

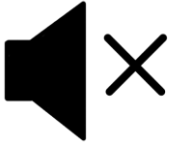
02 July, 2024

What's New in Informatica Cloud Data Governance and Catalog (CDGC), Cloud Data Marketplace (CDMP) and Cloud Data Quality (CDQ)? - July 2024 Release

- Amrish Thakkar, Sr Director, Product Management, R&D
- Kate Amory, Director, Product Management, R&D
- Ricardo Sobral, Sr Director, Product Management, R&D

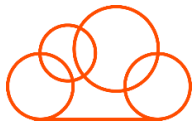
**Where data
& AI come to** 

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 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.

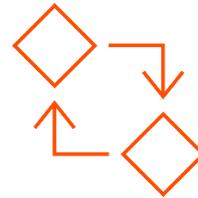
Feature Rich Success Portal



**Bootstrap trial and
POC Customers**



**Enriched Customer
Onboarding
experience**



**Product
Learning Paths
and Weekly
Expert Sessions**



**Informatica
Concierge**



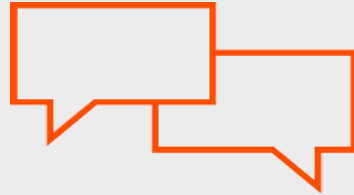
**Tailored training
and content
recommendations**

More Information



Success Portal

<https://success.informatica.com>



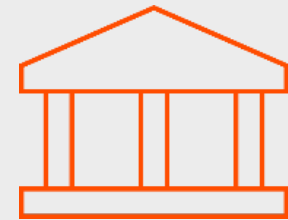
Communities & Support

<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 2024.07.M

- Inferred Lineage
- Lineage Enhancements
- Dashboard Improvements
- Rename of Catalog Sources
- Audit Report
- Data Discovery Enhancements
 - Glossary Confidence Score in UI
 - Glossary Prefix / Suffix extended for datasets
 - PAM
- Hierarchical representation for technical assets
- Data Quality Failure Notifications
- Usability Improvements
 - Search Improvements
 - Bulk Import changes
- Scanner Additions and enhancements

Inferred Lineage

Inferred Lineage

- Automatically detects dataflow between user provided source and target.
 - Dataset level
 - Data Element level
- Allows object filtering for more precise matching.
- Allows user curation of generated dataflows.
- Accepted dataflows are visible as lineage in CDGC application.
- With July, supported for Relational Sources only



Build lineage for dataflows happening through manual process, custom inhouse solutions, non-supported ETL, programming languages or data replication tools.

Benefits:

- **Efficiency** : By Avoiding time consuming manual lineage process using custom lineage.
- **Reduce risk** of user errors in manual custom lineage
- **End-To-End Visibility** and **enhanced data governance**.

Inferred Lineage

Metadata Command Center

Start typing search terms for suggestions.

INFA CDGC PM

Configure

Reference IDsWorkflowsIDMC MetadataLineage

Rule Based Linking

Name	Description	Sources
New Rule Based...	Lorem ipsum	EDW_Retail_Prod.HAWK1
QLabs Orders	All the orders from this lab	EDW_Retail_Prod.HAWK1
HR Reporting	All employee data for HR reporting	EDW_Retail_Prod.HAWK1
Customer Datasets	Top customers data	EDW_Retail_Prod.HAWK1

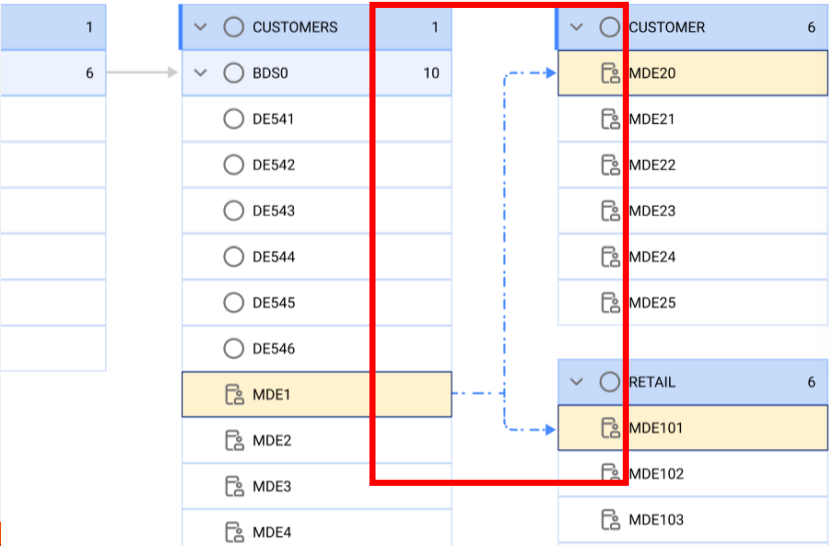
TOTAL LINKS2534ACCEPTED2534REJECTED0

Lineage Links (2534)

Find

Links Selected: (9)Accept AllReject All

Source	Target	Rule Type	Status	Actions
EDW_Retail_Prod.HAWK1	Snowflake_Retail.Contoso	Name Matching	Auto - Accepted	...
CUSTOMER_STG	CUSTOMER	Name Matching	Auto - Accepted	...
AddressLine1	AddressLine1	Name Matching	Auto - Accepted	...
AddressLine2	AddressLine2	Name Matching	Auto - Accepted	...
BirthDate	BirthDate	Name Matching	Auto - Accepted	...
COUNTRY	COUNTRY	Name Matching	Auto - Accepted	...
CompanyName	CompanyName	Name Matching	Auto - Accepted	...
CustomerKey	CustomerKey	Name Matching	Auto - Accepted	...
CustomerType	CustomerType	Name Matching	Auto - Accepted	...
FirstName	FirstName	Name Matching	Auto - Accepted	...



