Informatica

12 Aug, 2025

# Automated Data Lineage with CLAIRE and Metadata Rules


- Michael Sabotka, Principal Solutions Architect, IPS
- Ratish Neelakandan, Sr Principal Services Solutions Architect, IPS


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.

# Speakers

**Ratish Neelakandan**
Senior Principal, IPS

**Michael Sabotka**
Principal Solutions Architect, IPS

# Agenda

1. Why Data Lineage Matters
2. Lineage in CDGC
3. Lineage Decision Tree
4. CLAIRE Assisted Lineage
5. Use Cases
6. Demo

# Why Data Lineage Matters

Improved Data Transparency and Trust

Enhanced Regulatory Compliance

Efficient Impact Analysis

Accelerated Troubleshooting and Issue Resolution

Optimized Data Governance

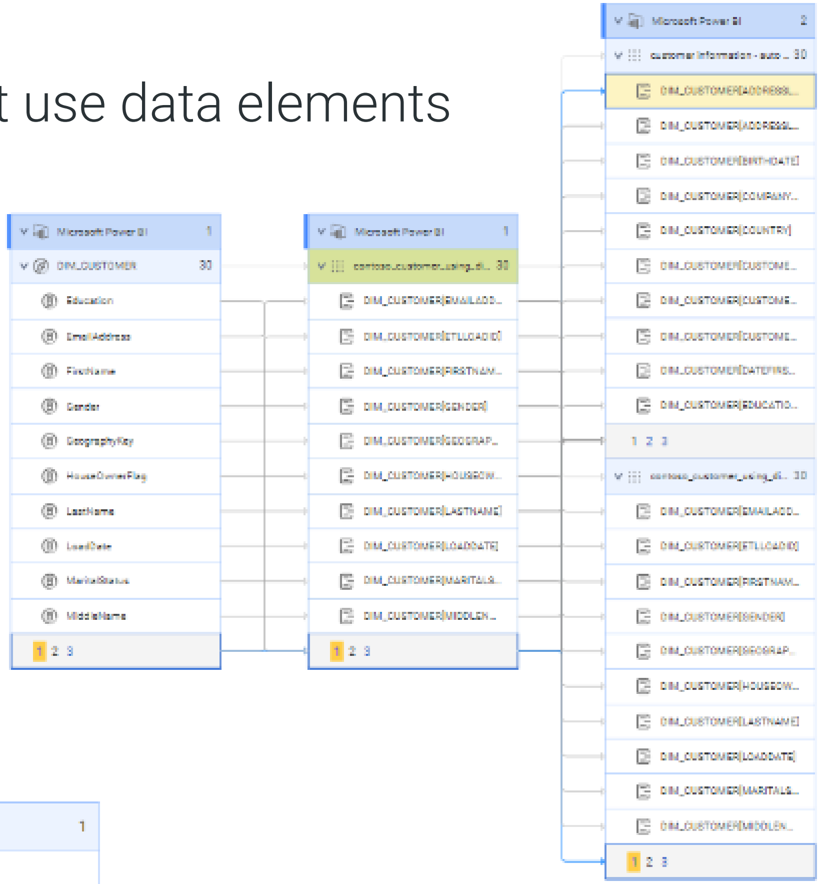
Informed Decision-Making

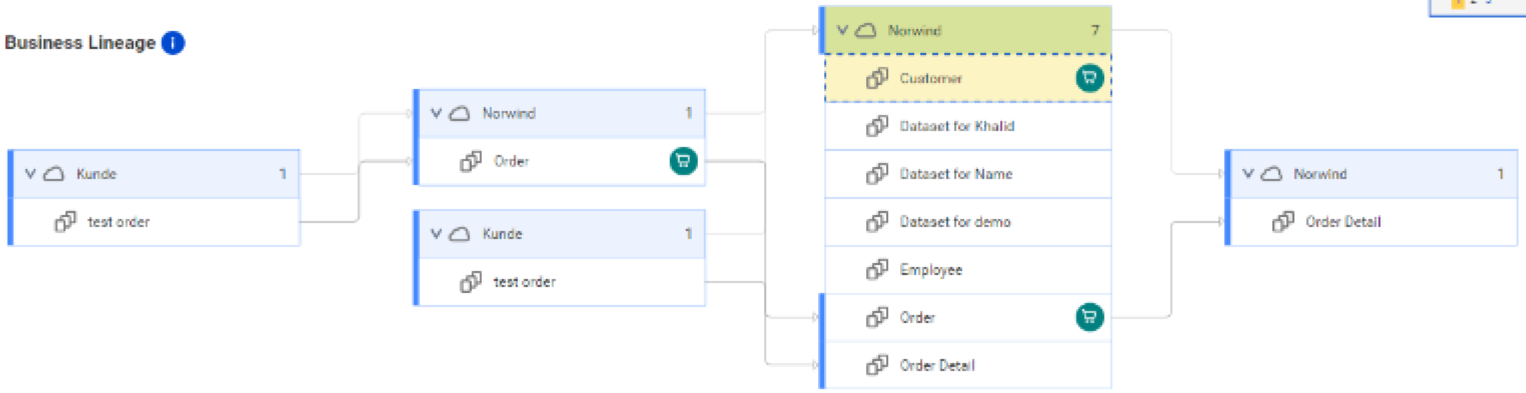
Facilitates Data Modernization and Migration

# Lineage in Cloud Data Governance

## In General Lineage presents

- Data flow across data containers, databases, and tables that use data elements
- Where and how data is used in a particular system
- Dependencies of the data
- Processes that the data undergoes

