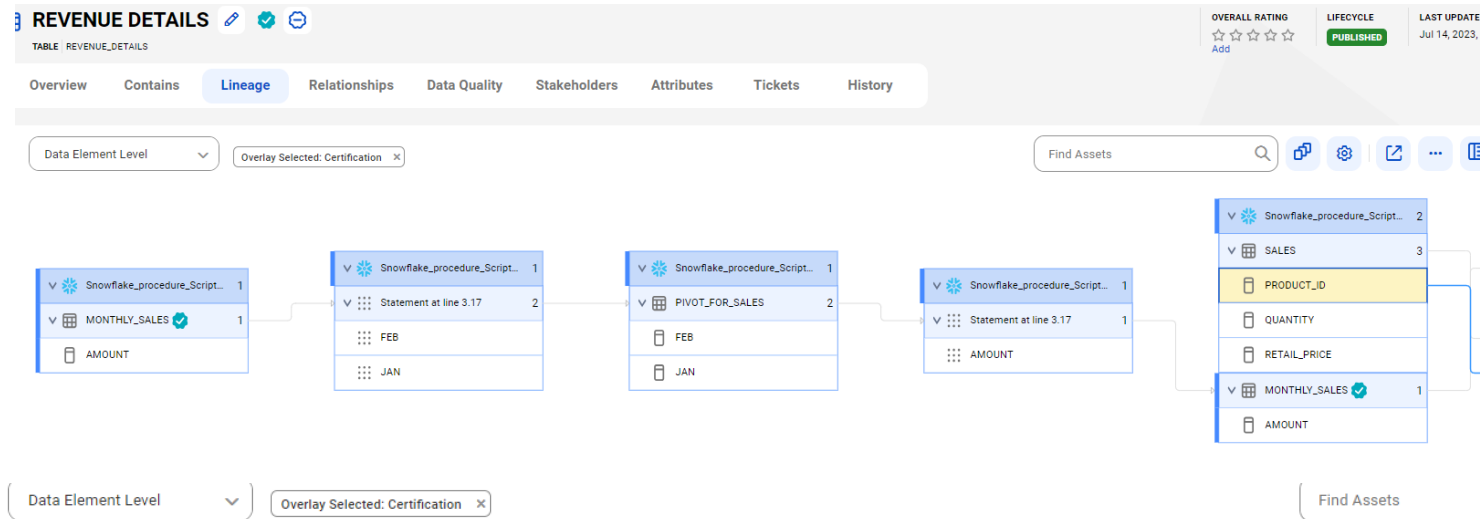
2023.07 – Lineage with Loop back

- **Loop is now shown in lineage.**
  - Loop option is now available in lineage “settings”.
  - Direct Sibling Data Flow : Improved lineage representation for data flows between direct siblings.
  - Indirect Sibling Data Flow : Addressed lineage rendering challenges for data flows between indirect siblings.
  - Parent-Child Data Flows :Enhanced visualization of data flows between parent and child assets.
  - Single or Multiple Hop loop is also handled.

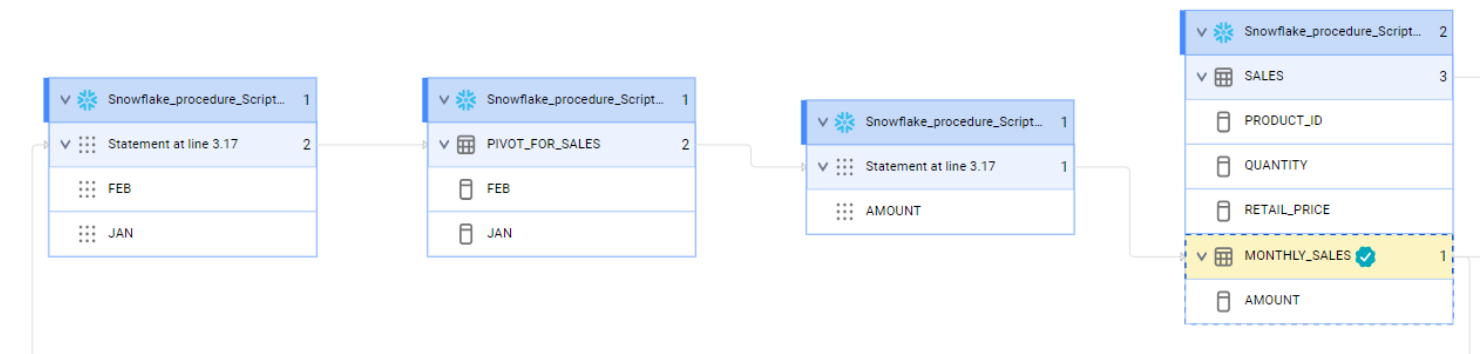
# Lineage Enhancements

2023.07 : Loops are shown in lineage diagram

## Without Loops



## With Loops

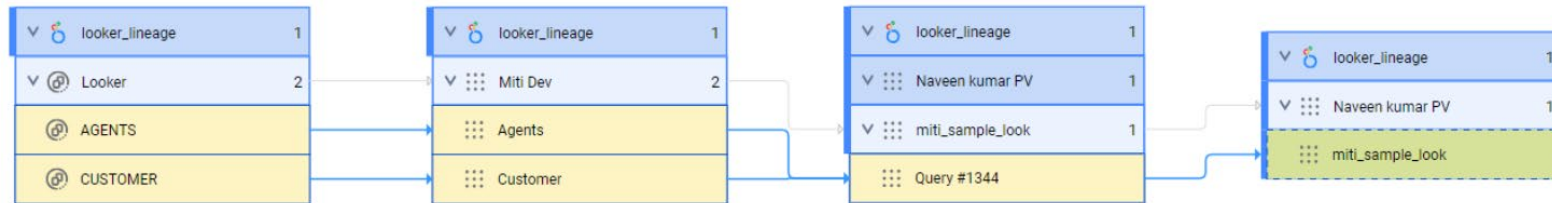




# Lineage Enhancements

2023.07 : Loops are shown in lineage diagram

## Without Loops



## With Loops



# Other improvements

# Dynamic Search Filter

- Provides additional filters for the user to refine the search results
  - These filters will show up when the user applies Asset Type filter
  - Provides all common fields when multiple asset type is selected
- Until now some of these filters could only be applied using the Search Query
- All fields except the Text fields are filterable
  - Includes Dropdown, Boolean, Date, Number