Lineage Improvements

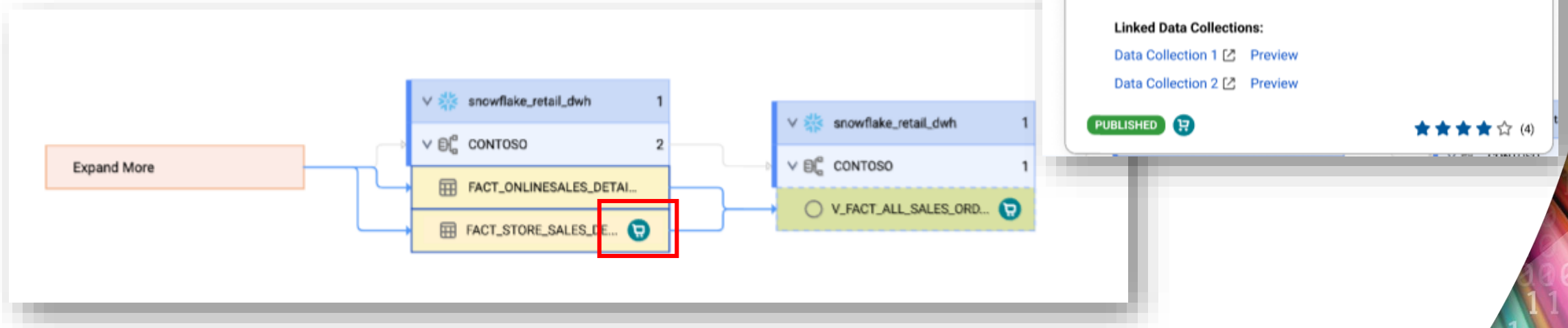
Data Marketplace Overlays

- Data Marketplace **Overlay** available as an option on both Business / Technical Lineage
- Allows user to quickly determine whether a specific asset upstream/downstream has been published to Data Marketplace

Overlays ⓘ

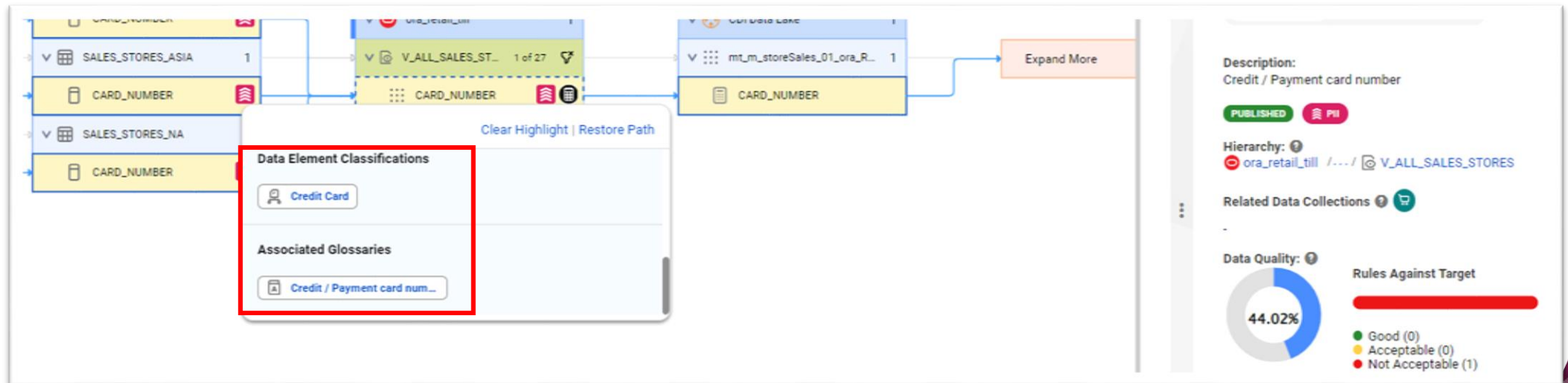
☐ Glossary	Data Quality	Others
☐ Business Term	☐ Aggregated	☐ Business Name
☐ Domain	☐ Accuracy	☒ Certification
☐ Metric	☐ Validity	☐ Calculation Expression
☐ Subdomain	☐ Completeness	☒ Data Marketplace
☐ Business Impact	☐ Consistency	
☐ Policy	☐ Uniqueness	
☐ Process	☐ Timeliness	

[Apply](#) [Cancel](#)



Other Lineage Improvements

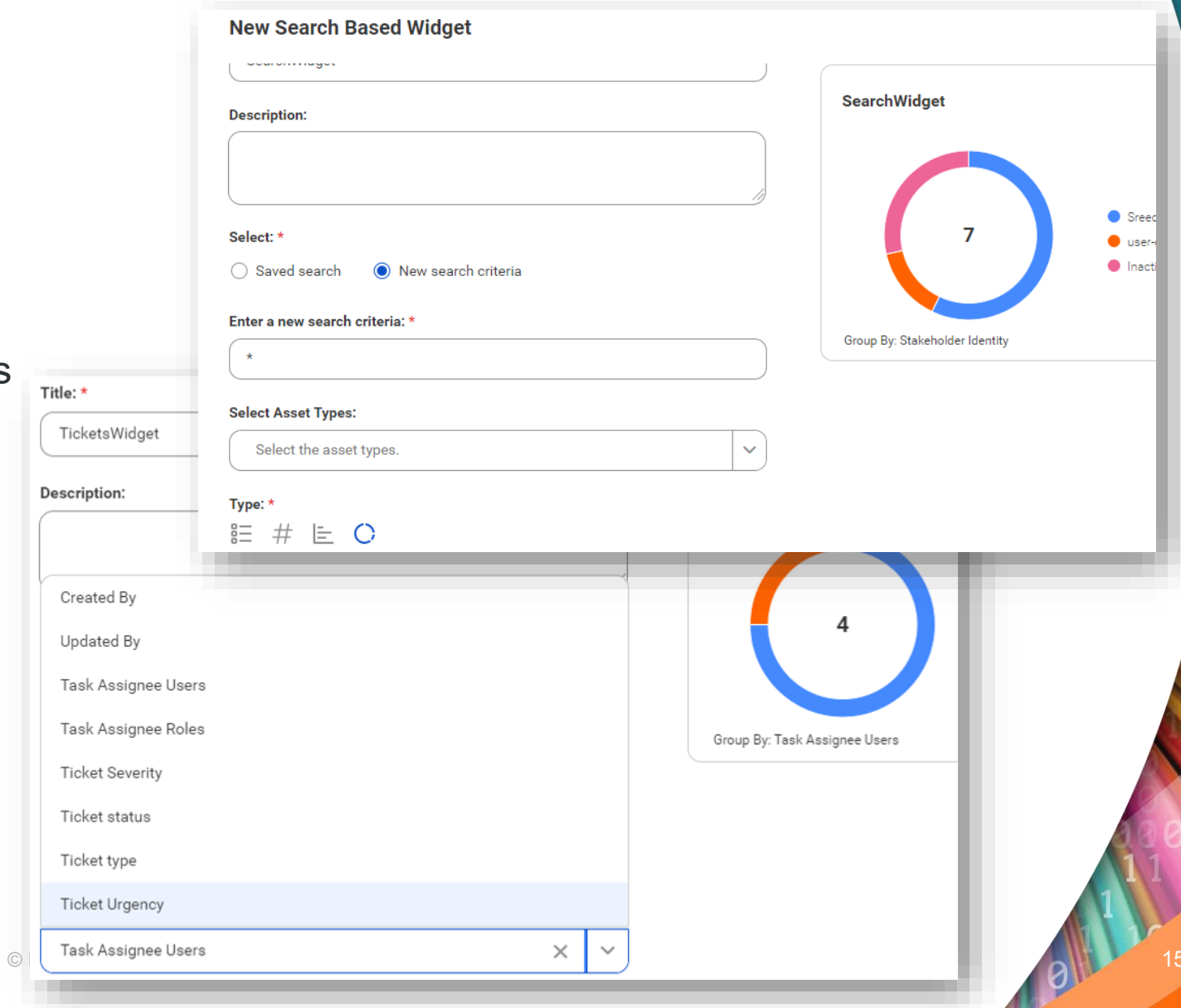
- Tooltip on hover over asset, Associated **Data Classification & Glossary** is displayed by default.



