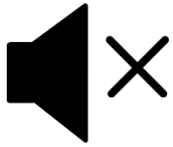June 20th, 2023

Best Practices for Implementing Cloud Data Governance and Catalog

- Kristin Feeback, Senior Consultant, IPS
- Steven Fleishman, Solution Architect, IPS

Informatica

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 INFASupport YouTube channel and 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.

Informatica

Bootstrap trial and POC Customers

Enriched Customer Onboarding experience
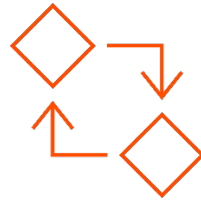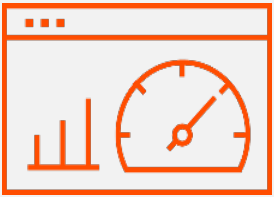
Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

Tailored training and content recommendations

Informatica

More Information
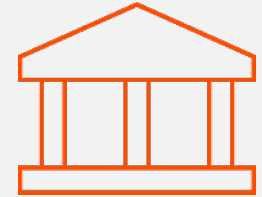
Success Portal

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

Informatica®

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

Agenda

1 Introduction	

1 Introduction

2 Technical Considerations

3 Roles and Responsibilities

4 Foundation Design

5 Data Quality Capabilities within CDGC

6 Question and Answer

Informatica

Technical Considerations

Account Verification and Services Screen

Welcome Email

Welcome to the Informatica family! We are so glad that you're here! Before you can dive into the great features we have to offer, please verify your account:

Verify me now

You have 48 hours to verify your account. If the link expires, please click here to get a new account confirmation email.

Services Available

© Informatica. Proprietary and Confidential.

Secure Agent Sizing
Requirements for CDGC

If agent status is inactive, communication Informatica Cloud Servers can be **blocked** by:
- Windows Firewall
- Virus Scanner
- Content Filter

Secure Agent Group

- **High Availability**
- **Load Balancing**
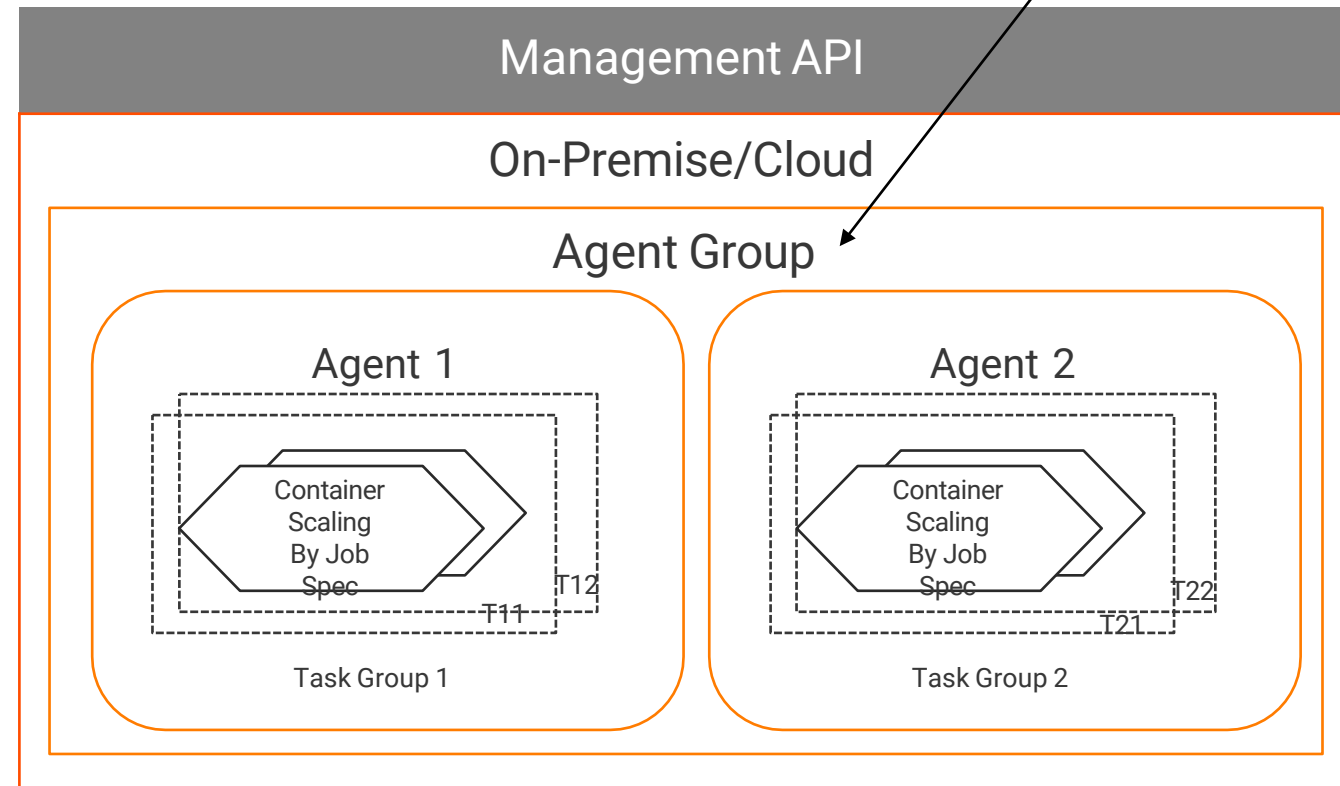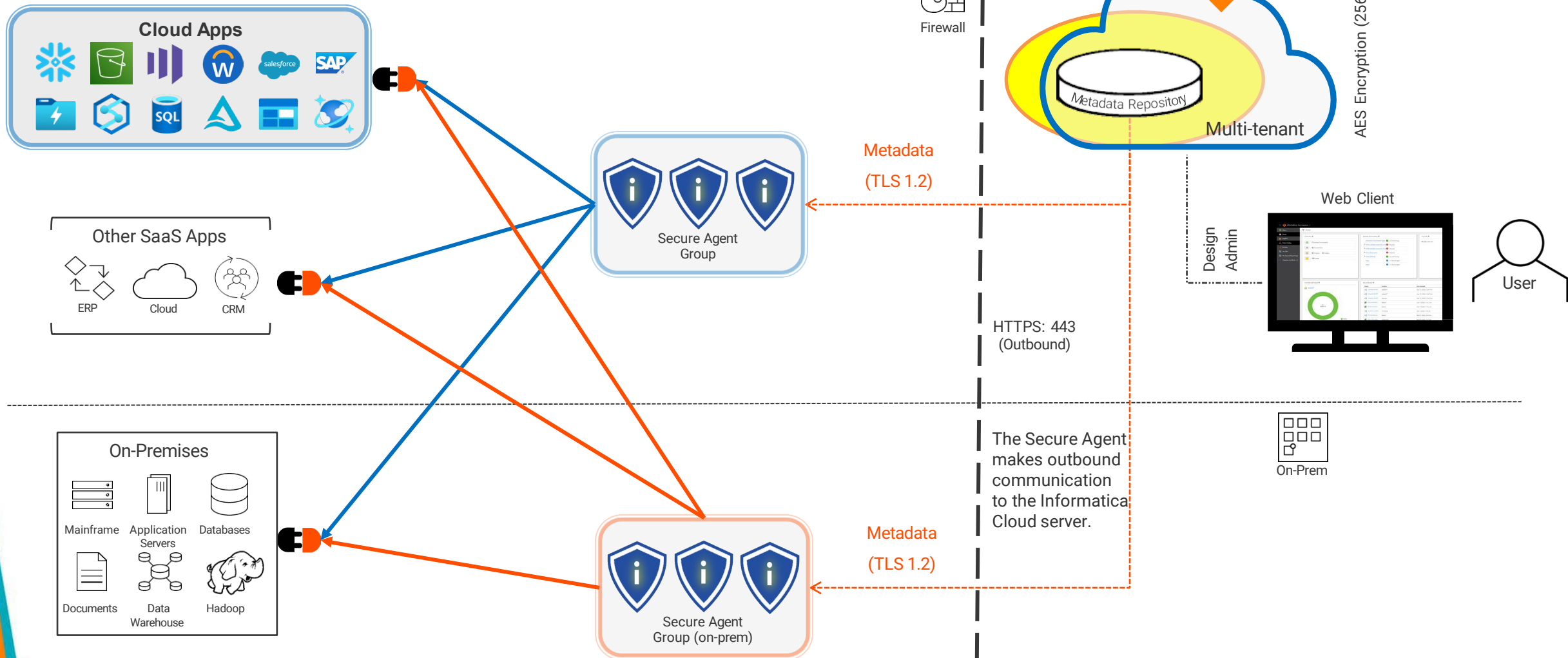- **Job isolation**