The screenshot displays the Informatica search interface. On the left, the 'Filters' panel is active, showing a list of asset types under 'Asset Type (131)'. A red box highlights the 'Additional Filters' section, which includes 'Critical Data Element' (with 'True', 'False', and 'No Value' buttons) and 'Format Type' (with a 'Select' dropdown). On the right, the 'Search Results (188)' panel shows a list of results. A red box highlights the first result, 'Item ID', which is a 'BUSINESS TERM' at the 'Top Level'. Below it, other results like 'Quantity of Product Sold', 'Date Ordered', and 'Department' are visible. The bottom of the interface shows pagination: '1 - 25 of 188 Items', 'Page 1 of 8', and 'Items Per Page: 25'.

# Ticket Improvements

- Data Quality failure Ticket
  - Ability to create Data Quality failure ticket for DQ Rules Occurrence to start the remediation process
  - Tickets are assigned to stakeholder from the Data element / Rule occurrence on which it is created
  - Stakeholder will be able to do Data Preview for the data elements if enabled in MCC
  - No workflow can be attached to this ticket type in this release
- Ability to Cancel ticket
  - Stakeholders and Super Admin will be able to close the tickets using the RESOLVE button
  - Super Admin and requesters can withdraw the raised ticket using the CANCEL button
  - Email notifications are enabled to inform the requester/stakeholder on the ticket status
- Driving Discussions on both Open and Resolved tickets
  - Comments can be added to both open and resolved ticket to communicate with the requester or the Stakeholder
  - All users having access to Ticket Manager can be involved in discussion on tickets from the comment panel.
  - All users can be tagged in the comment panel.

# Hierarchy Deletion

Part of EBF on top of 2023.07

- Promotes rapid deletion of multiple objects from CDGC, where appropriate
- Hierarchy delete will be available to Super Admins only.
- Super Admin users will be able to delete an object and the complete hierarchy below it
  - It will delete all relationships (inbound and outbound), associated Tickets (closed only)
  - Any open tickets in the hierarchy must be closed before hierarchy delete permitted.
- Valuable for Governance team if they wish to alter the direction of a Governance program e.g., with a reorganized operating model to help promote longer term engagement and adoption

# Profiling Support

- Profiling and Data Quality Support Added for
  - SAP ECC
  - SAP S/4 HANA
- Profiling enhanced to support CSV File with Shift-JIS codepage

# AWS Private Link Support

## Feature description

- Support the AWS Private link architecture implemented by IDMC platform
- Without Private link
  - Communication between Customer environment and IDMC platform is happening through public internet
  - Communication are encrypted with SSL
  - IP addresses should be whitelisted by customer to allow communication with the IDMC platform
- With Private link
  - Communication will be routed through the AWS backend between customer VPC and Informatica VPC
  - Communication will continue to be encrypted with SSL
  - More complex setup is needed as described in the [Platform documentation](#)
- Support Coverage
  - For CDGC, Serverless scan is not certified with private link, all other functionalities are supported

# Scanners



# Scanner Summary

## Scanners Added

- Erwin Data Modeler
- SAP Power Designer
- ER/Studio Data Architect
- AWS Glue (Tech preview)
- Atlas (Tech Preview)
- Microsoft Purview (Tech Preview)
- HDFS/Hive (Cloudera Data Platform)
- Additional DB Scripts
  - Microsoft Azure SQL Server, Microsoft Azure Synapse, Teradata FastLoad, Teradata FastExport, Teradata MultiLoad

## Scanners enhanced

- ADF addons
- PowerBI Dataset Field Desc
- Snowflake Scripting
- Cloud Data Integration and Informatica Power Center: Addl Connectors
- Namespace and PK/FK support for SAP scanners
- DB2 LUW (Support for programmable objects)
- Teradata (Support for programmable objects)

