Informatica

02 April, 2024

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

- Amrish Thakkar, Senior Director, Product Management

- Kate Amory, Principal Product Manager, Product Management

- Markus Loechel, Senior Manager, Product Management

- Sean O'Shea, Senior Manager, Product Management

Where data
& AI come to LIFE

Housekeeping Tips
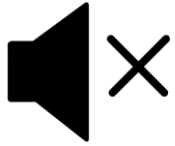
- Today's Webinar is scheduled for 1 hour

- 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.
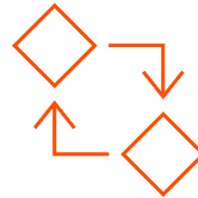
Feature Rich Success Portal
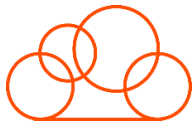
Bootstrap trial and POC Customers

Enriched Customer Onboarding experience

Product Learning Paths and Weekly Expert Sessions
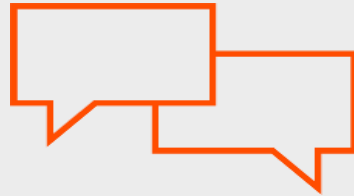
Informatica Concierge

Tailored training and content recommendations
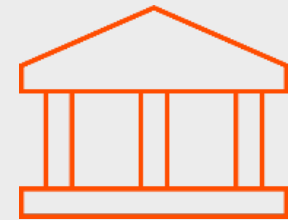
More Information

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.

Informatica

Cloud Data Governance and Catalog

Where data
& AI come to LIFE

Features in CDGC 2024.04.M

- **Model and Relationship customization**

  - Including DQ Dimension, Data Classification Sensitivity

- **Updated Audit History**

- **Manual Data Element**

- **Obsolete Lifecycle for Business Assets**

- **Saved Searches**

- **CDGC to CDMP Integration**

- **Data Discovery Enhancements**

  - Limit Glossary Association to specific Glossaries hierarchy for a Catalog Source

  - Improved Discovery for Packaged applications (SAP, SFDC)

  - Increase Max Length of Classification Rules

  - Rename Data Element Classification

- **Scheduling Catalog Source Capability**

- **Lineage Enhancements**

  - Save Lineage Preferences

- **Usability Improvements**

  - Workflow requester

  - DQ Rule Occurrence for non-supported DE

  - CDQ Rule reference as part of Bulk Import

- **Language Support: Dutch for CDGC**

- **Azure Private Link support**

- **Scanner Additions and enhancements**

- **Profiling**

Informatica®

Model & Relationship Customization

Model Customisation

Attributes and Relationships now fully customisable

- Most OOTB attributes will now be customizable as custom attributes

- Customized in MCC as per screenshots

- Values can be added, renamed, deleted

- Among other things, you can customize

  - DQ Dimensions (max 12 values permitted)

  - Data Element sensitivity values (with new icon set)

Relationship Types

- Extending range of relationship labels between existing assets

  - (doesn't allow creation of relationships between new asset pairs)

- All relationships between same asset types shown in one page

- Add / Remove

- Controlled – won't let users remove relationship types in use

Asset Audit History

Audit History

Reworked back end to provide reliable audit trail of asset activity

- All CDGC objects now have History tab

    - Captures all asset activity

- Filters for searching for specific actions

- Additions/deletions of child assets rolled up to the parent history

- Tickets also have updated look

Manual Data Element

Manual Data Elements

New Business Asset for documenting assets unsuitable for technical catalog

- Business Data Sets only
  - Mix of technical and manual data elements in same business data set

- Allows business users to describe data they consume data for different use cases, e.g. where technical data is changed.

- Members of single business data set

- Relationships
  - Lineage /Glossary / Data Element Classification

Lifecycle: Obsolete

New Lifecycle Value - Obsolete

Complete asset lifecycle journey, pre-deletion

- Allows upload of assets already in an Obsolete state – denotes they still have historical importance (e.g. future regulatory enquiry), not yet ready for deletion.

- May be particularly useful for customers migrating from Axon

- Implemented as fixed values, will be customisable in a later release

Saved Searches

Saved Searches

- Users can save search queries with applied filters.

- Reusable Search definitions

 - Eliminates the need for users to repeatedly enter complex queries

- Users can add custom names to saved searches for easy retrieval.

- Saved Searches are private to the user who created them.