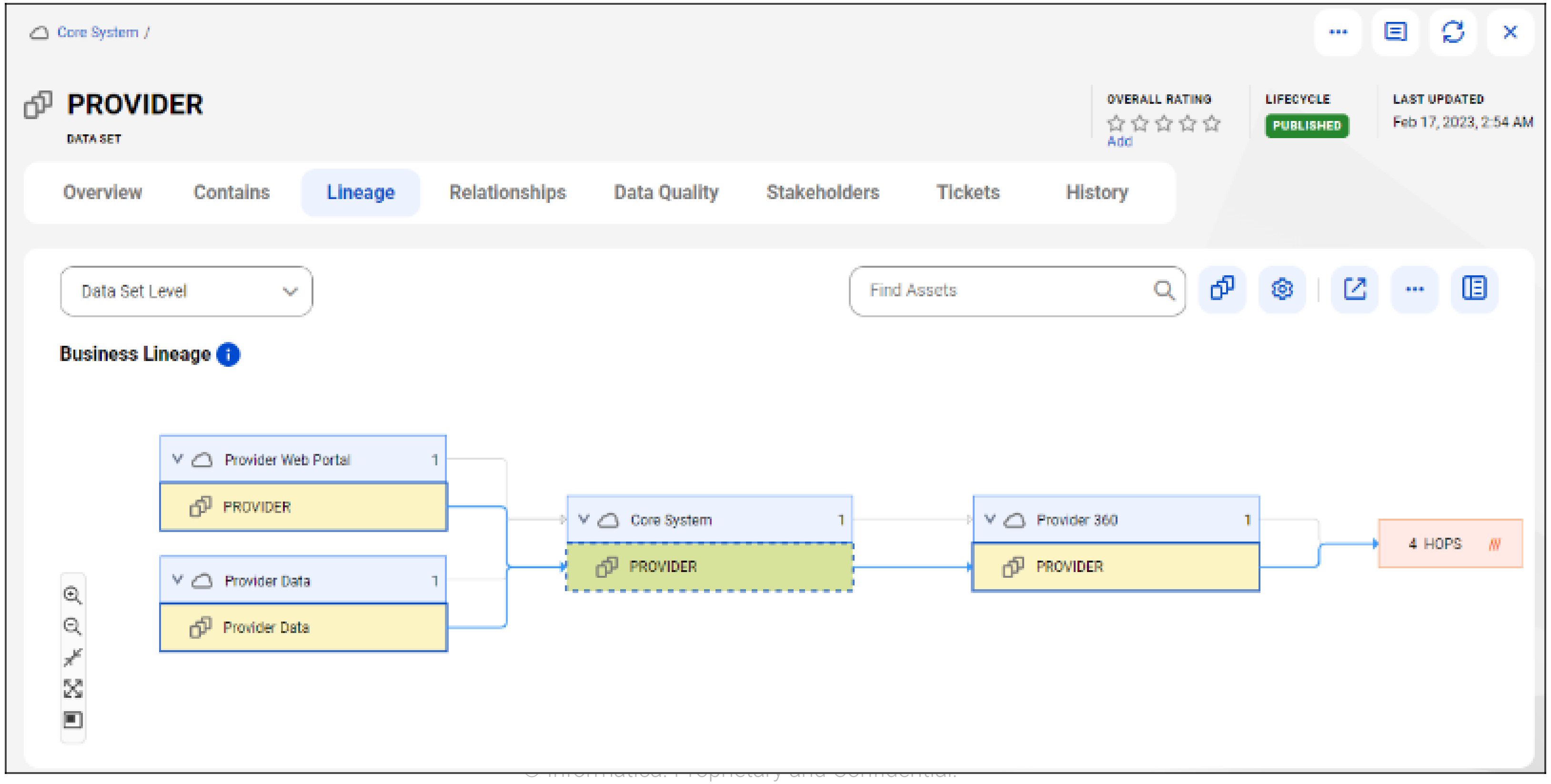




# Lineage in Cloud Data Governance

## Business Lineage

- Business lineage illustrates data flow and connections between data sets and system assets. It helps business analysts verify data sources and ensure data reliability, showing how data moves between data sets and their parent systems.