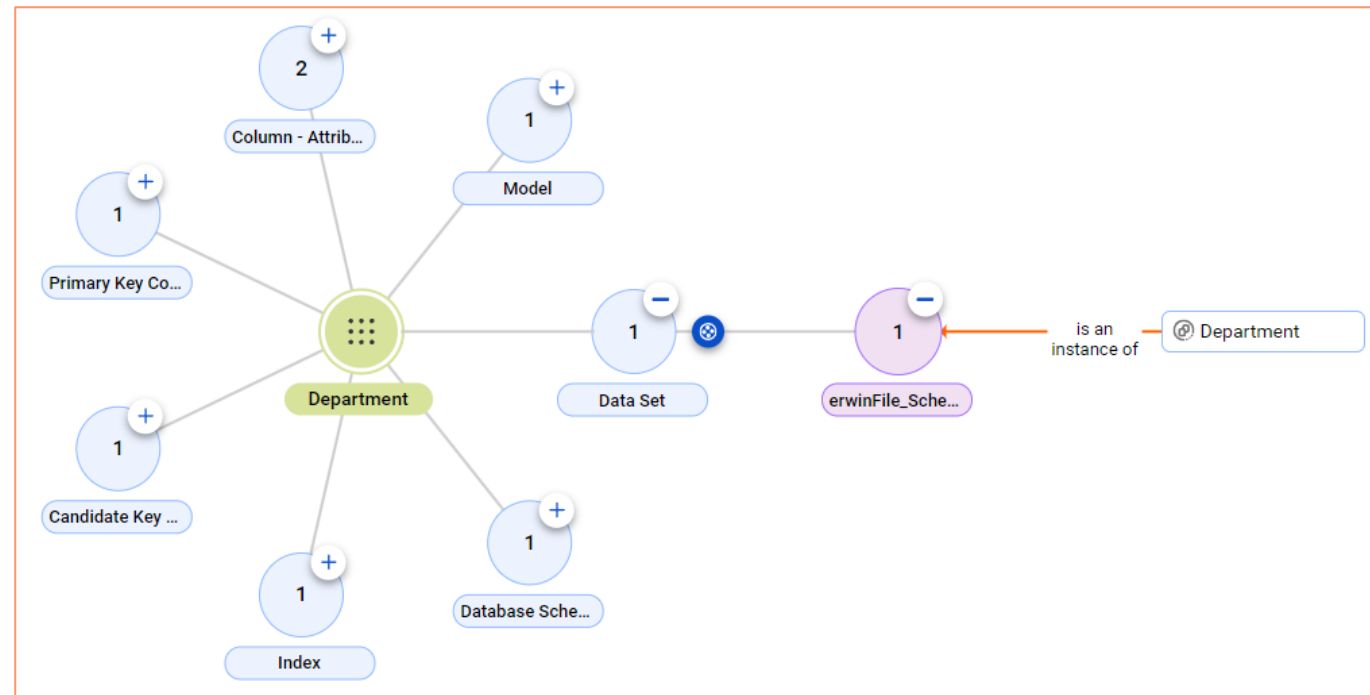
# Scanners: erwin Data Modeler scanners



## erwin Data Modeler File

A data modeling tool used to find, visualize, design, deploy and standardize data assets.

- erwin Data Modeler is a data modeling tool used to find, visualize, design, deploy and standardize data assets
- For tables and views from physical model reference objects are generated to make it possible to stitch them with database assets already present in the catalog
- The scanners extract from erwin models both logical and physical assets together with relations between them.
- There are two possible approaches
  - First is to save the models metadata to the output file using erwin Data Modeler tool
  - Second one is to connect directly to the erwin Mart Server which is a web-based repository of erwin models



# Scanners: SAP PowerDesigner scanner



**SAP PowerDesigner**

A collaborative enterprise data modeling tool

- SAP PowerDesigner is a collaborative enterprise data modeling tool that enables enterprises to visualize, analyze, and manipulate metadata for effective enterprise information architecture.
- The scanner reads models metadata from the physical data model (PDM) files prepared using SAP PowerDesigner tool and placed on the Secure Agent machine
- The scanner extracts physical assets and relations available in the model.
- For tables and views from physical model reference assets are generated to make it possible to stitch them with database assets already present in the catalog

## Connection Information

Catalog Source Type:

SAP PowerDesigner

File: \* ?

E:\SAPPDQA\sappd\_test.pdm

## Configuration Parameters

Apply binary characters filter: ?

☐ True ☒ False

Import UDPs: ?

As metadata



Append volumetric information to the description field: ?

☐ True ☒ False

Remove text formatting: ?

☒ True ☐ False

# Scanners: IDERA ER/Studio Data Architect scanner



**IDERA ER/Studio Data Architect**

A data modeling tool for designing and understanding databases.

- IDERA ER/Studio Data Architect is an enterprise tool for data architects to design and document data assets
- The scanner extracts models metadata from .DM1 files prepared using ER/Studio Data Architect tool and placed on the Secure Agent machine
- User can choose which models should be imported: “All models”, “Logical models” or “Physical models”
- The scanners can extract both logical and physical assets and relations between them
- For tables and views from the physical model reference assets are generated to make it possible to stitch them with database objects already present in the catalog

**Configuration Parameters**

Model selection: ?

All models

×

▼

Import UDPs: ?

As metadata

×

▼

Import owner schemas: ?

☒ True ☐ False

Remove text formatting: ?

☒ True ☐ False

Miscellaneous Options: ?

# Scanners: Apache Atlas scanner (Tech Preview)



Apache Atlas (Preview)

A scalable and extensible set of core foundational governance services

- Apache Atlas is the governance and metadata framework for Hadoop. Apache Atlas has a scalable and extensible architecture that can be plugged into many Hadoop components to manage their metadata in a central repository
- To extract metadata from Apache Atlas environment the following items should be provided:
  - Base URL to connect to the Atlas REST API
  - Kerberos principal used for authentication
  - Kerberos keytab file and configuration file located on the Secure Agent machine

## Connection Information

Catalog Source Type:

Apache Atlas (Preview)

Base URL: \* ?

http://testatlasinstance.informatica.com:31000/

Principal: \* ?

atlas/vmforsapna.informatica.com@INFAKRB.INFADEV.COM

Keytab File Path: \* ?

/data/atlas.service.keytab

Configuration File Path: \* ?

/data/krb5.conf

# Scanners: Apache Atlas scanner (cont...)



Apache Atlas (Preview)

A scalable and extensible set of core foundational governance services

- The following metadata is extracted by the initial version of the scanner from an Apache Atlas source system:
  - Atlas Server
  - Hive Process
  - Sqoop Process
  - Calculation (extracted only when there is column level lineage available in Atlas catalog)
- The initial version of the scanner creates reference assets for the following data sources:
  - Apache Hive
  - Hadoop Distributed File System (HDFS)
  - Oracle
- Apache Atlas catalog source is available for preview – GA is planned for later release

# Scanners: Microsoft Purview scanner (Tech Preview)



**Microsoft Purview (Preview)**

A unified data governance, risk, and compliance solution

- A unified data governance, risk, and compliance solution to manage and govern on-premises, cloud and software as a service (SaaS) data by enabling data discovery, traceability, and searchability
- The scanner extracts Microsoft Purview collections hierarchy and creates reference assets for data sources existing in the collections.
- To extract metadata from Microsoft Purview catalog the following service principal details should be provided:
  - Client ID
  - Tenant ID
  - Client Secret
  - Account Name

## Connection Information

Catalog Source Type:

Microsoft Purview (Preview)

Client ID: \*

xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx

Tenant ID: \*

xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx

Client Secret: \*

.....

Account Name: \*

ScannersQAPurview

# Scanners: AWS Glue Scanner (Tech Preview)



## AWS Glue (Preview)

A serverless ETL service from Amazon Web Services to discover, prepare, and combine data from multiple data stores

- AWS Glue (Tech Preview) is a new Scanner to extract Python Glue jobs related metadata

### Connection Information

Catalog Source Type: AWS Glue (Preview)

Athena Connection: \*

athena-us-west-2

Access Key: \* ?

.....

Security Key: \* ?

.....