My Saved Searches

- Centralized storage for all saved searches

- Easy access and management of saved searches

- Mark search queries as favorites
  and sort based on favorites

Dashboard Integration

- Ability to create dashboard widgets for quick access

- Customizable dashboard layout for personalized experience

CDGC to CDMP integration

Integration with Cloud Data Marketplace

- When viewing an asset that is linked to a Marketplace Data Product, you will see:
  - The teal badge indicating Data Marketplace presence
  - Links to the Data Products the asset resides within
- You will also see these in the preview panel from search results
- Clicking the 'info' tooltip next to a Data Product will open a preview showing key information about that Data Product

Glossary Association & Data Classification

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.

- Provides option to consider **entire glossary** or **subset of glossary**.

  - Allows you to select top-level glossaries

  - Only the business terms or metrics in selected hierarchies will be considered for glossary association

- Benefits :

  - More accurate associations

  - Better performance

Informatica®

Improve classification for packaged application (SAP, SFDC)

- In packaged applications such as SAP and Salesforce metadata objects names do not provide business context.

- To improve the overall classification of such sources, business names are extracted as part of metadata extraction and in generated classification instead of object names business names are considered for CLAIRE classification generation.

- Benefits :

    - This avoids invalid classification names.

    - Classification tags are more accurate

Increased max length of classification rules

- In Data Element Classifications length of the user classification rule is increased from 512 characters to 5000 characters.

- User can now add more conditions to fine tune classification rules.

- User can also validate rule length in MCC rule creation UI.

Informatica

Rename Data Classifications

- User can now rename data element classifications and data entity classifications rules in MCC.

- Allows user to modify rule names for OOTB rules or user defined rules.

- Allows user to use rule names in local languages.

- Avoids unnecessary rule cloning and re-curation activity.

Other Enhancements

Schedule Catalog Source Capabilities

- In Catalog Source Configuration, user can now **add multiple schedules** in same catalog source configuration.

- User can schedule each capability or selected capabilities at different execution frequency or frequencies.

- Allows governance administrator to better manage capability dependencies.

- With the ability to schedule less source-intensive activities such as metadata extraction more frequently, administrators can ensure that essential metadata is always up-to-date without consuming excessive resources.

- Administrators can plan source-intensive jobs like profiling or data quality checks at different times based on factors such as data availability, system load, or business requirements. This allows for better optimization of system resources and ensures that these tasks are executed when they will have the least impact on other processes.

Save Lineage Preference

- Users can tailor their lineage viewing experience by saving preferences such as overlays, loops, or display link options.

- This customization ensures that users can focus on the specific aspects of lineage that are most relevant to their tasks or interests.

- Preference can be saved for technical lineage, business lineage or data element level lineages.

- User can reset saved preference to default at any time.

- User choice are preserved for lineage of all the assets in the catalog.

  - User will see same saved overlay options by default for any asset.

Informatica

Data Quality Enhancement

- Allow association of Data Element with Rule Occurrence for Non DQ enabled sources

    - This allows for users to upload scores manually for the Catalog Source where DQ is not enabled

- Score propagation job to execute whenever a new score is pushed into Rule Occurrence via Bulk Import to give you an updated view quickly

- Ability to upload the Technical Rule Ref during Rule Occurrence bulk upload if autocatalog is switched on in MCC.

- Ability to Override Rule Occurrence schedule with the MCC Schedule defined for DQ Capability

Workflow Enhancement

- Ability to add Ticket creator / Ticket requestor as the task owner in workflow configuration

- Involvement of the requestor creates a feedback loop within the approval process enabling them to address any issues or questions raised by the approvers promptly.

Profiling Updates

- **PAM Additions**
    - Hive
    - PostGreSQL
    - Microsoft Fabric Warehouse
- **Random Sampling for AWS Redshift, Oracle and MS SQL Server**

Informatica®

Scanners Updates

New Scanners

- Microsoft Fabric OneLake (Preview)

- Microsoft Fabric Lakehouse (Preview)

- DBT (Preview)

- SAP S/4HANA Cloud (Preview)

- IBM DataStage

Informatica®

Scanners Updates 2024.04
Scanner Enhancements

- **Microsoft SQL Server Analysis Services enhancements**
  - Direct repository mode extraction (Tabular & Multidimensional)
- **Enable connection-aware scans for**
  - Microsoft PowerBI
  - Microsoft SQL Server Reporting Services
  - Tableau
- **Salesforce**
  - Support for Master-Details relationships
  - Connection assignment support
- **Databricks enhancements**
  - Databricks Workflows support
- **IDMC Metadata (Auto Cataloging)**
  - Support for CDI parameter files