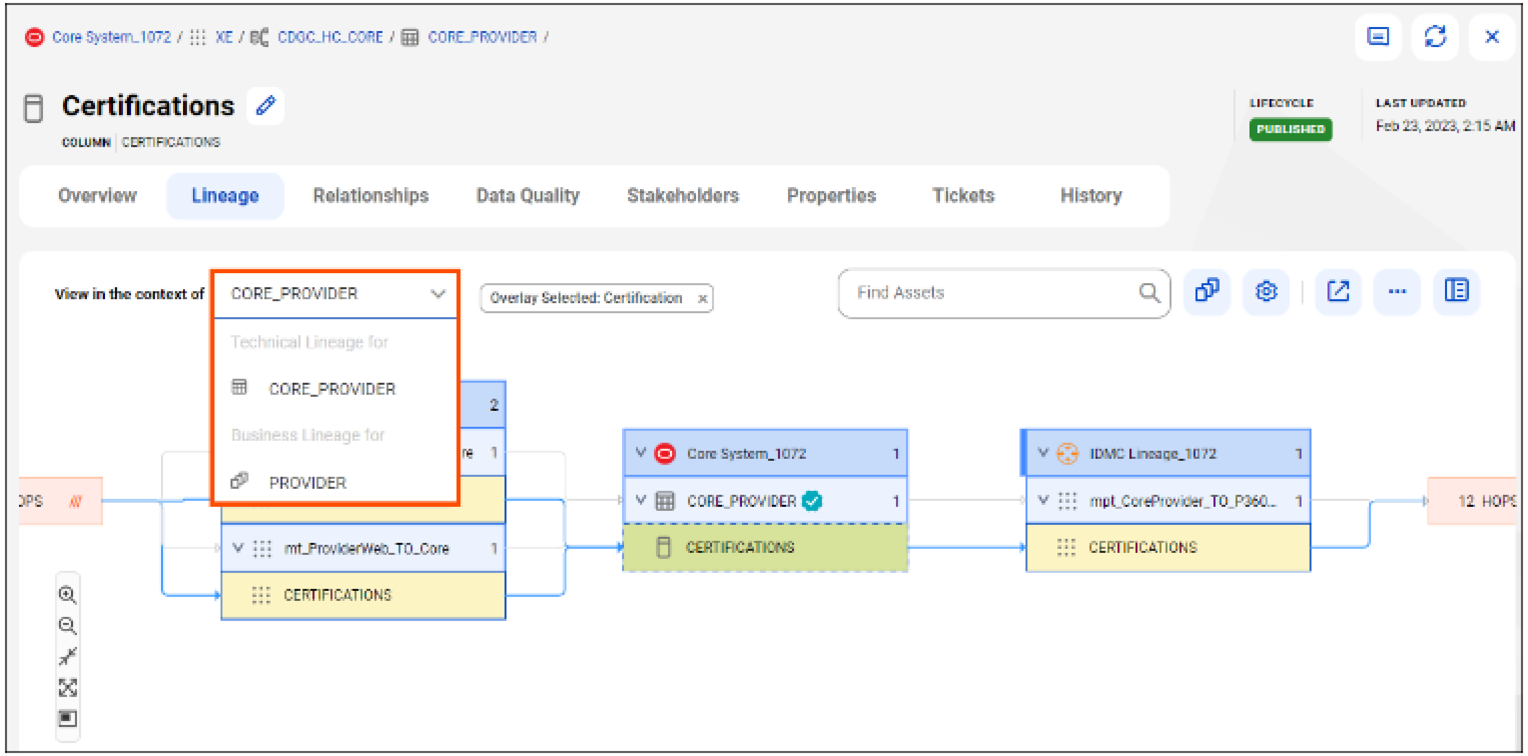



Informatica

# Lineage in Cloud Data Governance

## Technical Lineage

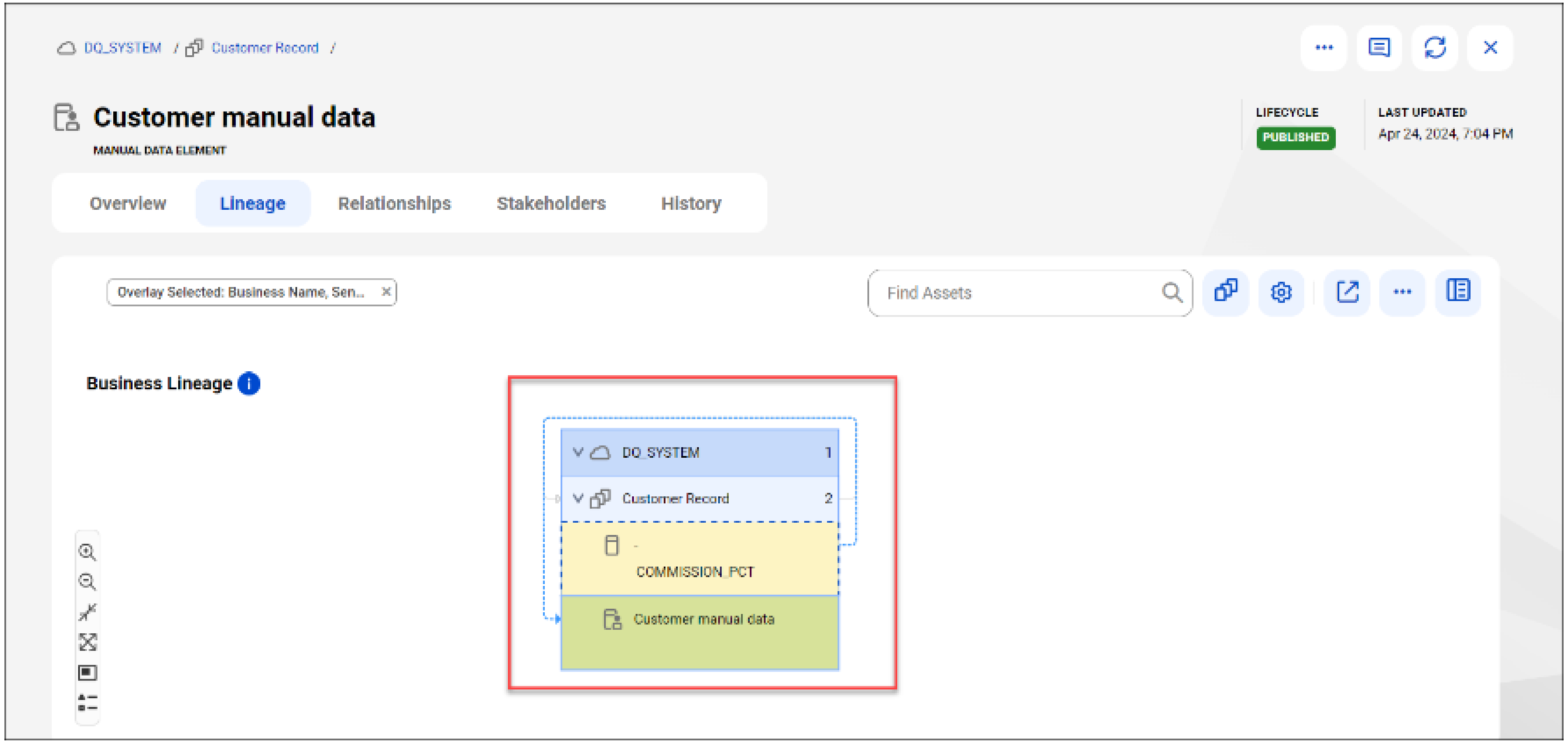
- Technical lineage provides a detailed view of a data element's journey, from source to destination, including transformations. Data architects use it to examine data at the granular level. If linked to a business data set, both technical and business lineages can be viewed for comprehensive insights.