- Recommended Prerequisites
  - 16 CPU cores
  - 64 GB RAM
  - 200 GB HDD
  - SSD disks are strongly recommended
- Easy to deploy
  - Download and install the secure agent
  - Register agent using generated token
- Scale with IDMC agent grouping capabilities
  - 2 or more SA for sources more than 100k assets
- [All agents can execute all the capabilities (Sizing Guidelines)](#)
  - Metadata extraction
  - Profiling
- How To: [Plan Secure Agent(s) with Best Practices](#)

Management API

On-Premise/Cloud

Agent Group

Agent 1
Container Scaling By Job Spec
T12
T11
Task Group 1

Agent 2
Container Scaling By Job Spec
T22
T21
Task Group 2

Informatica

Network Architecture

Cloud Apps

Other SaaS Apps
- ERP
- Cloud
- CRM

On-Premises
- Mainframe
- Application Servers
- Databases
- Documents
- Data Warehouse
- Hadoop

Customer VNET | Public

Firewall

IDMC Services

AES Encryption (256 bit)

Metadata Repository

Multi-tenant

Secure Agent Group

Metadata (TLS 1.2)

HTTPS: 443 (Outbound)

The Secure Agent makes outbound communication to the Informatica Cloud server.

Secure Agent Group (on-prem)

Metadata (TLS 1.2)

Web Client

Design Admin

User

On-Prem

- Only metadata and profile results are sent back to Informatica Cloud
- For metadata in IDMC, Administrator can restrict to view Sensitive Data & Profile Statistics

23

Informatica

Secure Agent Connectivity

Oper

1. T

2. T

This e   ices can connect to IICS
serve

Job doesn't initialize

Base IICS Services

"failed to deploy task"

Cloud Data Profiling

Secure Agent Configuration | Concurrent Jobs

Update **maxDTMProcesses** to increase number of concurrent mapping tasks

Connection Type	Formula to set value	Example on 16 CPU Server
File-based or ODBC connections	0.75 * Number of logical CPUs	0.75 * 16 = 12
Cloud data warehouse or cloud data lake connectors	0.33 * Number of logical CPUs	0.33 * 16 = 5

Custom Configuration Details

Service	Type	Sub-type	Name	Value
Data Integration Server	Tomcat		maxDTMProcesses	12

Informatica®

Secure Agent Configuration | Memory

① Additional memory for metadata scanning
Increase **maxHeap_size**

Custom Configuration Details

Service	Type	Sub-type	Name	Value
Metadata Foundation Ap ▾	PLUGIN_CFG ▾	▾	maxHeap_size	16384

---

② Agent-wide virtual memory for all IICS Services
INFA_MEMORY
Out of the box it's 512m
Must be set **at least** to 2048M
'-Xms512m -Xmx8192m -XX:MaxPermSize=384m'

System Configuration Details   Reset All

Service: Data Integration Server
Type: Tomcat JRE

Type	Name	Value
Tomcat JRE	INFA_SSL	
Tomcat JRE	INFA_MEMORY	'-Xms32m -Xmx512m'
Tomcat JRE	JRE_OPTS	'-Xrs'
Tomcat JRE	JAVA_LIBS	

③ Memory on the Mapping level
JVMOption1
Improve performance
Reduce writing to disk
'-Xmx8192m'

System Configuration Details   Reset All

Service: Data Integration Server
Type: DTM

Type	Name	Value
DTM	JVMClassPath	'pmserversdk.jar'
DTM	JVMOption1	'-Xmx8192m'
DTM	JVMOption2	
DTM	JVMOption3	

Informatica

Don't Forget to Import Pre-Defined Content

Import Data Quality Bundles with Pre-Built Content

1

- Organization
- Licenses
- SAML Setup
- Metering
- Settings
- Users
- User Groups
- User Roles
- Runtime Environm...
- Serverless Environ...
- Connections
- Add-On Connectors
- Schedules
- Data Services Rep...
- Add-On Bundles
- Swagger Files
- Logs
- Advanced Clusters
- File Servers

2

Add-On Bundles

Installed Bundles | Copied Bundles | **Available Bundles**

Available Bundles (5 items found)   data quality

Name	Description	Publisher	Version	Bundle Type	Last Published
Cloud Data Quality Automation B...	With this release, Informatica makes available an Informatica Cloud Data Quality Au...	DQ Content Team	1.0	Public	Jun 1, 2020, 2:36:16 AM
Cloud Data Quality Crisis Respon...	The Crisis Response Accelerator can help you find and fix errors and inconsistencies i...	DQ Content Team	1.2	Public	Oct 14, 2020, 4:32:00 AM
Data Quality Americas Bundle	The Data Quality Americas Bundle is a set of pre-configured data quality assets that y...	DQ Content Team	2.0	Public	Jan 25, 2022, 7:25:38 AM
Data Quality Financial Bundle	Use the Financial Bundle to verify and enhance data from organizations in the financi...	DQ Content Team	1.0.1	Public	Feb 7, 2023, 12:06:13 AM
Data Quality GBR Bundle	The Data Quality GBR Bundle is a set of pre-configured data quality assets that you c...	DQ Content Team	1.1	Public	Feb 15, 2023, 2:27:51 AM

3

Bundle Content (230)

Name	Description	Type
c_dq_Validate_County	Validates if a given input is a valid US county.	Cleanse
mplt_Assign_DQ_Mailability_Score_Description	This mapplet provides a brief description or meaning for the Mailability Score. Tags: Accuracy	Mapplet
c_SSN_Max_Index_For_Group	Returns index for a group.	Cleanse
p_Parse_Name_Components_FML_USA	Parses input into First Name, Middle Name, Last Name for USA.	Parse
mplt_CAN_Phone_Number_Parse	This mapplet parses a phone number that follows the standard format for Canada from a string of text ...	Mapplet
p_Assign_Firstname_Gender_Can	Assigns gender on the basis of Firstname for Canada	Parse
dedupe_US_IMO_Indiv_Name_Address_Match	Matches individual name and address with Address Part1 as key.	Deduplicate
usa_company_names_std_infa	This reference table contains list of company name standards in US.	Dictionary
mplt_BRA_Company_Suffix_Standardization	Mapplet standardizes the company suffix if one is present. Tags: Brazil, Corporate, Integrity/Conform...	Mapplet
v_USA_Validation_Discrete_Address	Verifies US address in discrete format.	Verifier
rs_Title_Case	Converts given input to titlecase.	Rule Specification
mplt_USA_Phone_Number_Validation	This mapplet validates the area code for US phone numbers. Quality scores are: 0 - NULL or blank in...	Mapplet

Check Metadata Command Center Source to see if Connection is required in Administrator

Data Governance & Catalog Scanners

Scanners in Red requires administrator connection required in Administrator Service

Data Integration
Cloud Data Integration | PowerCenter | Microsoft SSIS | Talend

Azure
Azure SQL DB | Azure Synapse | Azure ADLS Gen 2 | Azure Blob | Azure Data Factory

BI & Analytics
Tableau | PowerBI | SSRS | QlikView | Qlik Sense | Databricks Notebooks | Cognos | Microstrategy | Looker

On-premises
Oracle | SQL Server | IBM D... MySQL | Teradata | Postgr... JDBC | Netezza | MongoDB | ... | Local/Shared Filesyste...

New Catalog Source

① Registration   ② Configuration   ③ Associations   ④ Schedule

General Information

Name: *   Enter the Catalog Source Name

Description:

Connection Information

Catalog Source Type: Snowflake

Connection: *

AWS
AWS S3 | AWS Redshift | ...S RDS (Oracle, MS SQL Server, ...eSQL and MySQL) | AWS Athena | DynamoDBl

Google & OCI
Google BigQuery | GCS Oracle ADB

CDL/DW
Snowflake DW | Databricks Delta Tables

Applications
SAP BW | SAP BW/4HANA | SAP ECC | SAP S/4HANA | SAP BO | SAP HANA DB Salesforce | Marketo | Dynamics CRM | Workday | Informatica MDM B360

Informatica

Roles and Responsibilities

Why are you using CDGC?

What was the deciding factor for wanting to start/bring data governance to the next level

Need to think about how the program wants users to interact with the solution and in what ways they may want to operationalize it.