Dashboard Improvements

Dashboard Improvements

- Search Widget
 - Asset Type is no longer mandatory providing more flexibility to create dynamic dashboards.
 - Additional filters and Group by options
- Ticket Widget
 - Additional Group by options



New Search Based Widget

Search Widget

Description:

Select: *

☐ Saved search ☒ New search criteria

Enter a new search criteria: *

Title: *

TicketsWidget

Description:

Select Asset Types:

Select the asset types.

Type: *

☐ ☒ ☐ ☐

Created By

Updated By

Task Assignee Users

Task Assignee Roles

Ticket Severity

Ticket status

Ticket type

Ticket Urgency

Task Assignee Users

SearchWidget

Group By: Stakeholder Identity

7

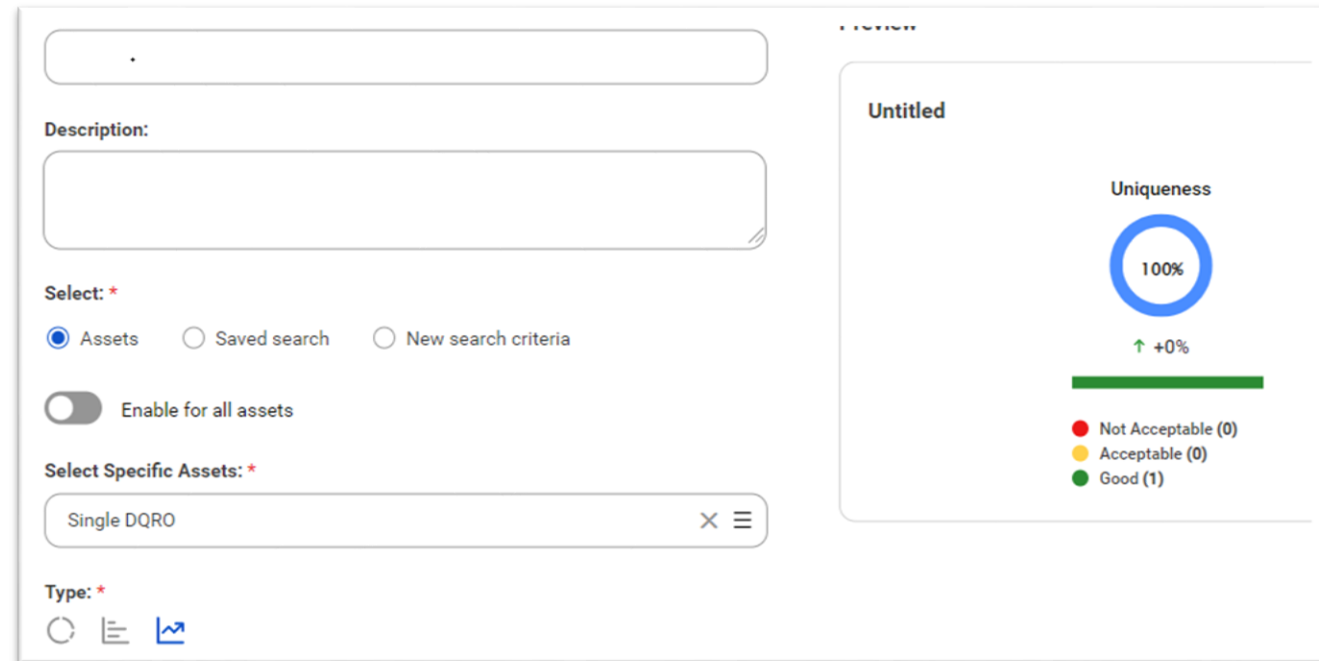
SearchWidget

Group By: Task Assignee Users

4

Dashboard Improvements

- Data Quality Widget
 - **Score Chart** and **Bar chart** is now supported for single asset such as Glossary, Dataset, Systems, Process, Policy, Table, Views, Files, Reports etc.
 - Saved Searches is added to DQ widget (only for Donut chart)



Glossary Association & Data Classification

Glossary Confidence Score

- The CDGC UI now includes a feature that displays the CLAIRE confidence score for glossaries.
- When you hover over a glossary from the data elements page, the confidence score of assigned will be shown
- Shown for Auto Accepted and recommended glossaries.
- This allows user to review and take action based on the confidence score.

The screenshot displays the CDGC UI interface. At the top, there are two sections: 'Glossaries' and 'Data Element Classifications'. Both sections show 'Accepted (1)' items. The 'Glossaries' section has a button labeled 'First Name' with a close icon. The 'Data Element Classifications' section also has a button labeled 'First Name' with a close icon. A tooltip is visible over the 'Item ID' button in the 'Glossaries' section, displaying the following information:

- Item ID
- Type: Business Term
- Confidence Score: 80%
- CLAIRE™ Recommendation
- Lifecycle: Published
- Description: Each item as an item number that uniquely identify.