- **Filter enhancements in:**
  - Lake family scanners
  - Redshift
  - MySQL
  - PostgreSQL
- **Authentication enhancements in ADLS Gen 2**
  - Shared Key Authentication
  - Managed Identity Authentication
- **Microsoft SQL Server linked servers support**
- **Oracle DBLinks support**
- **External Tables (Poly Base tables) support**
  - Microsoft SQL Server
  - Microsoft Synapse

Informatica

Summary of Features in CDGC 2024.02.S

- **Browse Improvements**
  - Find & Filter options
  - Row count to indicate Children
  - Browse option for Addl Asset types
- **Lineage Improvements**
  - Tooltips to show addl details,
  - Other usability enhancements
- **Improved Asset picker**
- **Quick Look enhancements**
  - Support for SAP sources
  - Support for Additional Technical Assets
- **Language Support**
  - Italian in CDGC
  - Japanese for QuickLook

- **Discovery Improvements**
  - Allow for both Auto Generated and OOTB/User Defined classifications
  - Ignore Prefix & Suffix in Glossary Association
  - Bulk Promotion of Auto Generated Classification
- **Discovery (IGA, Data Classification, Relationship Discovery) PAM:**
  - IICS, Hive, OneDrive
- **Profiling PAM**
  - Teradata, SAP HANA, SAP BW, SAP BW in Hana, External Tables (Oracle, Azure Synapse, Athena)
- **Scanners**
  - MariaDB, DB2 z/OS, Microsoft Fabric Data Warehouse

Cloud Data Marketplace

Data Marketplace

- Model Alignment with CDGC – removal of 'Linked Assets'

  - Asset page is a window to CDGC - no local names, editability or comments

  - No more 'onboarding' CDGC Assets as linked assets

- Enhanced experience for adding assets to data products

- Usability Enhancements

  - Shield Badge for CDAM

  - Cost Centre Description displayed

- Additional language support

  - Italian

  - Dutch

Scanners

Data Observability

Data Observability

Overview

- New functionality to detect anomalies within the data automatically

    - Row count

    - Various profiling metrics such as number of distinct values, max length etc.

    - Changes in data types for data elements

- Enable and schedule for catalogue sources of your choice in the MCC

- Available for all sources that support profiling in MCC

- Adds valuable information about data anomalies to CDGC via new Data observability events

Informatica

Overview

MCC Scan Job
(Data Profiling + DO enabled)

Profile Results
(New Storage)
Minimum 3 runs
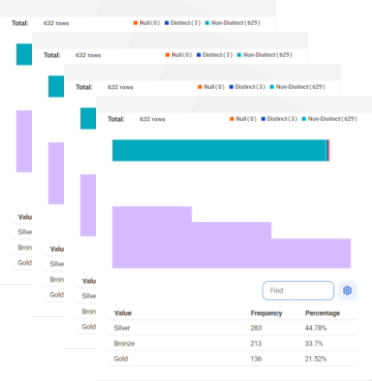
Results Comparison
& Event Generation

Visualize events within
CDGC

Data Observability

Configuration in the MCC

Configure and schedule

- New section in the source configuration "Data Profiling and Quality"

- Requires profiling without filters and in "full" scope

Data Observability

Dashboard widget

Customized visibility

- New widget type "Data Observability"

- Select the sources that should be observed

- Donut and list type

- Show all open events for selected sources
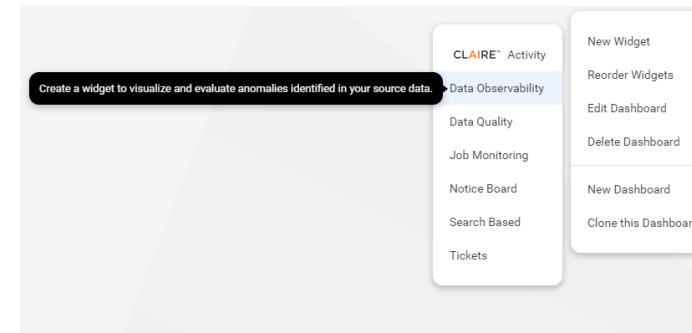
Data Observability

Technical asset view

Add visibility

- Overview with open events by severity

- New Data Observability tab
  - Donut charts grouped by metric and severity
  - List of events
  - Shows all events of this asset and child assets