Will there be a heavy data privacy, data quality, or master data management focus? Will priorities be split between multiple areas?
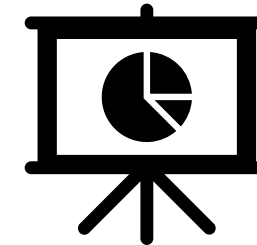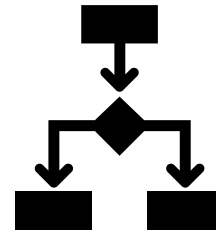
Preliminary Considerations
What needs to be done for the program to be functional?
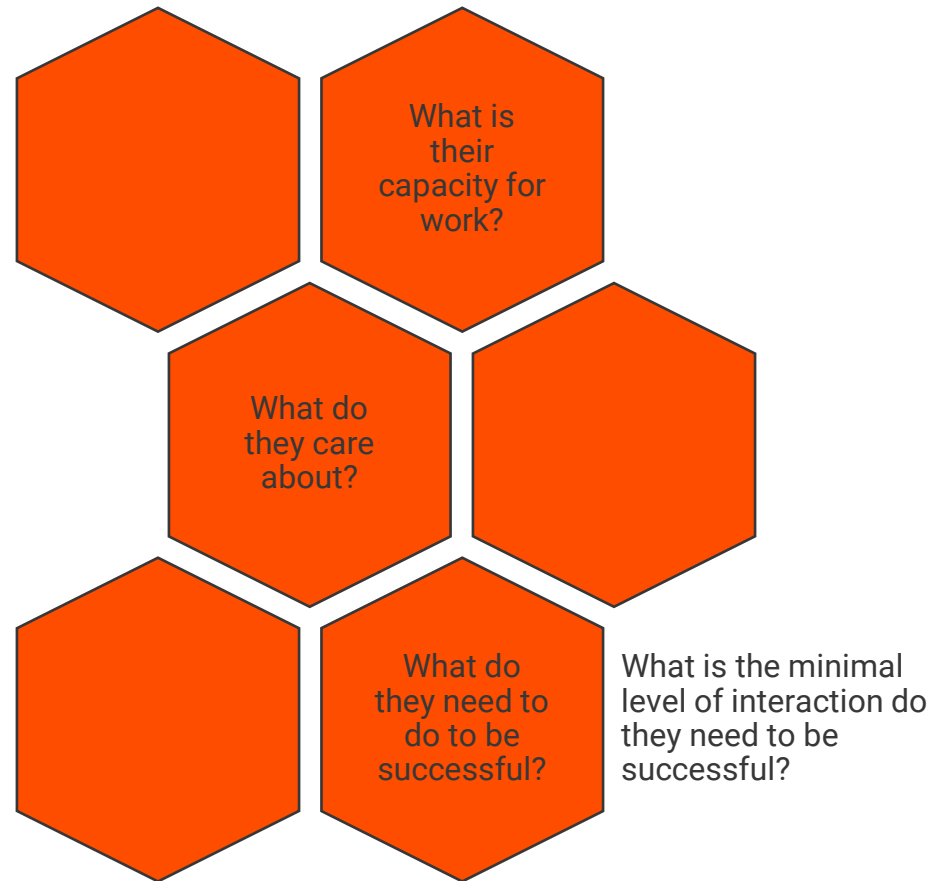
What operations need to be completed?
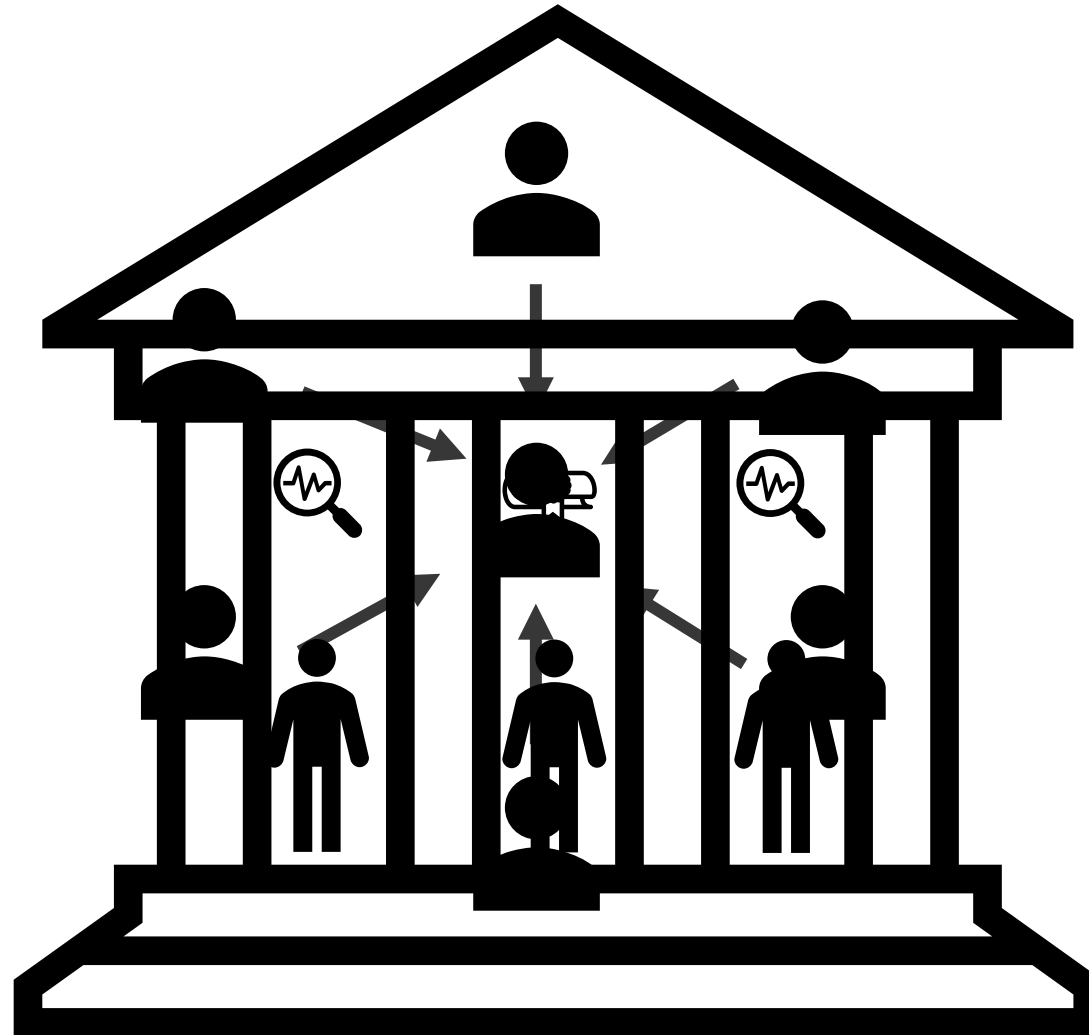
What is your current team's strengths and weaknesses?

Does your company already have a data first culture?