Below these sections, there is a table titled 'Data Element Classifications' with columns: 'Type', 'Glossaries', and 'Classifications'. The table has three rows. The first row is a header row. The second row shows 'Column' under 'Type', 'Account Inf... +2' under 'Glossaries', and 'First Name' under 'Classifications'. The third row shows 'Column' under 'Type', 'Account Information (85.3%)', 'Access Data (73.5%)', and 'First Name (86.8%)' under 'Classifications'. A tooltip is visible over the 'Account Information (85.3%)' cell, displaying the following information:

- Account Information (85.3%)
- Access Data (73.5%)
- First Name (86.8%)



Glossary Association Scope

- In Glossary association configuration for a catalog source, you can now limit the scope of association to a specific hierarchy or hierarchies.
 - This feature is extended to dataset as well with this release
- Provides option to consider **entire glossary** or **subset of glossary**.
 - Allows you to select **any top-level glossaries**
 - Only the **business terms or metrics** in **selected hierarchies** will be considered for glossary association
- Glossary Association for Prepackaged application (SAP/Salesforce)
- Benefits :
 - More accurate associations
 - Better performance

Glossary Association Scope

☐ All Glossary Assets ? ☒ Top-level Glossary Assets ?

⚠ Removes all existing glossary associations that are not within the scope of the selected glossaries.

Select Top-level Glossary Assets

Discovery and Classification PAM

- Classification and Glossary Association has been enabled for the following sources:
 - SAP BW
 - HDFS
 - SharePoint
 - SFTP File System
 - Microsoft Fabric OneLake
 - Microsoft Fabric Lakehouse

Other Enhancements

Audit Report

- Allows user to generate Audit report based on
 - Time Period
 - Asset Types
 - Changed By
 - Type of change
 - Or Combination
- Generates an Excel file with the changes matching the criteria
- Works on a single Asset or across multiple Assets

The screenshot shows the 'Export Audit History' window in the Informatica Data Governance and Catalogue application. The interface includes a sidebar with navigation options like 'New', 'Home', 'Browse', and 'Export Audit Hist...'. The main panel has a title bar 'Export Audit History' with an 'Export' button. Below the title bar, a message states: 'Download the Audit History as a Microsoft Excel (.xlsx or .xls) file. If you export 10000 or more assets, the Microsoft Excel file is compressed as a .zip archive.' The form contains several fields: 'Export by:' with radio buttons for 'Asset Type' and 'Assets' (selected); 'Assets:' with a text input field and a menu icon; 'Date Range:' with a dropdown menu showing 'Previous Month'; 'Operation:' with a dropdown menu; 'Changed By:' with a dropdown menu; and 'Export File Name:' with a text input field containing 'Audit_Export_2024_05_29_121316'. A 'Collapse' button is visible at the bottom left of the main panel.

Search improvements

- "Current User" variable available as part of the Search query and filter (May 2024 release)
 - This allows one to create and save searches with Current User and that can be leveraged as part of the default Dashboards made available by Admin for the user based on their Role
- Dynamic Filters: (May 2024 release)
 - When search result returns single asset type – additional filters are available by default (without clicking asset type)
 - Catalog source name filter is available by default
- Search improvements for partial string, begins with and ends with search

The screenshot displays the 'assets' search interface. The main panel shows a 'Filters' section with a search bar and various asset type filters: Column (2.6M), View Column (675.4K), Table (150.7K), Data element (57.9K), Field (57.8K), DataSource Field (32.7K), Foreign Key (22.5K), Function Module Table Field (21.3K), Calculation (18.2K), and Table Field (17.7K). Below these is a 'Catalog Source Type (61)' filter with a 'Select' dropdown. At the bottom, there are 'Additional Filters' including 'Catalog Source Name (1000)'. A sidebar on the right shows a 'Created By (5)' filter with 'Current User (sdhanuka2)' selected, and 'Updated By (2)' and 'Stakeholders (4)' filters, also with 'Current User (sdhanuka2)' selected. The sidebar includes 'Clear' buttons for each filter group and radio buttons for 'Select' and 'No stakeholders'.

Hierarchical representation for technical assets (May 2024)

- Data Sets that contains hierarchical information will now be represented as proper tree structure for data elements.
- This is available in contains tab as well.
- Example File Types: XML, XSD, Parquet, Avro.

FileSystem Demo / ... / IICS_Demo /

(Add Business Name)

XML FILE | emp.xsd

Overview

Contains

Lineage

Relationships

Data Quality

Data Obs

Data Elements (1)

Name	Type	Glossaries	C
▼ root (1)	Element		
▼ Emp_Details (3)	Element		
> address (8)	Element		
EmpAttribute	Attribute		
▼ employee (3)	Element		
Dependents	Element		
Employeeid	Element		
▼ Name (2)	Element		
Firstname	Element		
Lastname	Element		

Bulk Import enhancements

- Automatic Assignment of Business Names and Descriptions to Technical Data Elements
 - Users can now automatically assign business names and descriptions to technical data elements during glossary association.
 - Users can choose to propagate name and description from accepted or non-rejected glossaries; or manually add them to a technical data element using the technical asset export sheet.
 - Users can include glossary reference IDs in Business Name, Glossary: Accepted, or Glossary: Rejected columns to differentiate among duplicate glossary assets in the catalog.
- Specify Process duration in Weeks, Days or Hours
 - When users opt for the bulk import operation to create a Process asset, they can now specify the duration in weeks, days or hours.

Autocataloging of CAI Design Time Metadata

- This release will allow users to Autocatalog Design Time Metadata from Cloud Application Integration (CAI)
 - Scope of the Design Time objects cataloged is:
 - App Connections
 - Guides
 - Process Objects
 - Processes
 - Service Connectors
 - It will include following content:
 - Name, description, including creator, modifier, last updated date, creation date, dependencies

Other enhancements

- **Rename Catalog Source**

- User can now rename catalog sources in MCC.
- This capability allows them to update source names at any point during the data governance process, enhancing flexibility and accuracy in managing data sources.
- All catalog source capabilities will be available after next metadata extraction job execution.