Region: \* ?

us-west-2



### Filters

Specify metadata filters.

☐ No filters ? ☒ Filter conditions ?

> Show supported wildcards and examples

Include Metadata

JobName

CreateAgentTable\*



# Scanners: Hive (Cloudera Data Platform)



## Apache Hive

A distributed data warehouse system built on Apache Hadoop to provide data query and analysis using an SQL-like interface.

- Hive CDP is a new Scanner to extract metadata from Hive database catalog as well as HDFS filesystem.
- For Hive, with the current version it supports following datasets:
  - Tables
  - Views
  - External tables
  - constraints (PK, FK)

The screenshot displays the Cloudera Data Platform (CDP) interface. On the left, a sidebar shows navigation options: New..., Home, Browse, and Hive\_Kerberos (selected). The main content area is divided into two sections. The top section, titled 'Hive\_Kerberos' (APACHE HIVE CATALOG SOURCE), has tabs for Overview, Hierarchy, Relationships, Stakeholders, Attributes, Tickets, and History. The bottom section, titled 'HDFS\_Kerberos' (HADOOP DISTRIBUTED FILE SYSTEM CATALOG SOURCE), also has similar tabs. The 'Overview' tab for HDFS\_Kerberos is active, showing a list of file systems and a detailed view of 'Flat File (4)'. The file system list includes File System (1), Flat Field (46), Flat File (4), and Folder (4). The 'Flat File (4)' view shows a table with columns: Name, Description, and Path. The table lists four files: CUSTOMER\_Semicolon1.csv, ForAll\_InferredAssociationFile123.csv, ORDER\_DETAILS\_COLON1.csv, and ORDERS\_PIPE1.csv, all with descriptions '-' and paths starting with 'HDFS\_Kerberos / ... / Nestec'.

File System	Count
File System	1
Flat Field	46
Flat File	4
Folder	4

Name	Description	Path
CUSTOMER_Semicolon1.csv	-	HDFS_Kerberos / ... / Nestec
ForAll_InferredAssociationFile123.csv	-	HDFS_Kerberos / ... / Nestec
ORDER_DETAILS_COLON1.csv	-	HDFS_Kerberos / ... / Nestec
ORDERS_PIPE1.csv	-	HDFS_Kerberos / ... / Nestec

# Scanners Updates: Microsoft Azure Data Factory



## Microsoft Azure Data Factory

A cloud-based ETL and data integration service that is hosted on Microsoft Azure.

- Operational metadata support
- Support for Delta Lake as a source/target in data flows
- Support for Databricks Notebooks executed from within ADF processes

```

    "id": "/SUBSCRIPTIONS/18EC9AD2-B35C-4FB1-B967-542FE58360DF/RESOURCES/cd35-47fc-813c-a732a4ee931b",
    "runId": "065fdaaa-cd35-47fc-813c-a732a4ee931b",
    "debugRunId": null,
    "runGroupId": "065fdaaa-cd35-47fc-813c-a732a4ee931b",
    "pipelineName": "ForEach_Day_SM Staging To DL",
    "parameters": {
      "pYear": "2021",
      "pDateKey": "2"
    },

```

▼

Operational Metadata Config

Should process operational metadata: \*

?

☒ Yes ☐ No

Operational metadata time range mode:

?

Last 30 days

X

▼

Should process the latest operational metadata run only:

?

☐ Yes ☒ No

Resource group name: \* ?

# Scanners Update: CDI/Power Center

## IICS/PowerCenter enhancements:

- Added support for set of new connectors:
  - Generic JDBC Adapter Reader (IICS)
  - SAP BW Reader (IICS)
  - RESTv2 Reader (IICS)
  - RESTv2 Writer (IICS)
  - SAPBexQuery Reader (IICS)
  - Salesforce Reader (PWC)
  - Salesforce Writer (PWC)



### Informatica Intelligent Cloud Services

An on-demand subscription cloud service that provides access to Data Integration and Business 360 assets.



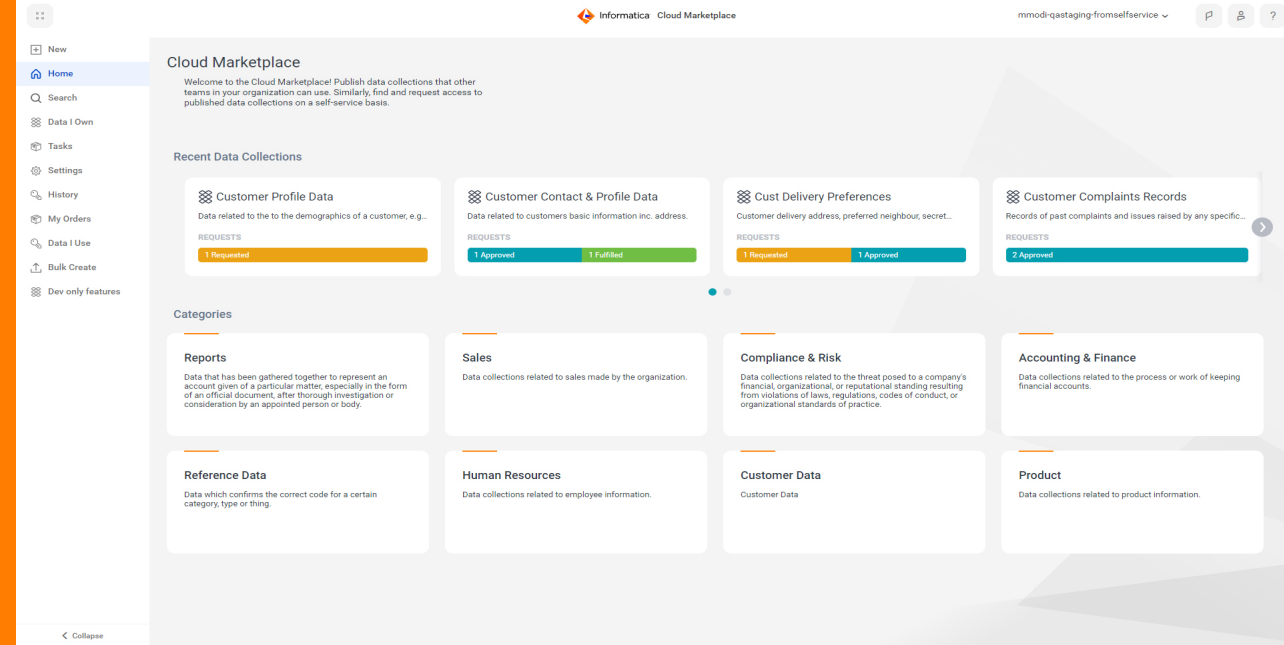
### Informatica PowerCenter

An ETL tool that is used to extract data from multiple sources, transform, and load them into different target databases.