Informatica®

Preliminary Considerations
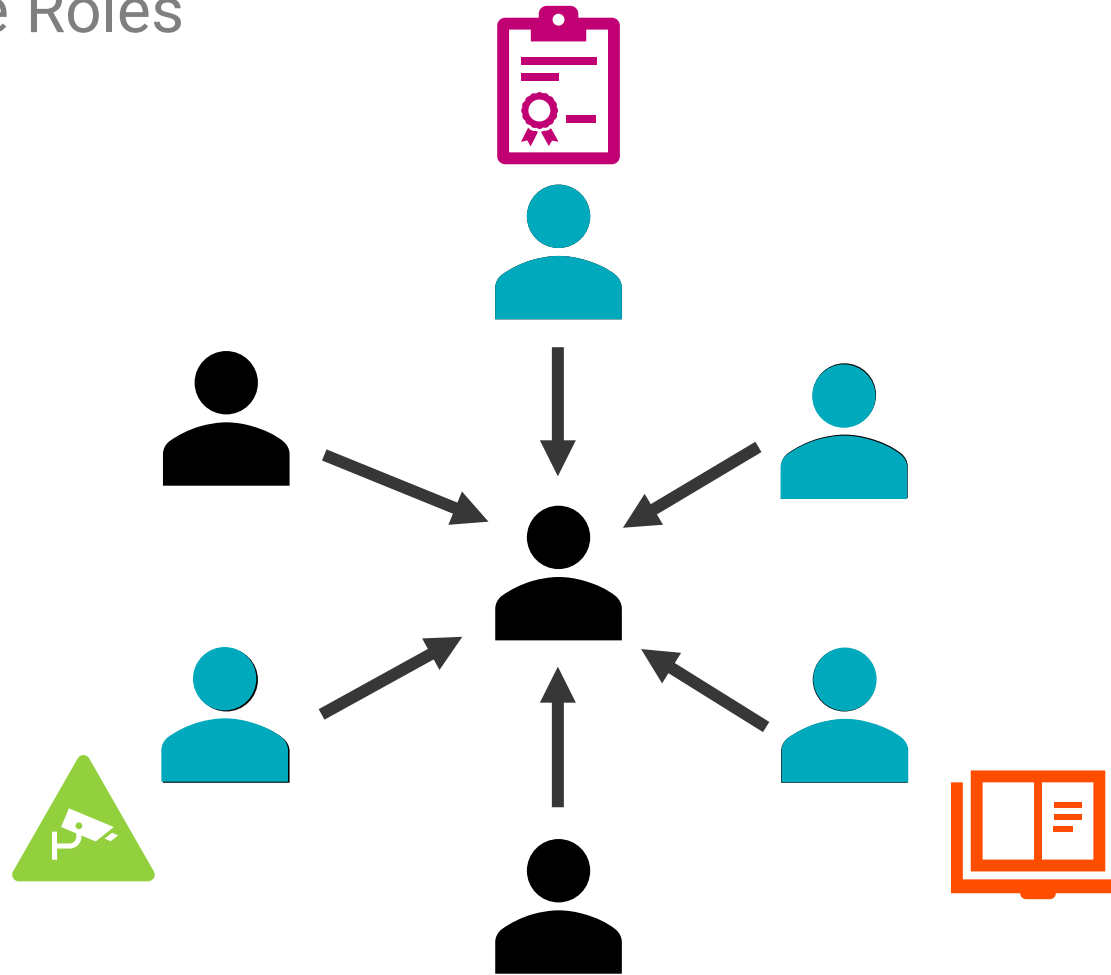## Who will be the primary stewards and owners?

What is their capacity for work?

What do they care about?

What do they need to do to be successful?

What is the minimal level of interaction do they need to be successful?

User interactions

User Interactions
Functions within the Roles

© Informatica. Proprietary and Confidential.

Informatica®

Catalog Data Steward

Communicates technical workings of data, fosters the goal of data re-use, and designs policies and processes to improve the underlying integrity of information to the organization. Responsible for uploading and curating catalog assets.

Asset Permissions

Services: Metadata Command Center

Assets   Features

Asset Type	Create	Read	Upd...	Delete	Run	Set Permi...
AI Model	☐	☑	☐	☐	☐	☐
Application Options	☐	☐	☐	☐	☐	☐
Business Reports	☐	☑	☐	☐	☐	☐
Business Term	☐	☑	☑	☐	☐	☐
Catalog Source Configura...	☐	☐	☐	☐	☐	☐
Consumer Access	☐	☐	☐	☐	☐	☐
Custom Catalog Source Ty...	☐	☐	☐	☐	☐	☐
Custom Models	☐	☐	☐	☐	☐	☐
Data Categories	☐	☐	☐	☐	☐	☐
Data Collection Orders	☐	☐	☐	☐	☐	☐
Data Collections	☐	☐	☐	☐	☐	☐
Data Quality	☐	☑	☐	☐	☐	☐
DataSets	☑	☑	☑	☐	☐	☐
Delivery Targets	☐	☐	☐	☐	☐	☐
Delivery Templates	☐	☐	☐	☐	☐	☐
Glossary Domain	☑	☑	☑	☐	☐	☐
Glossary SubDomain	☑	☑	☑	☐	☐	☐
Metric	☑	☑	☑	☐	☐	☐
Policy	☐	☑	☐	☐	☐	☐
Process	☐	☑	☐	☐	☐	☐
System	☑	☑	☑	☐	☐	☐
Technical Assets	☑	☑	☑	☐	☐	☐

MCC Permissions

Services: Metadata Command Center

Assets   **Features**

Feature Name	Enabled
Access Metadata Command Cen...	☑
Manage Access Control	☐
Manage Connection Assignments	☑
Manage Custom Attributes	☑
Manage Data Classifications	☑
Manage Reference Data	☑
Manage System Settings	☐
Manage Workflow Settings	☐
Monitor Jobs	☐
View Custom Attributes	☑
View Data Classifications	☑
View Reference Data	☑

Workflow Inbox

Services: Human Tasks

Assets   **Features**

Feature Name	Enabled
View Tasks	☑

CDGC Permissions

Services: Data Governance and Catalog

Assets   **Features**

Feature Name	Enabled
Access Data Governance And C...	☑
Curate Automatic Glossary Assoc...	☑
Curate Data Classifications	☑
Export	☑
Import	☑
Participate in Change Approvals	☑
Stakeholdership	☑
View Business Assets	☑
View Data Classifications	☑
View Profiled Stats	☑
View Sensitive Data	☑
View Technical Assets	☑
View Unpublished Content	☑

Informatica

Governance Data Steward

Responsible for working with business and technical stakeholders to ensure high levels of data quality, integrity, availability, trustworthiness and security. Creates and curates business assets.

Asset Permissions

Services: Metadata Command Center

Assets | Features

Asset Type	Create	Read	Upd...	Delete	Run	Set Permi...
AI Model	☐	☑	☐	☐	☐	☐
Application Options	☐	☐	☐	☐	☐	☐
Business Reports	☑	☑	☑	☐	☐	☐
Business Term	☑	☑	☑	☐	☐	☐
Catalog Source Configura...	☐	☐	☐	☐	☐	☐
Consumer Access	☐	☐	☐	☐	☐	☐
Custom Catalog Source Ty...	☐	☐	☐	☐	☐	☐
Custom Models	☐	☐	☐	☐	☐	☐
Data Categories	☐	☐	☐	☐	☐	☐
Data Collection Orders	☐	☐	☐	☐	☐	☐
Data Collections	☐	☐	☐	☐	☐	☐
Data Quality	☐	☑	☐	☐	☐	☐
DataSets	☑	☑	☑	☐	☐	☐
Delivery Targets	☐	☐	☐	☐	☐	☐
Delivery Templates	☐	☐	☐	☐	☐	☐
Glossary Domain	☐	☑	☑	☐	☐	☐
Glossary SubDomain	☐	☑	☑	☐	☐	☐
Metric	☑	☑	☑	☐	☐	☐
Policy	☑	☑	☑	☐	☐	☐
Process	☑	☑	☑	☐	☐	☐
System	☐	☑	☐	☐	☐	☐
Technical Assets	☐	☑	☐	☐	☐	☐

CDGC Permissions

Services: Data Governance and Catalog

Assets | **Features**

Feature Name	Enabled
Access Data Governance And C...	☑
Curate Automatic Glossary Assoc...	☑
Curate Data Classifications	☑
Export	☑
Import	☑
Participate in Change Approvals	☑
Stakeholdership	☑
View Business Assets	☑
View Data Classifications	☑
View Profiled Stats	☑
View Sensitive Data	☑
View Technical Assets	☑
View Unpublished Content	☑

Workflow Inbox

Services: Human Tasks

Assets | **Features**

Feature Name	Enabled
View Tasks	☑

Informatica®

Role Interactions

Customer Birthdate
BUSINESS TERM

OVERALL RATING
☆ ☆ ☆ ☆ ☆
Add

CRITICAL DATA ELEMENT
NO

LIFECYCLE
PUBLISHED

LAST UPDATED
Oct 27, 2022, 6:24 AM

Overview | Hierarchy | Relationships | Data Quality | Stakeholders | Tickets | History

Description

date customer was born

Attributes

Parent:

Person

Reference ID:

KF-BT-002

Stakeholders

Legal Governance Data Owners
Governance Data Owner

Alexandra Sommers
Governance Data Owner

Kristin Feeback
Governance Data Steward

View All

Informatica

Role interactions

© Informatica. Proprietary and Confidential.

Foundation
Design

Foundation for the Future

CDGC's strength for the glossary lies in depth of information and its relationships.

CDGC should be kept in a standard model to maximize use and growth.

Keep in mind that the use case is just giving you the first set of information to input information.

Harder to change the model once the use case information is built out.

Where do things belong?