- **Data Quality Notifications**

- DQ failure notifications for all rules available with July release
- When threshold/target are breached, stakeholders will be notified

- **QuickLook improvements**

- Support for German in QuickLook. Other languages to follow
- Support for Microsoft Office 365 and PDF Documents

Scanners Updates 2024.07

New Scanners

- Microsoft Azure Analysis Services (GA)
- SAS Base Libraries (IPS Accelerator)
- Added support for programmable objects (i.e., stored procedures):
 - Greenplum
 - Netezza
 - PostgreSQL
- dbt (Preview -> GA)
- IBM Netezza (base inventory) - May '24
- Greenplum (base inventory) - May '24

Scanners Updates 2024.07

Scanner Enhancements

- **Case-sensitivity support for relational sources**
- **Salesforce**
 - Support for Lookup relationships
 - Added relationships between List Views and Source Objects
- **Databricks enhancements**
 - Databricks Delta Life Tables support (SQL version only)
 - Support for RunJob Task in workflows
- **Informatica CDI scanner enhancements**
 - Support for AccessPolicy component
 - Support for SnowflakeCloudDataWarehouseV2 transformation
- **Lake scanners enhancements**
 - Support for S3 compatible storage MinIO
 - Enabled partitioned file configuration options for Azure Blob
 - Improved JSON metadata extraction to properly represent hierarchical structure in CDGC
- **Oracle enhancements**
 - Additional package filtering option
- **Tableau enhancements:**
 - Populating additional attribute with DataType on Calculation objects
 - Added support for PostgreSQL as a data source
 - Added support for referenced stored procedures and custom SQLs
- **Support for Azure Analysis Services and SQL Server Analysis Services compatibility level 1600**
- **Added connection-aware scan capability for SQL Server Analysis Services**
- **Microsoft Azure Data Factory enhancements**
 - Support for AppendVariable activity
- **Microsoft PowerBI enhancements**
 - Enhanced metadata model - extending Dataset hierarchy with Table level (Dataset->Table->Column)

Cloud Data Marketplace

New Asset Types from CDGC

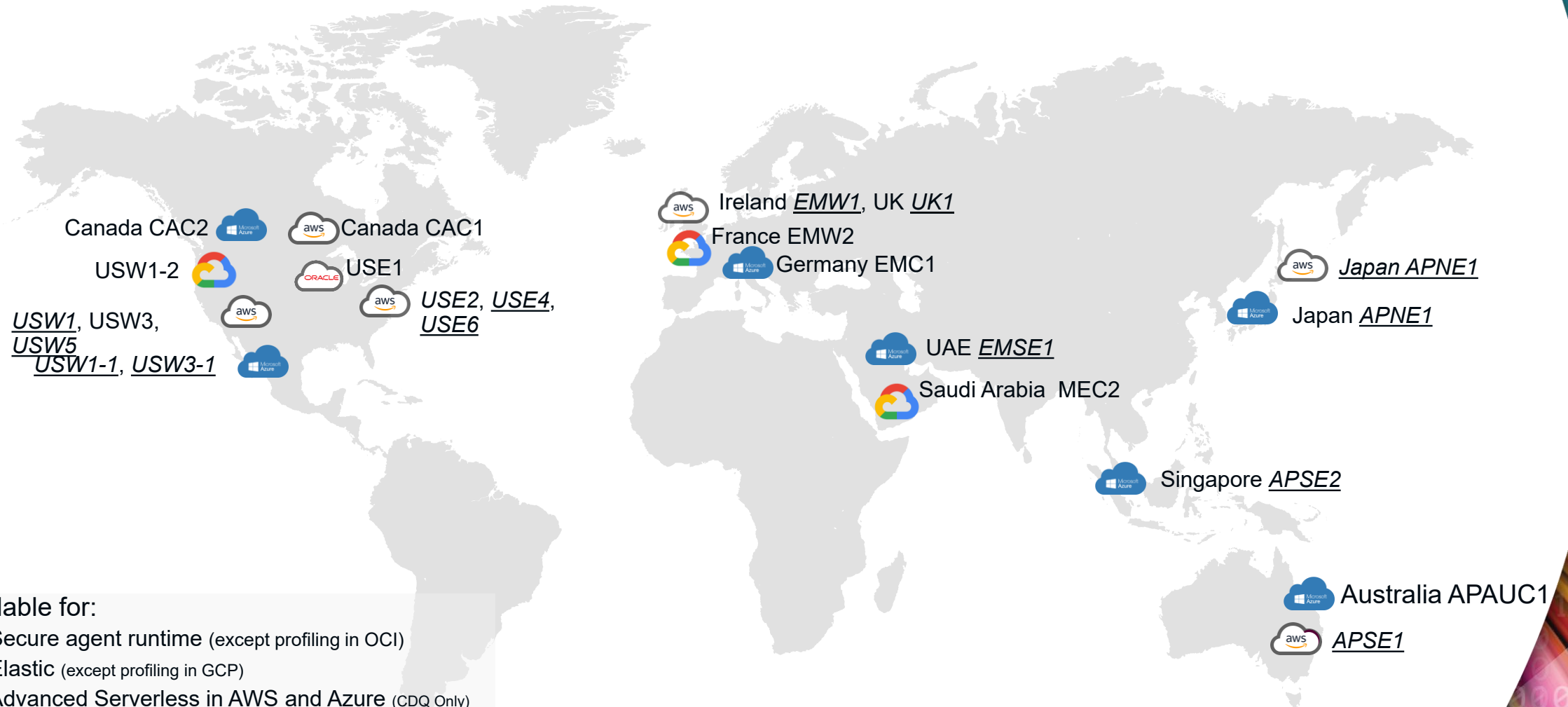
July 2023 Release Updates

- New Asset Types Available in Marketplace:
 - Qlik
 - Cognos
 - Unstructured Data (excel, word, etc.)
 - MicroStrategy
 - SAP Business Objects

Data Governance and Catalog Asset Types				ExternalSourceDQMetricsDC1171921...
NAME ↕	COMPATIBILITY ↕	STATUS ↕		DESCRIPTION
Model Query	Available	☐	Inactive	IBM Cognos Model Query
Query	Available	☐	Inactive	IBM Cognos Query
View	Available	☐	Inactive	IBM Cognos View
Delimited File	Available	☐	Inactive	IBM Cognos Delimited File
Data Set	Available	☐	Inactive	IBM Cognos Dataset
Hierarchical File	Available	☐	Inactive	A big data file is a structured set of related hierarchical r...
Unstructured File	Available	☐	Inactive	Unstructured file class like PDF, Microsoft Excel, and MI...
Logical Table	Available	☐	Inactive	MicroStrategy Logical Table
HTML Document Report	Available	☐	Inactive	MicroStrategy HTML Document Report
Report	Available	☐	Inactive	MicroStrategy Report
Intelligent Cube Report	Available	☐	Inactive	MicroStrategy Intelligent Cube Report
Logical Query Table	Available	☐	Inactive	MicroStrategy Logical Query Table
Data Set	Available	☐	Inactive	MicroStrategy Dataset
Warehouse Table	Available	☐	Inactive	MicroStrategy Warehouse Table
Document Report	Available	☐	Inactive	MicroStrategy Document Report
16 - 30 of 61 Object types				Rows: 15

Cloud Data Quality

Availability of Cloud Data Quality



Available for:

- Secure agent runtime (except profiling in OCI)
- Elastic (except profiling in GCP)
- Advanced Serverless in AWS and Azure (CDQ Only)
- Elastic Advanced Serverless (except profiling)
- Private link support

Cloud Data Quality

What's New in Cloud Data Quality July 2024?