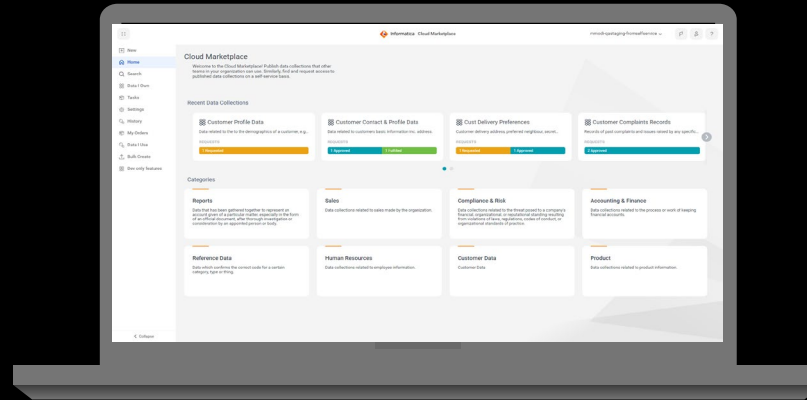
# Cloud Data Marketplace Summary of July Release Features

# Cloud-First, Cloud-Native, intelligent, governed Data Sharing for the Enterprise



# Informatica Cloud Data Marketplace

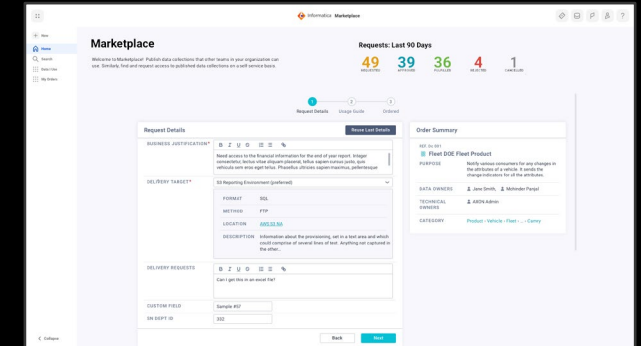
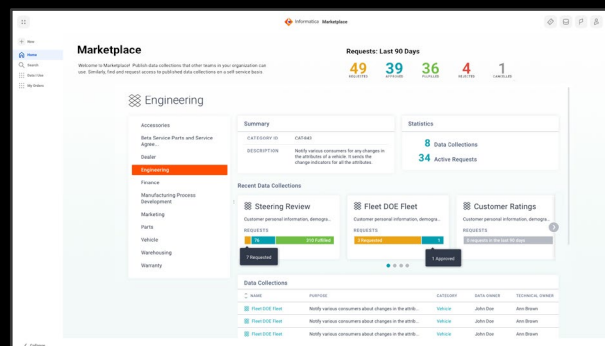
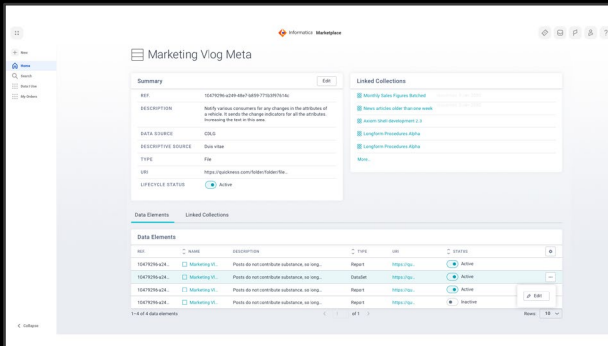
## Cloud First/Cloud Native, intelligent, enterprise-scale, governed Data Sharing



Package Diverse  
Data Assets

Convenient  
Data Shopping

Automated  
Delivery & Fulfillment



# July Release

Okta\_CDMP\_DCTechOwner\_Group

Select group members, and assign roles to define the application permissions.

Group Information

Name: Okta\_CDMP\_DCTechOwner\_Group

Description:

Mirrors the SAML group: Okta\_CDMP\_DCTechOwner\_Group

SummaryData AssetsDeliveryTerms of UseConsumersApproveFulfillment

Data Product Summary ⓘ

CancelSave

NAME \*EMEA Q1 Sales Data

PURPOSE \*

B I U G ☰ ☷ 🔗 +

Business A

Asset Type

Order

Request

Approve

View the orders that await your approval.

Orders

Data Assets

NAME	DESCRI...	DAT... <i>i</i>	DES... <i>i</i>	TYPE	SOU... <i>i</i>	SOU... <i>i</i>	TEC... <i>i</i>	STA...	CREATED	DATA QUALITY <i>i</i>	
DAss...	DAssetS...	DA_Ena...	DA_Ena...	DA_Ena...	-	-	-	E...	Jun 6, 2...	-	...
1.ASD	1.ASD	1.ASD	-	1.ASD	-	-	-	E...	Jun 19, ...	-	...