Data Observability

Lineage overlay

Add context and relations

- New overlay "Data Observability"

- Show data sets or elements with open events and the number of open events

- Preview for events
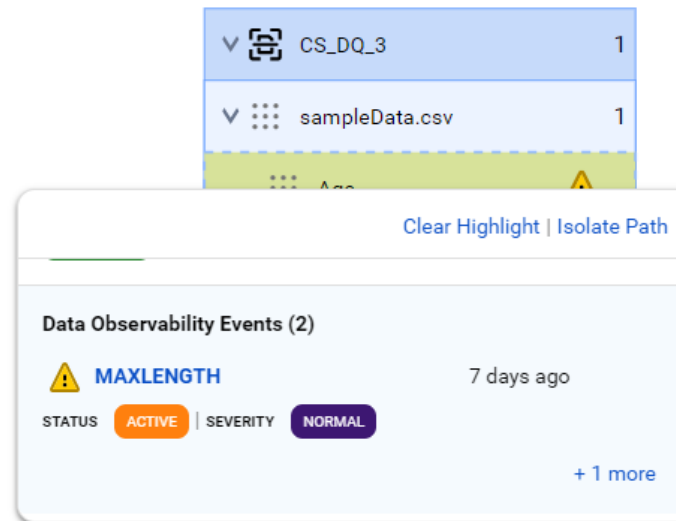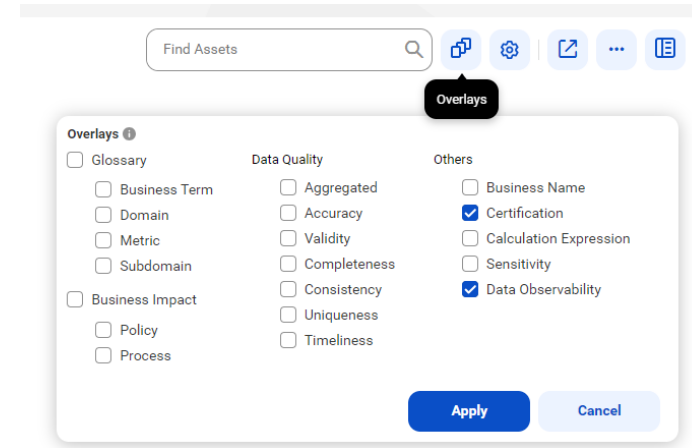
Data Observability