Data Quality

- ☐ Data Observability
 - ☐ Performance improvement
 - ☐ Increase limits to support more than 10,000 data elements
- ☐ IPU consumption
 - ☐ Asset level consumption
- ☐ Rule Specification:
 - ☐ Support for external java packages using java transformation



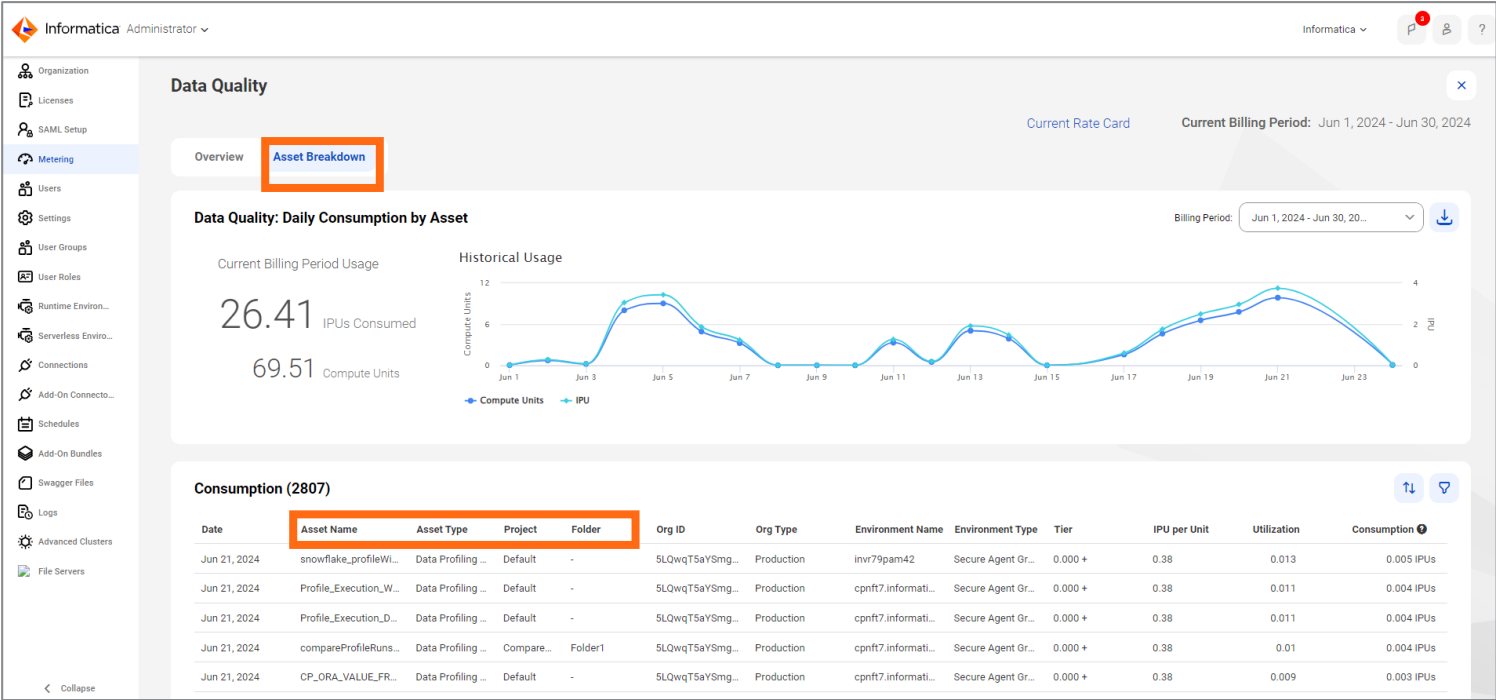
Data Profiling

- ☐ Change of behavior:
 - ☐ Statistics as percentage
 - ☐ Scorecard threshold
- ☐ Support custom queries in source mapplet
- ☐ Extended PAM for Source mapplet
 - ☐ Athena,
 - ☐ Microsoft Fabric OneLake
 - ☐ JDBC_V2
- ☐ Extend Profiling APIs to update advanced options

Chargeback meter for Cloud Data Qualit and Profiling

Administration -> Metering

- New “Asset Breakdown” tab
- Consumption show breakdown by “Project”, “Folder” and “Asset”
- Available for Native and Advanced execution



Chargeback meter for CDQ and CP

Detailed report download

- Consumption show breakdown by “Project”, “Folder” and “Asset”