Data Elements of 'DAssetSearchName221686058375151'

NAME <i>^</i>	DESCRIPTION	TYPE <i>^</i>	TECHNICAL ... <i>i</i>	TECHNICAL ... <i>i</i>	STATUS <i>^</i>	DATA ASSET	DATA QUALI... <i>i</i>
1. Element	-	-	-	-	ENABLED	DAssetSearc...	-
CDM Element	-	-	-	-	ENABLED	DAssetSearc...	-
DElementSea...	DElementSearch...	CDMP	-	-	ENABLED	DAssetSearc...	-
cdmp	-	-	-	-	ENABLED	DAssetSearc...	-

1 - 4 of 4 Data Elements

1 of 1

Rows: 10

Thank you



# Agenda

1

Introduction

2

Cloud Data Quality

3

Cloud Data  
Profiling

4

Demo

5

6

# What is new for Cloud Data Quality & Profiling?

## Cloud Data Quality

- Rule Specification - Advanced Mode
- Deduplicate Populations: Extended coverage for countries & locales (63 options now)
- New Data Quality Bundle for “Australia and New Zealand”
- Support for Private link on Azure ecosystem

## Cloud Data Profiling

- 
- 
-

# Rule Specification – Advanced Mode

## *Customer driven requirement*

### A Low-Code option for Rule Specification

- Reduce time to value for rule logic development, using Rule Advanced option
- Empowers business users address more complex use cases
- An easy-to-use scripting capability to write business rule logic
- Aligns to the market trend of Low-Code

### Productivity

- With an intuitive user interface, allows all users to implement complex logic
- Business users with Excel Macros experience can build rules in Advanced mode
- Generate rule externally & copy/paste

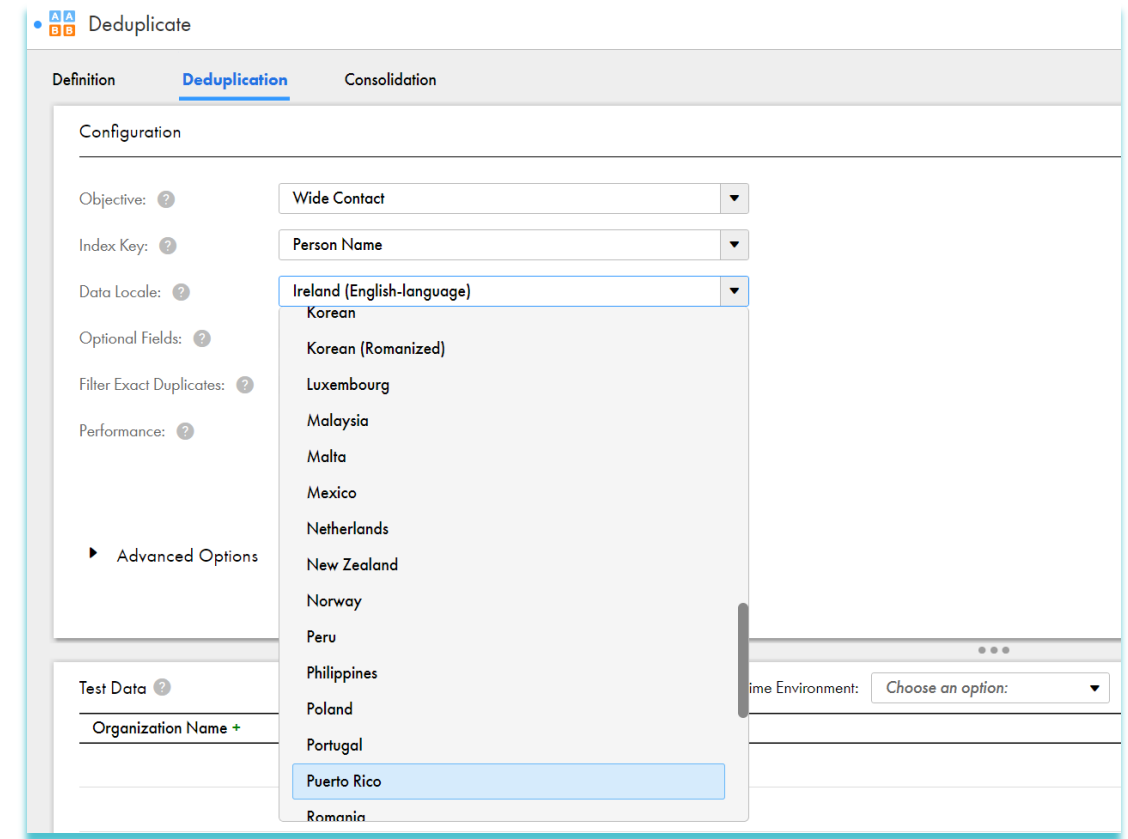
### Technical

- Script editor supports syntax highlighting, code suggestions, inline function/mapplet help/description to avoid to reference to the documentation
- Conditional logic, with Nested IF-THEN-ELSE capabilities
- Support to call other rules, dictionaries, mapplets and complex expressions in your expression logic
- Support for Decision Transformation

# Deduplicate Populations – New Countries and Locales

## 53 new populations

- **Added support for 53 new populations in Deduplicate:**
  - Total number of population now is 63 (10 existing)
- All the populations will list when user expands **Data Locale Dropdown.**
- All the dropdowns are now searchable so user can also search with the configuration field name and select
- **Populations order will be - Existing old populations followed by new populations in Alphabetic order**
- New population supports:
  - Deduplicate Test + Consolidation Test + Mapping execution in secure agent
  - Execution in Advanced DQ (Elastic), Serverless and Advanced (Elastic) Serverless



# Cloud Data Profiling

# Profiling of a “source mapplet”

## USE CASE

- Profile joined tables without creating a view on source
  - Example: Master and Details sources profiled in one profile execution
- Profile federated sources
  - Example: profile a flat file with codes joined with a relational source with supplementary data
- Compare two sources in one profiling
  - Join account table in Salesforce with Customer table in Oracle and compare statistics of columns

## SOLUTION

- Create a source mapplet and use it as a source for profiling

The screenshot displays the Informatica Data Integration Designer interface for a source mapplet named "Mapplet\_MultipleDiffSource". The top bar indicates the mapplet is "Valid".

**Design View:** A flow diagram showing the data flow. It starts with two input sources, "Master" and "Detail", which feed into "Filter\_Gold\_Records" and "Expr\_Clean\_Key" respectively. These two processed streams then merge into a "Joiner" component, which finally outputs to an "Output" component.