# Lineage in Cloud Data Governance

## Manual Lineage

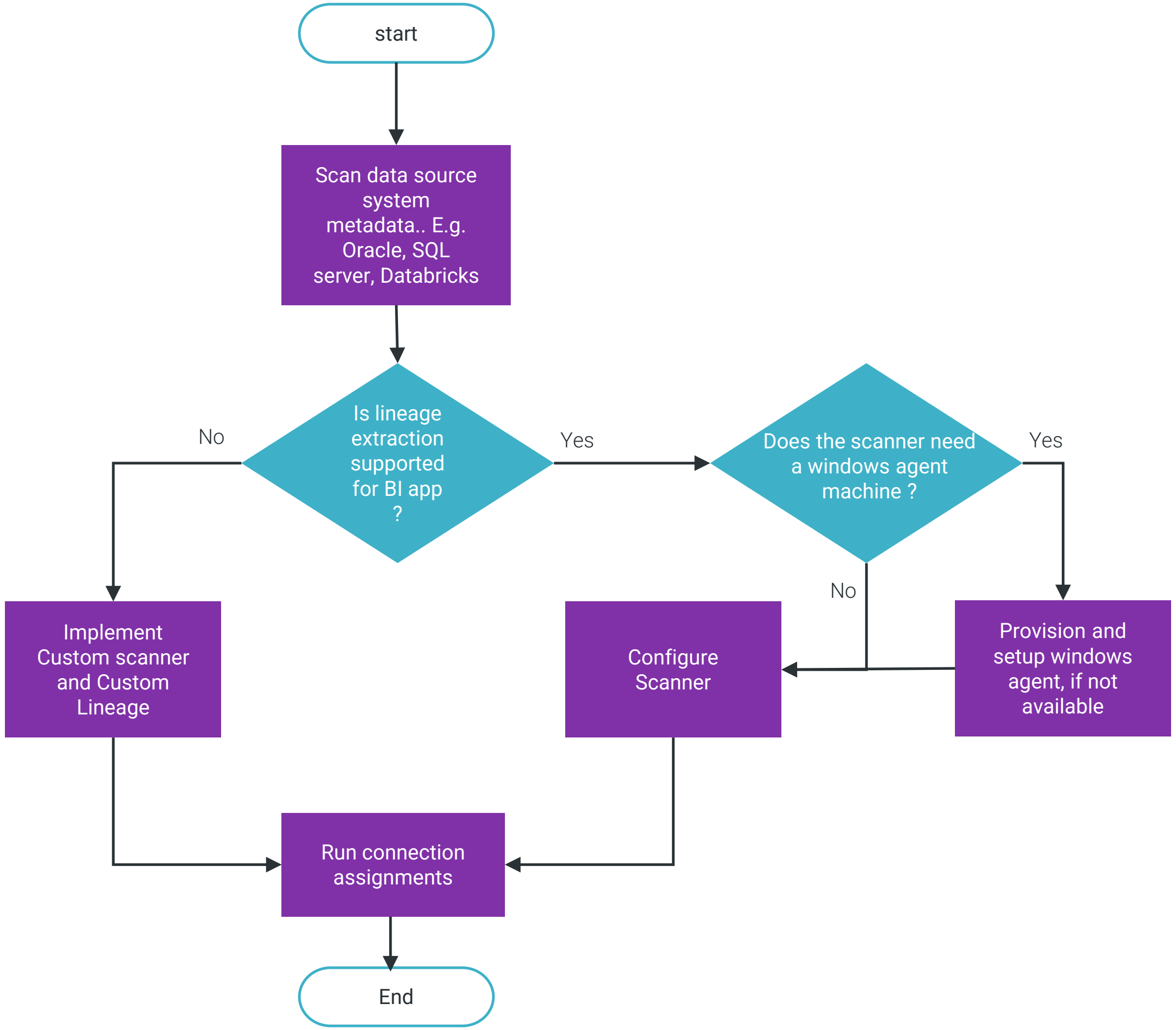
- If you want to establish a connection between a manual data element that you created and other business and technical assets for governance purposes, use the Create Manual Lineage option on the Relationships tab of an asset. Type is always “*is in Manual Lineage with*”