Informatica

Finding the start…

Starting concept…

© Informatica. Proprietary and Confidential.
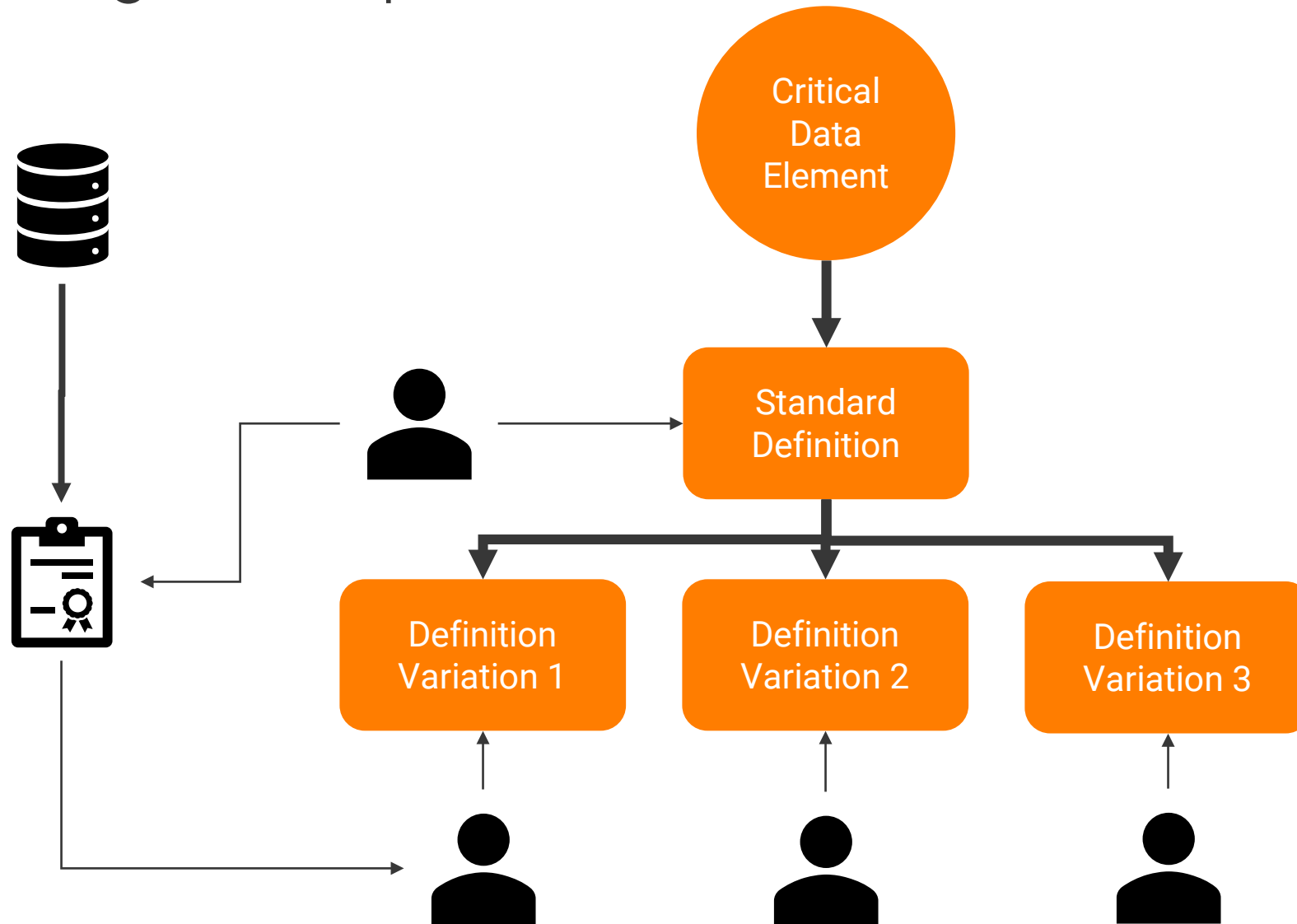
Collection of assets
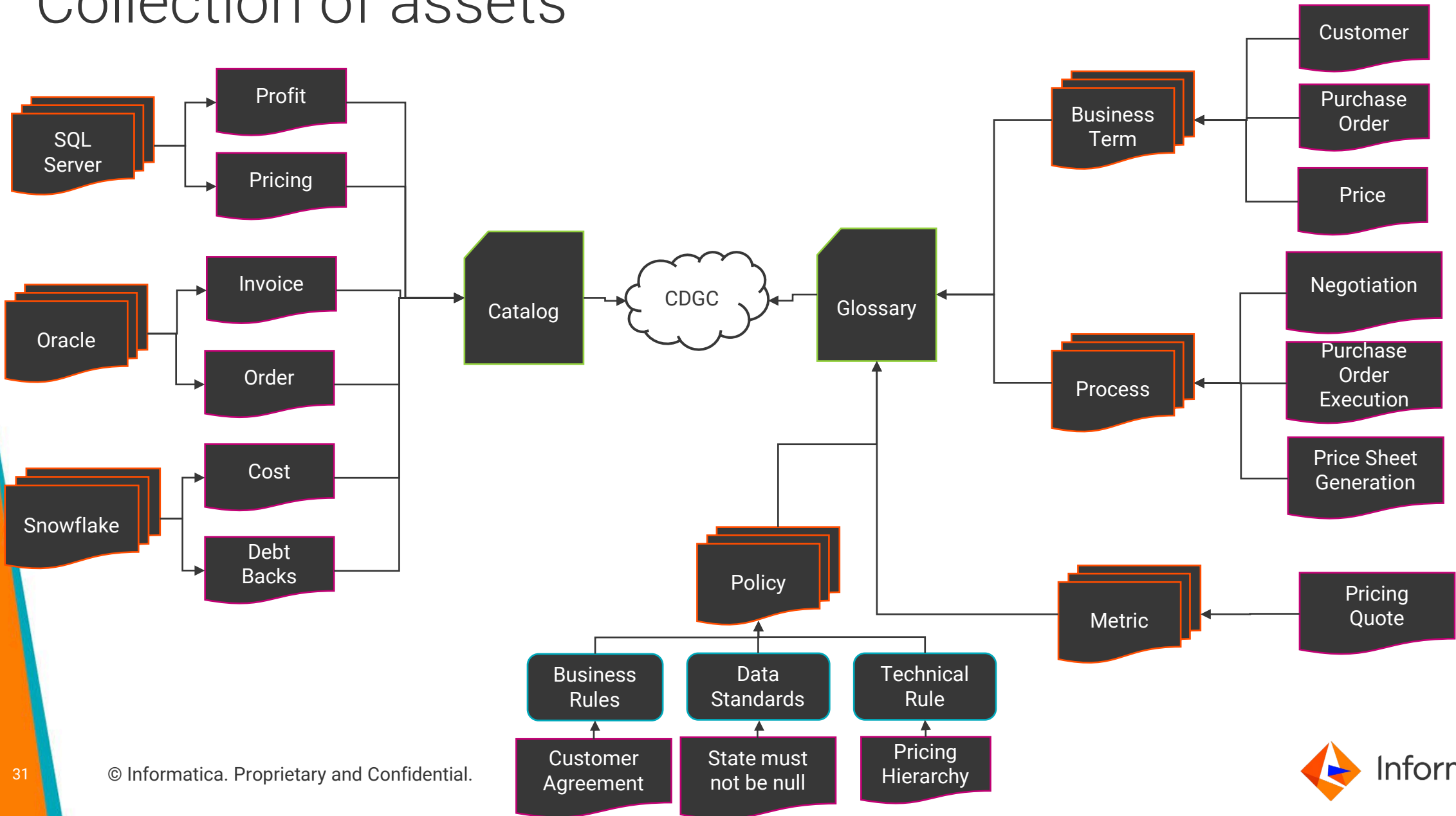
Data Quality within CDGC

Data Governance and Catalog After Enrichment
Profiling, Classification, Glossary Association, Relationship Discovery

Overview | **Contains** | Lineage | Relationships | Data Quality | Stakeholders | Properties | Tickets | History

Data Elements (10)

Find

# ↑	Name	Type	Glossaries	Classifications	CLAIRE™ Re…	Null | Distinct | Non-Distinct %	Source Data Type	Inferred Data Type
1	CUSTOMER_ID	Column	Party ID				VARCHAR2	Date (28.00%) +4
2	PARTY_NAME	Column					VARCHAR2	String (100.00%)
3	ADDRESS_LINE1	Column		Address			VARCHAR2	String (68.00%)
4	ADDRESS_LINE2	Column		Address			VARCHAR2	String (74.00%)
5	ADDRESS_LINE3	Column		Address			VARCHAR2	
6	ADDRESS_LINE4	Column		Address +1			VARCHAR2	Integer (76.00%) +3
7	CITY	Column	City	City			VARCHAR2	String (100.00%)
8	PROVINCE	Column					VARCHAR2	Integer (22.00%) +2
9	POSTAL_CODE	Column	Postal Code	Postal Code			VARCHAR2	String (99.00%)
10	COUNTRY	Column	Country	World Cities +1			VARCHAR2	Fixed Length String (100.00…