**Asset Details:**

- Name: Demo\_Profiling\_Source\_Mapplet
- Description: Profiling a mapplets with multiple sources, joiner, expression and filter transformations
- Location: \_RNS

**Source Details:**

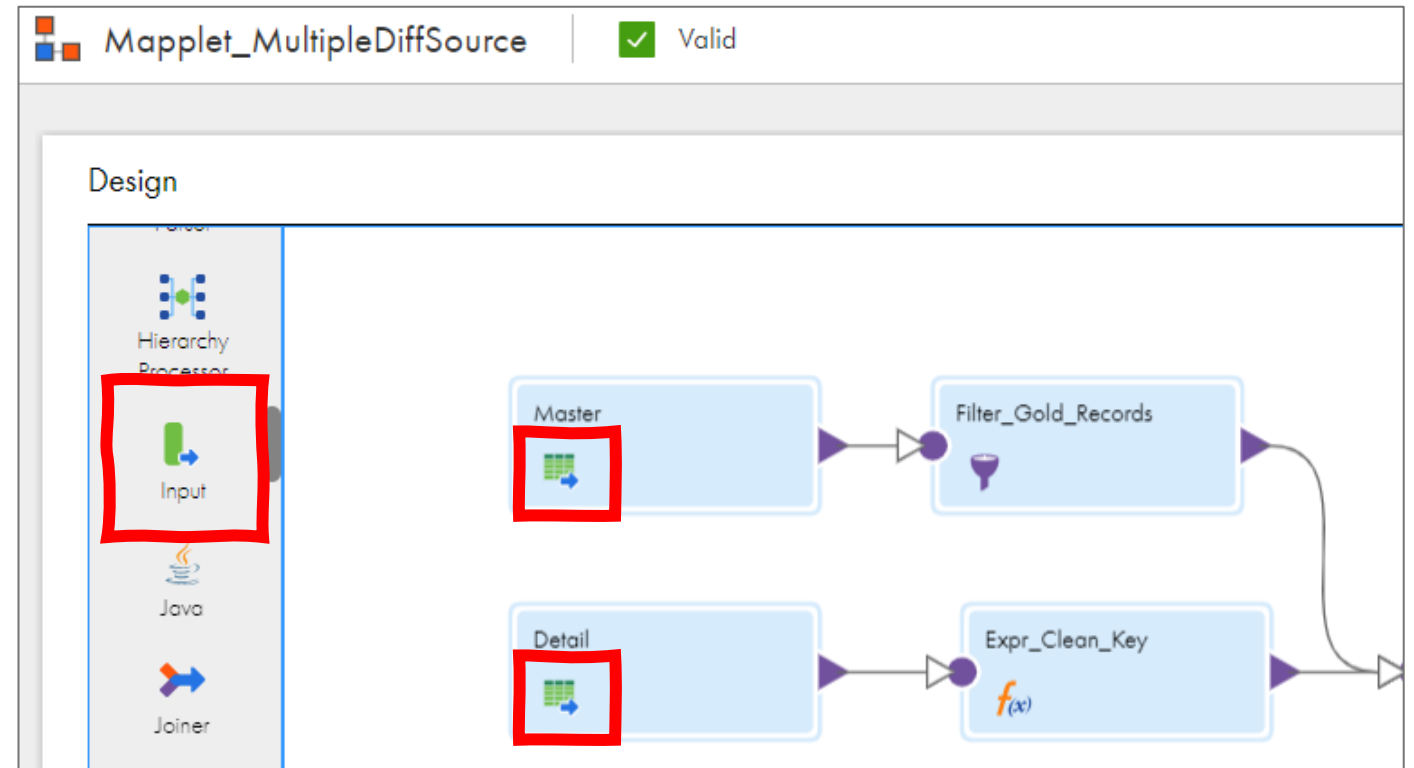
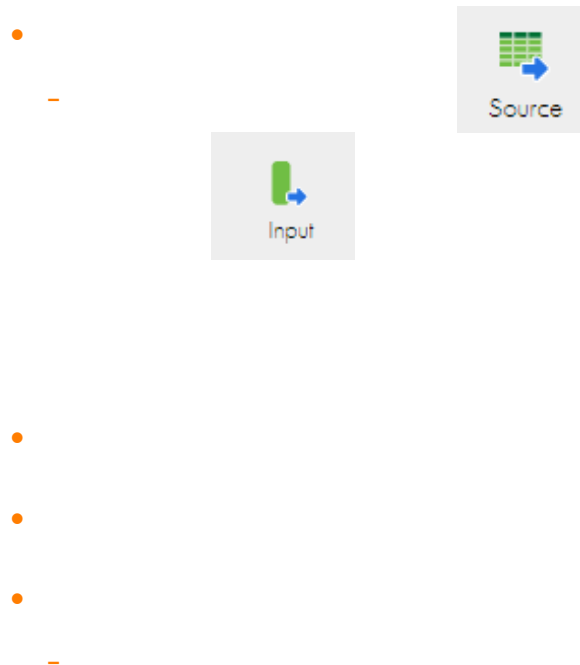
- Connection: Mapplet
- Source Object: Mapplet\_MultipleDiffSource [\_RNS]
- Enable CLAIRE capabilities: ☒
- Profile Settings: Drill down: ☒ Yes ☐ No

**Columns:**

Select columns to run the profile on. (Selected 3 of 3)

Columns	Type	Native Data Type	Precision	Scale	Nullable	Key
<input checked="" type="checkbox"/> Columns						
<input checked="" type="checkbox"/> Out_FIRSTNAME	string	string	25	0		
<input checked="" type="checkbox"/> Out_LASTNAME	string	string	25	0		
<input checked="" type="checkbox"/> Out_PHONE	string	string	20	0		

# What is a “source mapplet” and what Tx are supported?



# Cloud Data Profiling

## Support Data Quality Rules with multiple output ports in scorecards

The image displays two screenshots of the Informatica Cloud Data Profiling interface, specifically the 'Create Rule Occurrence' dialog box.

**Left Screenshot: Rule Outputs Selection**

The dialog box is titled 'Create Rule Occurrence' and has the subtitle 'Select rule outputs you want to score.' Below this, there is a section labeled 'Rule Outputs' with a table listing available outputs. The 'isValid' output is highlighted with a red box and a green checkmark, indicating it is available for selection.