HistoricalTableData_5LQwqT5aYSmgMPQGL04awv_cPnIT3XYAZiLqoXtd0KC_2024-06-24T19_03_54Z...																
File Home Insert Page Layout Formulas Data Review View Automate Help Acrobat																
B594 X ✓ fx CompareColumnsTest_SQL_40383																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Date	Asset Name	Asset Type	Project	Folder	Org ID	Org Type	Environment N	Environment Typ	Tier	IPU Per Unit	Usage	Consumption (IPU)			
6	2024-06-21T00:00:00	compareProfileRuns_Redshif	Data Profiling Task	CompareProfileRunsT	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.009825	0.0037335			
33	2024-06-21T00:00:00	compareProfileRuns_Redshif	Data Profiling Task	CompareProfileRunsT	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.00984694	0.003741839			
44	2024-06-21T00:00:00	compareProfileRuns_Snowfla	Data Profiling Task	CompareProfileRunsT	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.00989278	0.003759256			
116	2024-06-21T00:00:00	compareProfileRuns_Snowfla	Data Profiling Task	CompareProfileRunsT	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.00860889	0.003271378			
192	2024-06-20T00:00:00	TopNProfileRuns_Profile_894	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.00840278	0.003193056			
266	2024-06-20T00:00:00	SortOrderTopNProfileTaks_81	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.01264889	0.004806578			
323	2024-06-20T00:00:00	Profile_LastRunKey2	Data Profiling Task	ProjectName_1	FolderName_1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.02623389	0.009968878			
351	2024-06-20T00:00:00	compareProfileRunsTestWith	Data Profiling Task	CompareProfileRunsT	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.0167825	0.00637735			
369	2024-06-20T00:00:00	SortOrderTopNProfileRuns_2	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.01399028	0.005316306			
379	2024-06-20T00:00:00	Default_TopNProfileRuns_18	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.01396833	0.005307967			
467	2024-06-20T00:00:00	Default_TopNProfileTasks_88	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.01393306	0.005294561			
484	2024-06-20T00:00:00	compareProfileRunsTest_612	Data Profiling Task	CompareProfileRunsT	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.03633028	0.013805506			
550	2024-06-20T00:00:00	TopNProfileTasks_Profile_51	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.02796694	0.010627439			
594	2024-06-20T00:00:00	CompareColumnsTest_SQL	Data Profiling Task	CompareColumns_SC	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.00620028	0.002356106			
642	2024-06-20T00:00:00	CompareProfRuns4000Prec	Data Profiling Task	CompareColumns	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.01817	0.0069046			
649	2024-06-20T00:00:00	ProfilePermTest_49291	Data Profiling Task	PERMISSIONTEST	PERMTEST8	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.00823861	0.003130672			
667	2024-06-20T00:00:00	CompareColumnsTest_8382	Data Profiling Task	CompareColumnsPut	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.00964528	0.003665206			
842	2024-06-19T00:00:00	TopNProfileTasks_Profile_75	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.02791944	0.010609389			
845	2024-06-19T00:00:00	TopNProfileTasks_Profile_83	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.05769944	0.021925789			
906	2024-06-19T00:00:00	SortOrderTopNProfileRuns_7	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.01613639	0.006131828			
945	2024-06-19T00:00:00	SortOrderTopNProfileRuns_1	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.012575	0.0047785			
1003	2024-06-18T00:00:00	TopNProfileTasks_Profile_70	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.05100389	0.019381478			
1028	2024-06-18T00:00:00	TopNProfileTasks_Profile_11	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.04546861	0.017278072			
1090	2024-06-18T00:00:00	SortOrderTopNProfileRuns_7	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.01606278	0.006103856			
1102	2024-06-18T00:00:00	SortOrderTopNProfileRuns_6	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.01625972	0.006178694			
1115	2024-06-18T00:00:00	CompareColumnsTest_7776	Data Profiling Task	CompareColumns	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.014745	0.0056031			
1245	2024-06-17T00:00:00	Profile	Data Profiling Task	_zdt		5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.00231778	0.000880756			
1265	2024-06-17T00:00:00	validateRuleCreatedByInsight	Data Profiling Task	_Insights_UI_Autom	_Insights40896	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.10509611	0.039936522			
1326	2024-06-14T00:00:00	SortOrderTopNProfileRuns_9	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	dqcagent09.infor	Secure Agent Group	0.000 +	0.38	0.02168139	0.008238928			
1428	2024-06-14T00:00:00	CompareColumnsTest_5179	Data Profiling Task	CompareColumns	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	dqcagent09.infor	Secure Agent Group	0.000 +	0.38	0.02158694	0.008203039			
1460	2024-06-14T00:00:00	TopNProfileTasks_Profile_84	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	dqcagent09.infor	Secure Agent Group	0.000 +	0.38	0.05820778	0.022118956			
1629	2024-06-11T00:00:00	TopNProfileTasks_Profile_19	Data Profiling Task	TopNProfileTasks_Pro	TopNProfileTasks_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.04413944	0.016772989			
1680	2024-06-11T00:00:00	SortOrderTopNProfileRuns_9	Data Profiling Task	TopNProfileRuns_Pro	TopNProfileRuns_f	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.0172925	0.00657115			
1716	2024-06-11T00:00:00	CompareColumnsTest_3653	Data Profiling Task	CompareColumns	Folder1	5LQwqT5aYSmgMPQGL04awv	Production	invr79pam42	Secure Agent Group	0.000 +	0.38	0.01464278	0.005564256			
1808	2024-06-07T00:00:00	ora_10m_81col	Data Profiling Task	_Volume_Tests	1_Ors_linux	5LQwqT5aYSmgMPQGL04awv	Production	cpnft7.informatic	Secure Agent Group	0.000 +	0.38	0.22819194	0.086712939			

Support for external java packages using Java transformation

- **Use case:**

- Data Quality check requires a passive java transformation and external Java libraries

- **Solution:**

- Rule Specification supports maplet with Java Transformation that uses external packages
- Run test and mapping of the rule with that maplet executes successfully.

- **Prerequisites**

- Jar files of the external packages should be placed in secure agent machine and the path of the same should be specified in the Design time Classpath field in the Java transformation for compilation.
- Data Integration Server should be set with JVMClasspath
- For more information, check <https://docs.informatica.com/data-governance-and-quality-cloud/data-quality/current-version/rule-specification-assets/preface.html#q=external>

Profiling statistics shown as percentage

- **Use case:**
 - Increase precision of profiling results statistics presented as percentage
- **Existing Behaviour**
 - 99.99% presented as 99.99%, **anything greater than 99.99% presented as 100%**, for ALL percentage statistics
 - **Anything lower than 0.001% presented as 0%**, for ALL percentage statistics
- **New Behaviour (as of July 2024)**
 - If at least one record failed to pass DQ tests, the score will be 99.99 (even if it is 1 record in 200M or 99.99999999%)
 - If at least one record pass, you will get a score 0.01 (even if 1 record pass out of 200M Or 0.00000000001%)
 - Show 100% only when ALL rows passed
 - Show 0% only when ALL rows failed

Profiling statistics shown as percentage

CP_ValidatePercentageRoundOff_89024

Profile run 1 of 1 | 2 of 2 Columns | 2 of 2 Rules | 10000002 Rows (All rows) | Jun 24, 2024, 8:51 PM

Save

Run

...

X

Results

Definition

Rules

Schedule

Insights

Column: PrimaryRuleSet (RuleCheck_2) | Input(s): CMTITL

Back to Summary

2 distinct values (2 non-unique values, 0 unique values)

Sort By: Frequency Descending

#	Value	Frequency	Percentage	Length
1	INVALID	10000000	99.99%	7
2	VALID	2	0.01%	5

Details

Trend

Type	Trend	% Change
Null	.	
Distinct	.	
Non-distinct	.	

Data Types (2)

string(100)	Documented Data Type ?
String(7)	100%

 Informatica

© Informatica. Proprietary and Confidential.