Informatica®

Informatica Cloud Data Governance and Catalog
Data Quality Process

Metadata Command Center

Data Profiling

Data Governance and Catalog

Data Quality

Cloud Data Governance and Catalog
Review DQ scores

Cloud Data Governance and Catalog
(If adhoc) Start Rule Occurrence

Cloud Data Governance and Catalog
Create Data Rule Template with Automation enabled

Metadata Command Center
Ensure Data Quality is configured

Data Profiling
Validate rule specification

Cloud Data Quality
Create Rule Specification

Identify a data Ruleset

Data Profiling
Collect additional information on the data

Cloud Data Governance and Catalog
Connect data to a business term

Metadata Command Center
Run Metadata Extraction and enrichment, including Data Quality

Job won't run unless DQ is enabled

There are multiple ways to create DQ Rule Specifications

- Data Profiling - Generated from Data Quality Insights
- Data Governance and Catalog – Leveraging Natural Language Processing
- Data Quality – Create a new Rule Specification
- Administrator – Data Quality Bundles

Informatica®

Metadata Extraction - Data Quality/DQ Enabled

Metadata Command Center

① Registration ② Configuration ③ Associations ④ Schedule

✓ Metadata Extraction | Data Profiling and Quality | Data Classification | ✓ Relationship Discovery

Data Profiling

Data Quality

Enable Data Quality: ⬤

Use Data Profiling Parameters

Connection and Runtime

Runtime Environment: | vm-test-dgov-01 |

Parameters

Data Quality Automation: * ⊙   ⦿ Yes   ○ No

| Apply on all Data Elements |

Cache Result: * ⊙ | Agent Cache |

Sampling Type: * ⊙ | Limit N Rows |

No of rows to limit: * | 1000 |

Mapping Submission Timeout: ⊙ | 20 |

Maximum Precision of String Fields: ⊙ | 50 |

Text Qualifier: * ⊙ | Double Quote |

- Apply on data elements linked with business data set
- Apply on all data elements

Sampling options based on source
- All Rows
- Limit N Rows
- Random N Rows
- Random N Percentage
- Custom Query

✓ Metadata Extraction | Data Profiling and Quality | ✓ Data Classification | ✓ Relationship Discovery | ✓ Glossary Association

Enable Metadata Extraction: ⬤

Connection and Runtime

Runtime Environment: ⊙ | Informatica Managed Infrastructure |

Filters

Specify metadata filters.
○ No filters ⊙   ⦿ Filter conditions ⊙

> Show supported wildcard characters and examples

| Include Metadata | Database.Schema | Named | ABC.* ✕ |

Configuration Parameters

Process programmable objects: * | Yes |

Default variables values: ⊙ | |

Informatica

Associate Technical Asset to Business Term

© Informatica. Proprietary and Confidential.
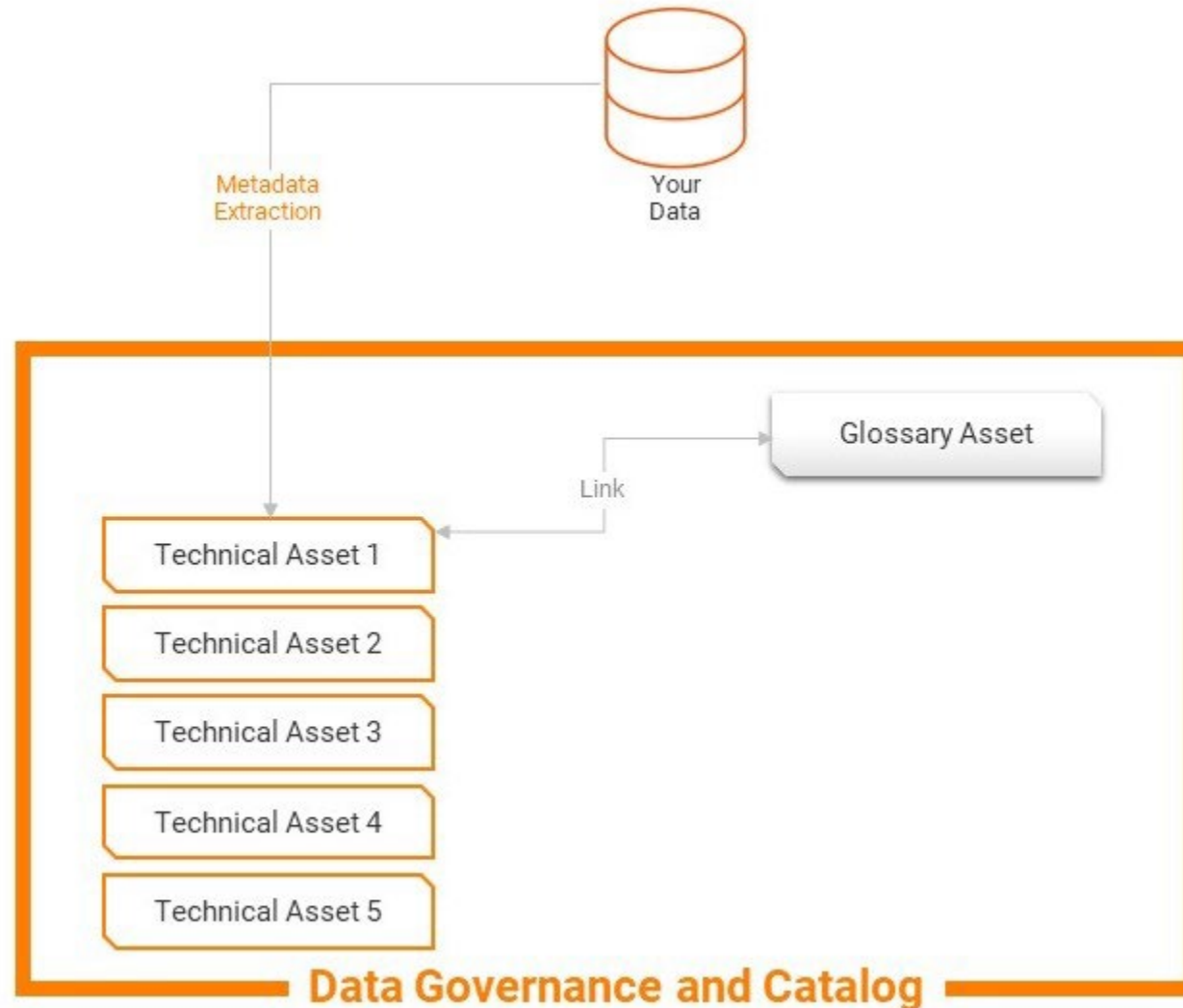
Business Term Associated to Four Columns

Data Governance and Catalog

State
BUSINESS TERM

Overview   Hierarchy   **Relationships**   Data Quality   Stakeholders   Tickets   History

6 All   **4** Column   **1** Data Quality Rule Template   **1** Subdomain

Column (4)
Data Quality Rule Template (1)
Subdomain (1)

Name	Type	How It Is Related
STATE	Reference Column	**STATE** Linked Glossary this Asset
STATE	Reference Column	**STATE** Linked Glossary this Asset
State Name Completeness	Data Quality Rule Template	**State Name Completeness** applies to this Asset
STATE	Reference Column	**STATE** Linked Glossary this Asset
STATE	Reference Column	**STATE** Linked Glossary this Asset
Postal Address	Subdomain	**Postal Address** is the Parent of this Asset

Informatica

Data Profiling – Create Connection and Run Profile

Collect Information & Create Rule Set

Data Profiling

| Profile 1 |

| Profile 1 |

Results   Definition   Rules   Schedule   **Insights**

View: [All Insight Types ▼] in: [All Columns ▼]   ☑ ☒ ↺ | ☐ Hide Rejected Insights