Name
demo_multi_input_port
isValid ✓ Available
out_min_salary
out_max_salary
rs_OUT_SALARY_ColumnLengthDeviation_1
IsValid

**Right Screenshot: General Details Configuration**

The dialog box is titled 'Create Rule Occurrence' and has the subtitle 'Specify general rule occurrence details and define valid threshold values for scorecard generation.' Below this, there is a section labeled 'General Details' with the following fields:

- Name: VALIDITY\_OUT\_FIRST\_NAME, OUT\_SALARY\_isValid
- Dimension: Validity
- Description: Advanced mode rule with multi-input and multi-output port
- Rule Name: demo\_multi\_input\_port
- Output Port: isValid (highlighted with a red box)

Below the 'General Details' section, there is a section labeled 'Rule Occurrence Thresholds' with the following values:

95 to 100 % Good

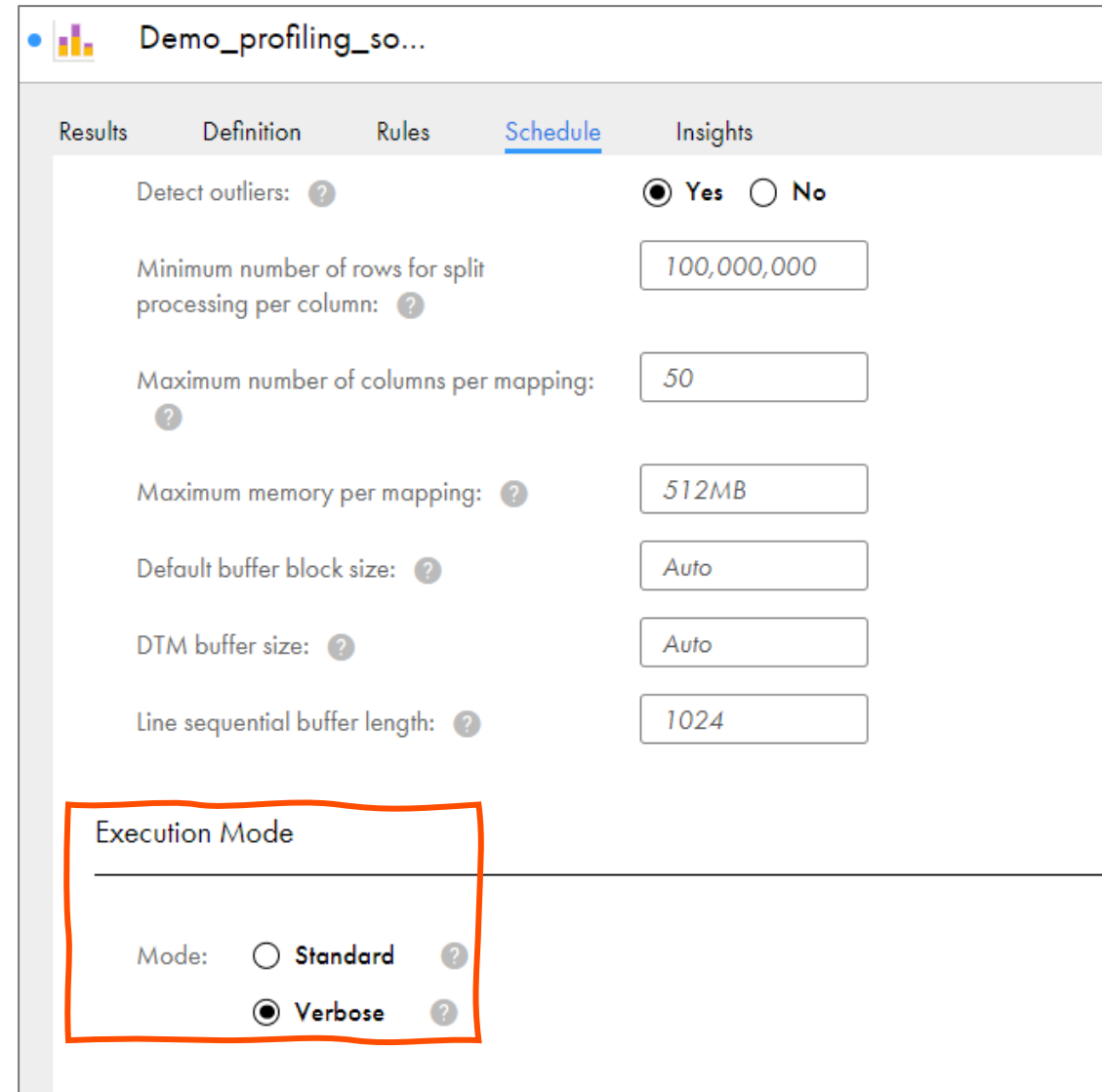


# Cloud Data Profiling

Verbose log option for profiling runs

- 
- 
- 
- 

Caution:



Demo\_profiling\_so...

Results Definition Rules Schedule Insights

Detect outliers: ? ☒ Yes ☐ No

Minimum number of rows for split processing per column: ? 100,000,000

Maximum number of columns per mapping: ? 50

Maximum memory per mapping: ? 512MB

Default buffer block size: ? Auto

DTM buffer size: ? Auto

Line sequential buffer length: ? 1024

Execution Mode

Mode: ☐ Standard ? ☒ Verbose ?

# Cloud Data Profiling

## Create Schedules from Profiling Definition

Profile103985

**i** The results from the latest profiling task are available. [Click here to update](#)

Results Definition Rules Schedule Insights

Runtime Environment

**i** Select the runtime environment that has access to the source object.

Runtime Environment:

Schedule Details

☐ Do not run this task on a schedule

☒ Run this task on a schedule:

Services:

Assets		Features					
Asset Type	Create	Read	Upd...	Delete	Run	Set Permi...	
Connection	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Elastic Configuration	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Folder	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Group	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
NTT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
OAuth Client	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Organization	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Privilege	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Project	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Role	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Schedule	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

New Schedule

Schedule Details

Name:

Description:

Schedule Options

Status:

Time Zone:

Frequency:

Interval:

Effective Days:

Time Range:

Repeat:

# Cloud Data Quality - Demo