39

Enhancements to thresholds in metrics/rule occurrences

Use case:

- 1 Only “GOOD” and “NOT ACCEPTABLE” scores AND Score is “GOOD” ONLY when all rows are Valid
- 2 Score is “GOOD” ONLY when all rows are Valid

Scores that fall into a range between two bands are included in the higher band.

- in example 1, the score 100, it will be “Good” and 0 will be “Acceptable”
- in example 2, the score 100 will be “Good”, 99.99 will be “Acceptable” and 0 will be “Not Acceptable”

1

100.0	to	100	%	Good
0.00	to	100.00	%	Acceptable
0	to	0.0	%	Not Acceptable

In this Example, if score < 100, score is “ACCEPTABLE”

2

100.0	to	100	%	Good
99.99	to	100.00	%	Acceptable
0	to	99.99	%	Not Acceptable

In this Example, if score < 100, score is “NOT ACCEPTABLE”

Cloud Data Profiling

What sources can you profile? (starting July 2024)

- Microsoft SQL and Azure Database with Active Directory Authentication
 - Microsoft Azure Synapse SQL (using Native driver)
 - Azure ADLS gen 2 and Azure Blob storage type
 - Oracle Database and Oracle Autonomous DB
 - Microsoft SQL and Azure Database
 - Amazon Redshift (using native V2 driver)
 - Amazon Athena
 - Snowflake (using native V2 driver)
 - Google Big Query
 - Delimited File on: File system | Amazon S3 | Azure ADLS Gen2 | Google Cloud Storage | **Microsoft Fabric OneLake**
 - ODBC
 - JDBC
 - SAP ECC, SAP S/4 Hana and SAP BW
 - Databricks Delta tables using IDMC connector
 - Secure agent
 - Advanced / Elastic
 - Salesforce
 - Salesforce Sales Cloud, Salesforce Service Cloud
 - Salesforce Verticals: Salesforce Financial Cloud, Salesforce Health Cloud
 - Applications on Force.com
 - Source mappers
 - Amazon Redshift, Amazon S3, Azure Data Lake Store, Databrick, (ODBC), Flat File, Google BigQuery, Google Cloud Storage, Microsoft Azure, Synapse, Microsoft SQL Server, Oracle, Oracle (ODBC), Oracle Object Source, Salesforce, SAP BW, SAP Table, Snowflake. **added Athena, Microsoft Fabric OneLake**
 - PARQUET, AVRO and JSON files on:
 - Amazon S3 - [Requires Advanced / Elastic]
 - Azure ADLS Gen2 - [Requires Advanced / Elastic]
 - **PARQUET, AVRO on Microsoft Fabric OneLake** - [Requires Advanced / Elastic]
- Deprecated**
- Azure Synapse, Snowflake, Redshift and Databricks using ODBC (still works, not recommended)
 - Databricks Delta tables using ODBC driver

Profiling of Source Mapplets

Support custom queries in source mapplet

- Use Case:
 - Profile of multiple tables based on a SQL query
- Solution:
 - Profiling of Source Mapplet where user can copy/paste a working SQL.
 - SQL can contain joins and any statement supported by the source

The screenshot displays the Informatica Source Mapplet configuration interface and its profiling results. The top section, titled 'Properties | Source', shows the configuration for a 'Query' source type connected to 'SNOWFLAKE_NATIVE_SRC'. The 'Query' field contains a complex SQL statement for batch processing. The bottom section, titled 'LDOProfile_Snowflake_Complex_Custom_Query_10490', shows the profiling results for the query. It includes a table with columns for 'Columns', 'Value Distribution', '% Null', '# Null', '% Distinct', '# Distinct', '% Non-di...', '# Non-dis...', and '# Patter...'. The table lists three columns: 'ORDER_NUMBER_OUT', 'BATCH_ID_OUT', and 'QUANTITY_OUT', each with corresponding statistical data.

Properties | Source

Details

Connection:* SNOWFLAKE_NATIVE_SRC (Snowfla... [New Parameter...](#)

Source Type:* Query

Validation Type: ☒ Mapping ☐ Mapping (Advanced Mode)

Query: [Edit Query...](#)

```
WITH batch_cte AS
(SELECT *, ROW_NUMBER() OVER (ORDER BY batch_id) AS rn
FROM (
  WITH RECURSIVE batch_split AS
  (SELECT batch_id, 1 AS quantity FROM INFADQ_DB.DQ_SCHEMA.batch
```

LDOProfile_Snowflake_Complex_Custom_Query_10490

Profile run 1 of 1 | 3 of 3 Columns | 0 of 0 Rules | 8 Rows (All rows) | Jun 7, 2024, 2:12 PM

Results Definition Rules Schedule Insights

View: Columns and Rules with: All Statistics

Columns	Value Distribution	% Null	# Null	% Distinct	# Distinct	% Non-di...	# Non-dis...	# Patter...
Columns								
ORDER_NUMBER_OUT		0%	0	75%	6	25%	2	1
BATCH_ID_OUT		25%	2	37.5%	3	37.5%	3	2
QUANTITY_OUT		25%	2	50%	4	25%	2	2

Extend Profiling APIs to update advanced options

Update the advanced options via API

- **Use case:**

- Profiling of a S3 files in a folder in source in /DEVPATH/, but after exporting/importing to another environment, the folder should be /PRODPATH/
- Ability to create/update the advanced options for the sources that have advanced options via API.

- **Solution:**

- create/update “Advance Option” properties for source via API
- POST CALL: [https://qa-pod1.rel.infaqa.com/frs/v1/Folders\('{folderID}'\)/Documents](https://qa-pod1.rel.infaqa.com/frs/v1/Folders('{folderID}')/Documents)
- PATCH CALL: [https://qa-pod1.rel.infaqa.com/frs/api/v1/Documents\({profileID}\)](https://qa-pod1.rel.infaqa.com/frs/api/v1/Documents({profileID}))
- Portion of payload where advance property details are given in payload.

```
advancedOptions\":{\"Allow Wildcard Characters\": \"false\", \"Compression Format\": \"None\", \"Download Part Size\": \"5242880\", \"Enable Recursive  
Read\": \"false\", \"Encryption Type\": \"None\", \"File Name\": \"\", \"Folder Path\": \"cpqabucket\\\\testdata\\\\n\", \"Hadoop Performance Tuning Options\": \"\", \"Incremental File  
Load\": \"false\", \"Multipart Download Threshold\": \"10485760\", \"Source Type\": \"File\", \"Staging Directory\": \"\", \"Temporary Credential Duration\": \"900\"}
```

Q&A

Where data & AI come to



Informatica