	Insight Statement	Score	Insight Type	Columns
☐	The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	STREET_ADDRESS2
☐	The number of tokens in the column has a high standard deviation.	Medium	Column Token Deviation	STREET_ADDRESS2
☐	The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	MIDDLE_NAME
☐	The number of tokens in the column has a high standard deviation.	Low	Column Token Deviation	MIDDLE_NAME
☐	The data may contain special characters	Low	Special Characters	MIDDLE_NAME
☐	Numeric values found outside the 95% standard deviation range.	Low	Number Value Distribution	FEDERAL_WITHHOLD_CODE
☐	The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	CELL_PHONE2
☐	The number of tokens in the column has a high standard deviation.	High	Column Token Deviation	CELL_PHONE2
☐	The length of the data values in the column has a high standard deviation.	Low	Column Length Deviation	FEDERAL_WITHHOLD_DESCRIPTI
☐	The number of tokens in the column has a high standard deviation.	Low	Column Token Deviation	FEDERAL_WITHHOLD_DESCRIPTI
☐	The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	ZIP_CODE
☐	The number of tokens in the column has a high standard deviation.	Medium	Column Token Deviation	ZIP_CODE
☐	The length of the data values in the column has a high standard deviation.	Low	Column Length Deviation	FIRST_NAME
☐	The number of tokens in the column has a high standard deviation.	Low	Column Token Deviation	FIRST_NAME
☐	The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	STATE_NAME
☐	The data may have inconsistent spelling.	High	Spelling Analysis	STATE_NAME

Test		Value	is	NULL		then	'Valid'
or	if	Child_0	is greater than or equal to	9		then	'Invalid'
or	if	Child_0	is less than or equal to	2		then	'Invalid'
or	if	Value	is not	NULL	🗑		
	and	Child_0	is less than	9	🗑		
	and	Child_0	is greater than	2	🗑	then	'Valid'
or	if		No rule statement is valid			then	Do Nothing

Informatica

Create the Business Rule

Create, Test, and Save a Rule Specification

Open the Data Quality Service

New

Select the type of asset you want.

Rule Specification	Dictionary	Verifier	Cleanse
Deduplicate	Parse	Labeler	

Valid State Length

Definition   **Configuration**

▶ Design

Properties: PrimaryRuleSet

General	Rule logic definition			
Inputs		Input	Operator	Condition
Rule Logic	if	in_State_Name	is not	NULL
	if	Result of # 1	is less than or equal to	4
Test	or	if	No rule statement is valid	

Define Action

Actions   Find

▶ Merge
▶ Input
▶ Functions
▶ Standardize
▶ Convert Case
▶ Rules
▶ Mapplets
▶ Expressions

In the Data Quality Service, this appears under "Pick an existing rule"

Items Technical Rule Reference

○ Create a Rule   ● Pick an existing rule

Explore ⌄   All Assets ⌄

All Assets (2 items found)   ⇅ ▽ state ✕ ?

Name	Type	Description	T..
Rule Check State Field Length	Rule	IC	
Valid State Length	Rule	IC	

Associate Rule Template to Glossary & Rule Specification
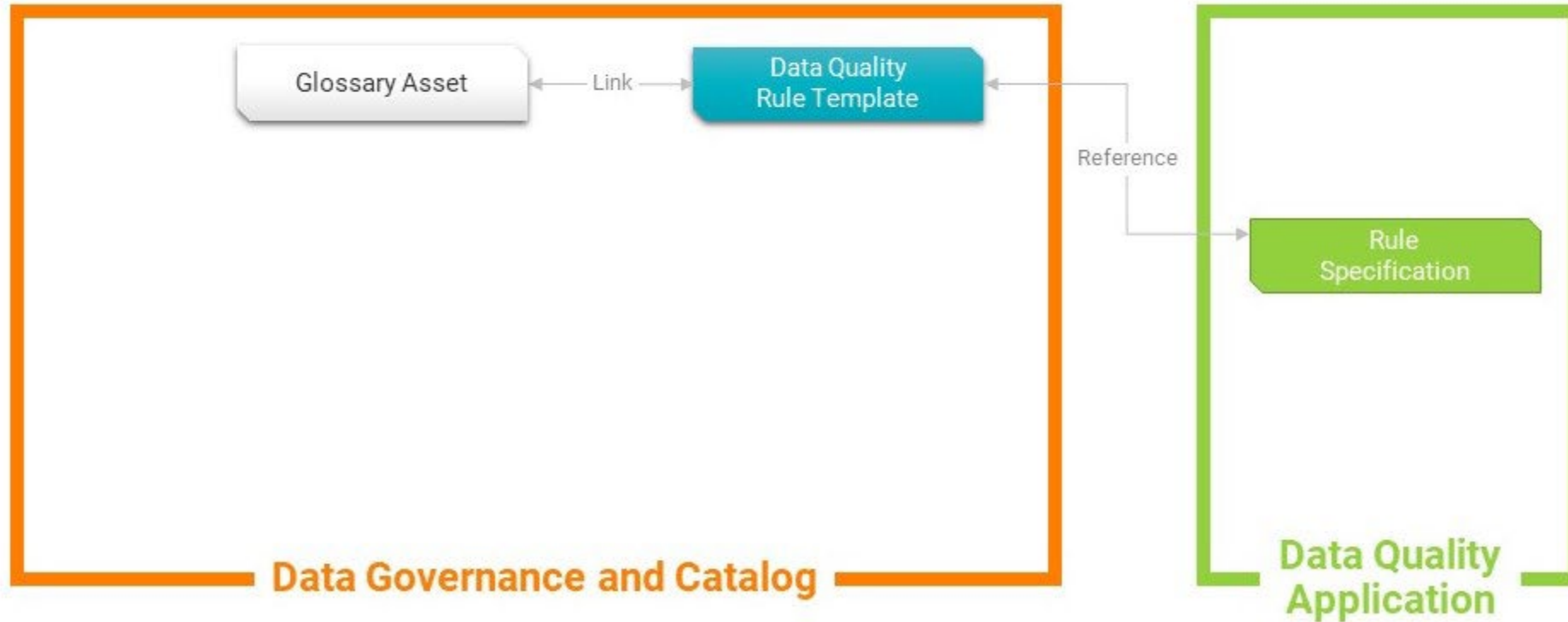
Data Quality Rule Template and Natural Language Processing

Data Governance and Catalog

Data Quality Rule Template

State Name Completeness
DATA QUALITY RULE TEMPLATE

Name: * — State Name Completeness

Description: H₁ B I U S ... — Enter a description for the rule template.

Reference ID: — Enter a unique identifier for the asset. If left blank, the Reference ID is automatically generated.

Dimension: * — Completeness

Measuring Method: * — Business Extract

Primary Glossary: * — State

Secondary Glossary:

Technical Rule Reference: * — Valid State Length

Criticality:

Automation: ☑

Target: * — 95

Threshold: * — 90

Frequency: — Daily

Items Technical Rule Reference

● Create a Rule ○ Pick an existing rule

Describe the rule ⓘ

State Name must not be less than 4 characters

Click here to use the description from the rule template.

View Recommendations

CLAIRE® Recommendations

Select to view details and create a new rule:

● IF Length(State) >= '4' THEN TRUE

CLAIRE® will generate and attach the following rule:

Rule Name — BT_Stat_Com_May_27_2023_133543

Description — This is a CLAIRE® generated rule based on the description State Name must not be less than 4 characters

Input — State

NLP texts to consider

You can use the following types of NLP rule descriptions. If the text you enter is insufficient or unclear, CLAIRE® might recommend that you rephrase the rule description.

Empty Value Rules

The following examples show descriptions of null value rules that CLAIRE® NLP technology can read:
- Sum may be NULL
- keep the section field empty
- Box value should be nullified
- Title should be Nothing
- Blank values are equal to null

Non-Empty Value Rules

The following examples show descriptions of rules that are not null values that CLAIRE® NLP technology can read:
- Age is always NOT NULL
- Name field cannot be blank or empty
- RFID values is not Null
- Radius must not be blank or empty
- doc number is not NULL

Comparison Rules

The following examples show descriptions of comparison rules that CLAIRE® NLP technology can read:
- Down payment number is greater or equal to thousand
- Down payment is going to either be 1000 or be above 1000
- Value cannot be less than 1,000

NLP texts to avoid

Avoid the following types of NLP rule descriptions. If the text you enter is insufficient or unclear, CLAIRE® might recommend that you rephrase the rule description.