Event page

All information to take action

- Summary of what changed

- List of stakeholders inherited from the data set or data element

- Stakeholders can take actions

  - Resolve

  - Reject
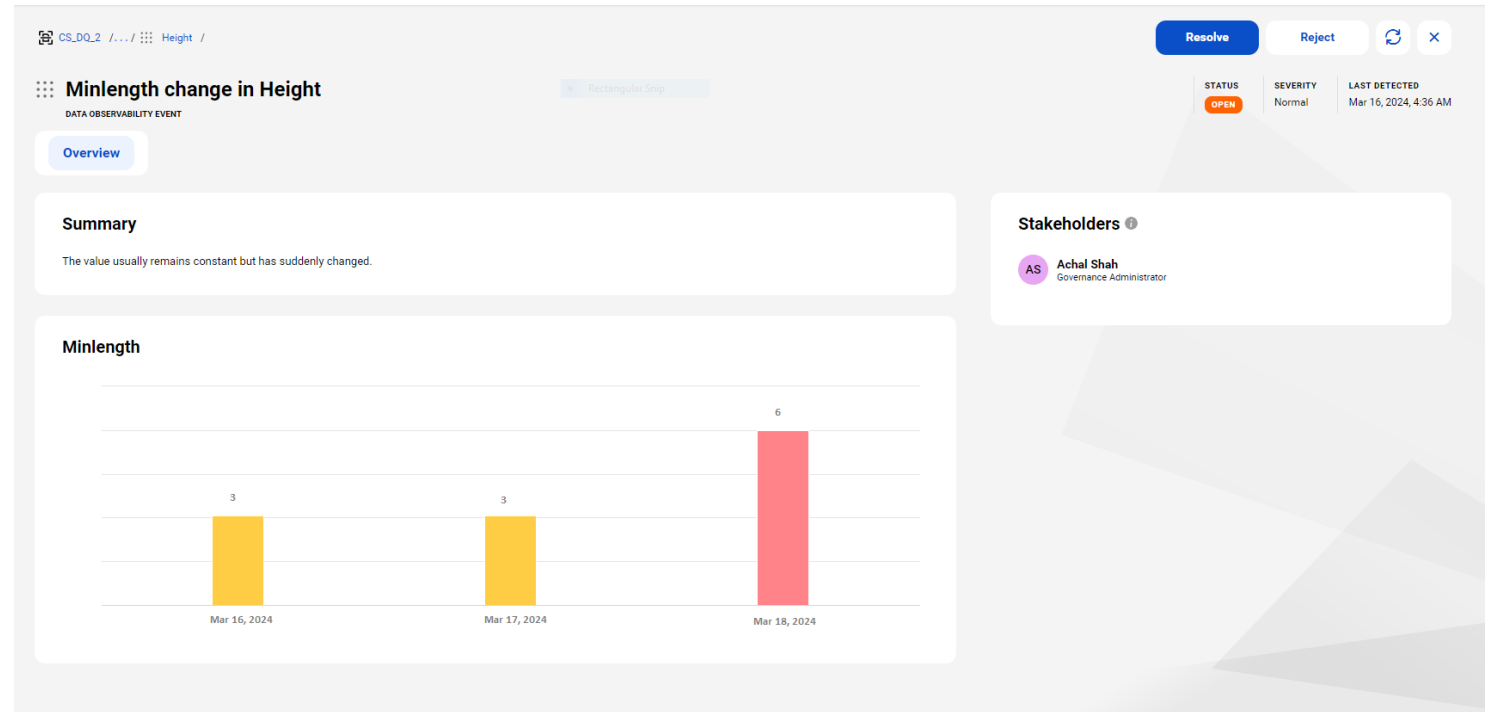
Cloud Data Quality

Cloud Data Quality

What's New in Cloud Data Quality and Profiling

Data Quality

- **Data Observability**

- **New Visual Design**

- **CDQ available in OCI POD**

- New Data Quality Bundle for EU

- Exception Management Validation for Advanced Rules

- Run test on sub-org using parent org shared secure agent

- **Support for Azure Serverless (Native)**

- Platform Updates

  - MS Visual C upgrade to 2019

  - Rocky Linux 8 – replacement CentOS

  - Cortex XDR 8.2

Data Profiling

- **New Visual Design**

- **Product Availability Matrix (PAM)**

  - Profiling **Databricks using Secure Agent**

  - **Athena** connectivity using **EC2 Role**

  - **MariaDB** using **JDBC_V2**

- New public API for lookup profiling tasks

- Improvements to source mapplet support - SQL overwrite and bugs

- Platform updates

  - MS Visual C upgrade to 2019

  - Rocky Linux 8 – replacement CentOS

  - Cortex XDR 8.2

Informatica

Informatica®

Designed for the business user

A New Visual Experience

Where data & AI come to LIFE

New Visual Design

- Objectives

  - Modernize the user interface

  - Improve UI response time

  - Align with other services already using the new design

  - Open doors for more future enhancements:

    - Accessibility

    - Dark Mode

    - Compact Mode
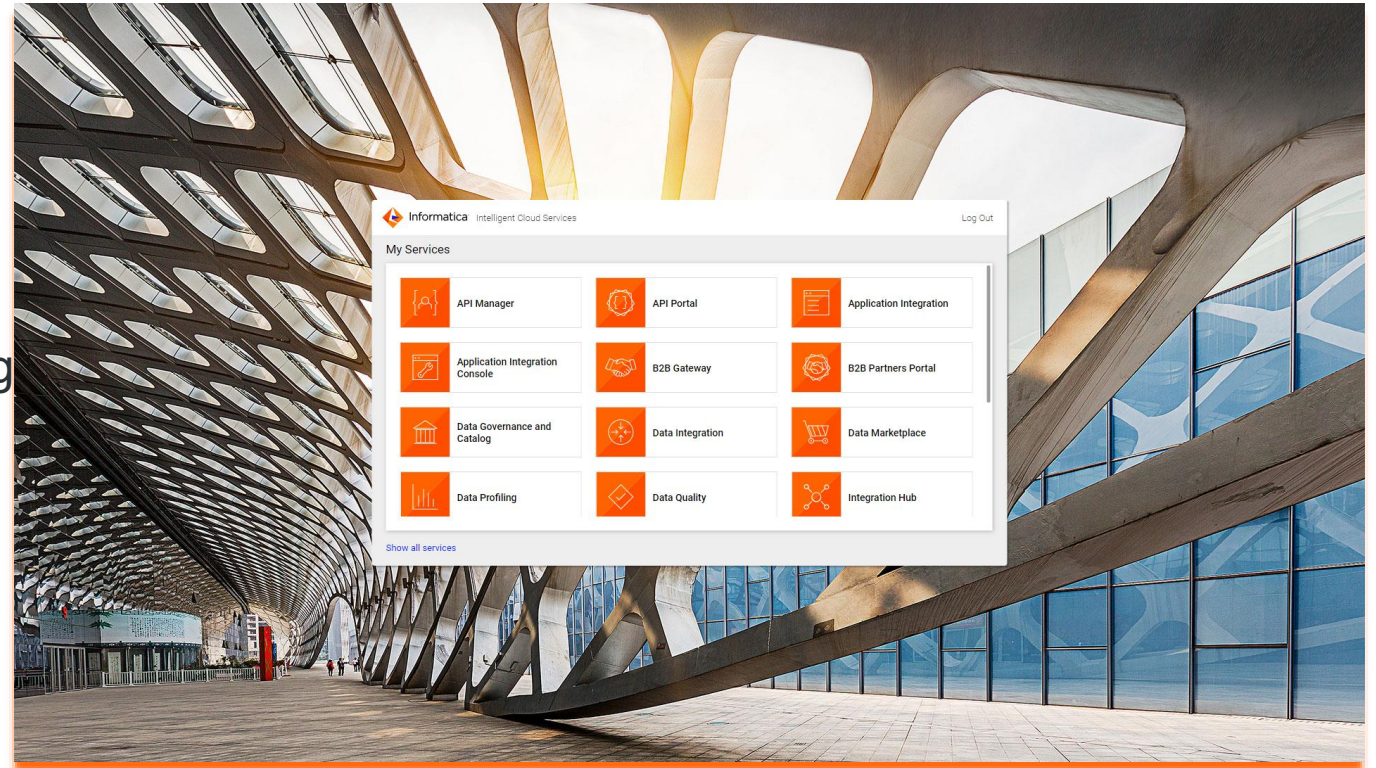
New UX Pattern Applied to CDQ Assets

New

Choose the asset type you want to create.

| Rule Specification | Dictionary | Verifier | Cleanse |
| Deduplicate | Parse | Labeler | |

New Visual Design

Cloud Data Profiling

New Visual Design

Cloud Data Profiling

New Visual Design

Rule Specification

New Visual Design

Rule Specification

Cloud Profiling

Profiling of Databricks

Secure agent – native mode

Use case

- Profiling of Databricks Delta Tables and External Tables

When to use Secure Agent – Native?	When to use Advanced/Elastic?
• When user requires data preview, drill down, filters, reduce the scope of profiled rows (First N), Data Quality Insights, SQL Override CONS • Cannot profile complex data types – content is treated as a string	• When user requires the profiling statistics for columns with **complex data types** (array, map, struct) is required CONS • Inability to profiled rows (First N),drill down, add filter, data preview, SQL Override

Detailed of features and functions supported in each runtime is here

Cloud Data Profiling

What sources can you profile? (starting April 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
- ODBC
- JDBC (NEW)
- 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 mapplets
- PARQUET, AVRO and JSON files on:
    - Amazon S3 - [Requires Advanced / Elastic]
    - Azure ADLS Gen2 - [Requires Advanced / Elastic]

Deprecated
- Azure Synapse, Snowflake, Redshift and Databricks using ODBC (still works, not recommended)
- Databricks Delta tables using ODBC driver

Cloud Data Quality

Rule Specification Advanced

Adding Exception Management validation

- A Rule Specification can only be configured with Exception Management (EM) metadata at the top-level.

- This is consistent for both Basic and Advanced Modes.

- Validation added to support only one Rule Specification can be configured with EM metadata

  - If the script in advanced rule refers another advanced rule which is configured with exception management, then once user clicks on the validate button, the validation message gets displayed.

Azure Serverless

- Azure Serverless is supported in Azure Pods only.

- User should have access to Azure Portal to setup Azure Serverless organization and Runtime.

- Create Informatica Organization from Azure Marketplace - '*Informatica Intelligent Data Management Cloud - An Azure Native ISV Service*'

Where data & AI come to
LIFE

Informatica

Where data & AI come to

LIFE

Informatica®