Double Negation Sentences

The following examples show descriptions of double negation sentences that CLAIRE® NLP technology cannot read:
- It is not defined city name to be not null
- Distance is not equal to or lesser than NOT NULL
- Date of birth cannot be not null
- It is not acceptable for Phone Number to be not NULL

Complex Sentences

The following examples show descriptions of complex sentences that CLAIRE® NLP technology cannot read:
- It's not acceptable for date to equal NULL
- Nothing should be populated in name field
- Joining Date should not contain the value of date but contains numbers
- Product is not equal to 2 or not larger than 1

Grammatically Incorrect Sentences and Incorrect Spellings

The following examples show descriptions of grammatically incorrect sentences and incorrect spellings that CLAIRE® NLP technology cannot read:
- First name should be $NULL
- Serial cannot contain numerics
- Last_name is not a number and alphasnumeric
- Product is nt equal to or nt larger then 1
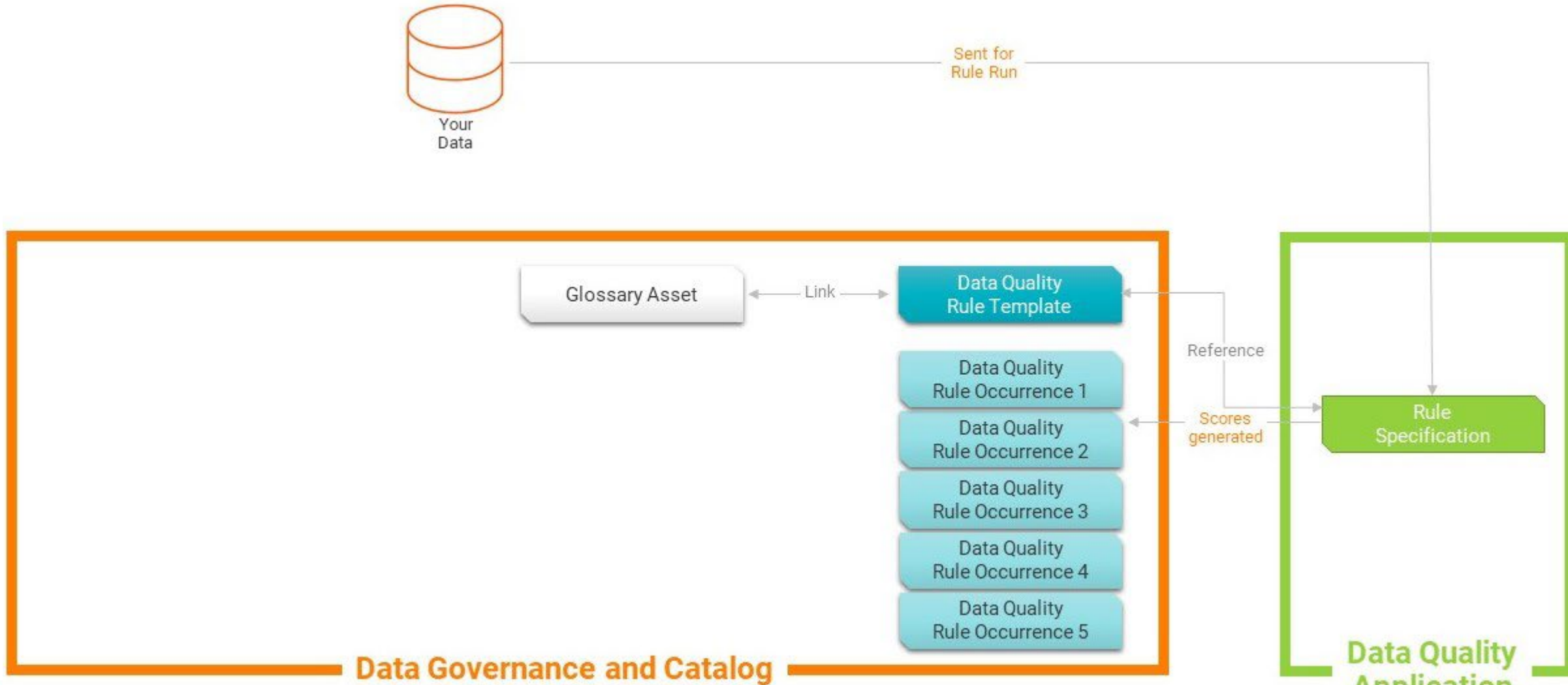
Informatica®

Automate Rule Occurrence Generation

© Informatica. Proprietary and Confidential.

Automate Rule Occurrence Generation

Data Governance and Catalog

State Name Completeness
DATA QUALITY RULE TEMPLATE

Name: * State Name Completeness

Description:

Enter a description for the rule template.

Technical Rule Reference: * Valid State Length

Criticality:

Automation: ☑

Data Quality Rule Automation ✕

You have selected automation for the rule template. This will create rule occurrences for the data elements that are associated with the primary glossary. After you save the rule template, you cannot unselect the automation option later.

Do you want to enable rule automation?

[Yes] [No]

Top Level

State Name Completeness
DATA QUALITY RULE TEMPLATE

Overview | Rule Occurences | Stakeholders | Tickets | History

Description

Confirms if the State Name has at least 4 characters

Primary Glossary: | Secondary Glossary:
State |

Reference ID: | Occurences:
DQT-2 | 4 Rule Occurences

State Name Completeness
DATA QUALITY RULE TEMPLATE

OVERALL RATING ☆☆☆☆☆ Add | LIFECYCLE PUBLISHED

Overview | Rule Occurences | Stakeholders | Tickets | History

Dimensions Latest Scores

Completeness (4)

7.14%

Not Applicable

Rule Occurrences (4)

Name	Dimension	Latest Score	Date	Tot
Auto-STATE-Completeness-State Name Completeness	Completeness	● 7.14%	May 28, 2023, 6:0...	14
Auto-STATE-Completeness-State Name Completeness	Completeness		May 28, 2023, 5:4...	
Auto-STATE-Completeness-State Name Completeness	Completeness		May 28, 2023, 5:4...	
Auto-STATE-Completeness-State Name Completeness	Completeness		May 27, 2023, 1:3...	

CRM_CUSTOMER
TABLE

OVERALL RATING ☆☆☆☆☆ Add | LIFECYCLE PUBLISHED

Overview | Contains | Lineage | Relationships | Data Quality | Stakeholders | Properties | Tickets | History

Dimensions Latest Scores

Consistency (2) 30% ↑+0
Completeness (3) 99.5% ↑+0
Validity (3) 100% ↑+0

Rules Against Target

● Good
● Acceptable
● Not Acceptable

Consistency (2)
Completeness (3)
Validity (3)

Count: 0 1 2 3

Rule Occurrences (8)

Name	Dimension	Latest Score	Date	Total Rows	Failed Rows	Primary Data Element	Secondary Data Element	Description	Related Rule Template
Auto-COUNTRY-Validity-Country - Validit...	Validity	● 100%	Jun 1, 2023, 1:15 ...	100	0	COUNTRY	-	country should exist in country referenc...	Country - Validity - v2
Auto-CUSTOMER_ID-Validity-Party ID - V...	Validity	● 100%	Jun 1, 2023, 1:15 ...	100	0	CUSTOMER_ID	-	Validity - Party ID - should be integer	Party ID - Validity
Auto-POSTAL_CODE-Consistency-Posta...	Consistency	● 0%	Jun 1, 2023, 1:15 ...	100	100	POSTAL_CODE	-	Postal Code - Consistency	Postal Code - Consistency
Auto-CITY-Consistency-City - Consisten...	Consistency	● 60%	Jun 1, 2023, 1:15 ...	100	40	CITY	-	City - Consistency	
Auto-COUNTRY-Validity-Country - Validity	Validity	● 100%	Dec 20, 2022, 8:0...	100	0	COUNTRY	-	country should exist in country referenc...	
Auto-COUNTRY-Completeness-Country -...	Completeness	● 100%	Dec 19, 2022, 1:3...	100	0	COUNTRY	-	Country - Completeness	
Auto-POSTAL_CODE-Completeness-Pos...	Completeness	● 99%	Oct 10, 2022, 9:0...	100	1	POSTAL_CODE	-	Postal Code - Completeness	Postal Code - Completeness
Auto-CUSTOMER_ID-Completeness-Part...	Completeness	● 100%	Oct 7, 2022, 4:35 ...	100	0	CUSTOMER_ID	-	Party ID - Completeness	Party ID - Completeness

Managing the Data Quality Automation Process

Driven by Metadata Command Center and Data Governance and Catalog

isAutomated option on rule templates in Data Governance and Catalog	Data quality option in Metadata Command Center	Data quality automation option in Metadata Command Center	Result
Yes	Yes	Yes	Create rule occurrences for all data elements that are associated with glossary business assets.
Yes	Yes	No	• Does not create new rule occurrences for data elements or update an existing rule occurrence on data elements. • Does not affect the execution of the existing rule occurrences in Data Governance and Catalog
Yes	No	Not applicable	• Does not create any rule occurrences for data elements. • Data quality execution stops for existing rule occurrences that are associated with assets of a particular catalog source.
No	Yes	Yes	• Does not create rule occurrences for data elements. • Does not affect the execution of the existing rule occurrences in Data Governance and Catalog

Informatica®