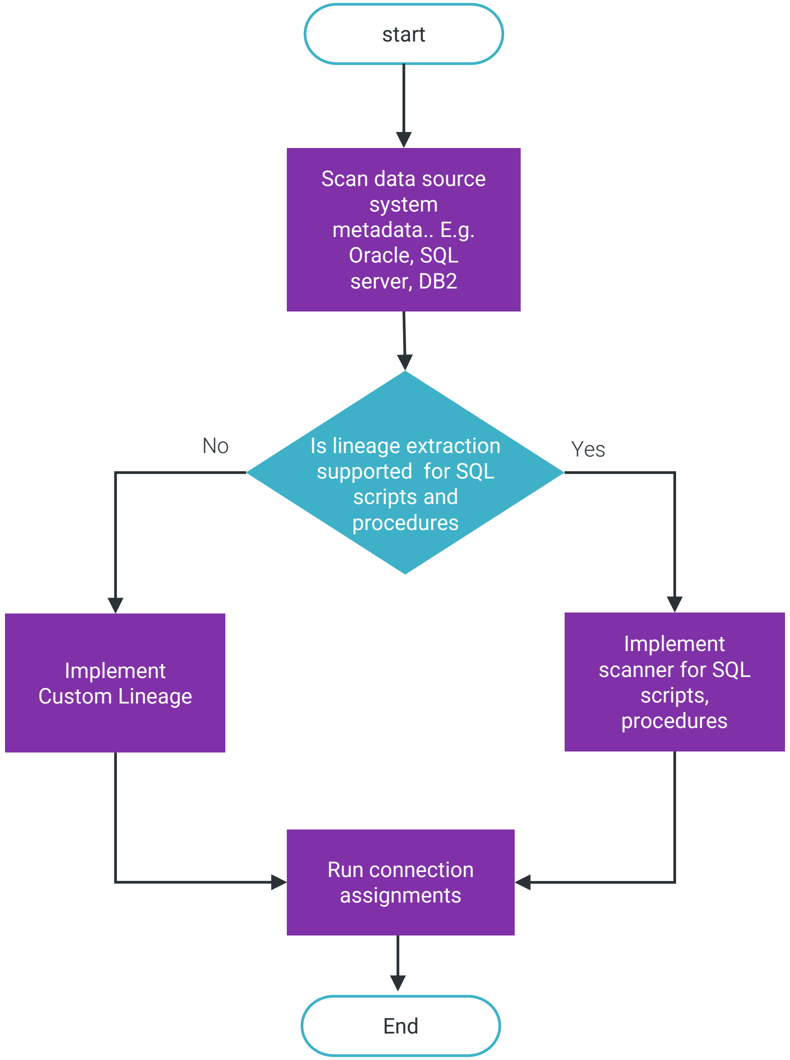

Informatica

# Scanner Decision Chart - Lineage

## Metadata/Lineage Scans (from BI Sources)

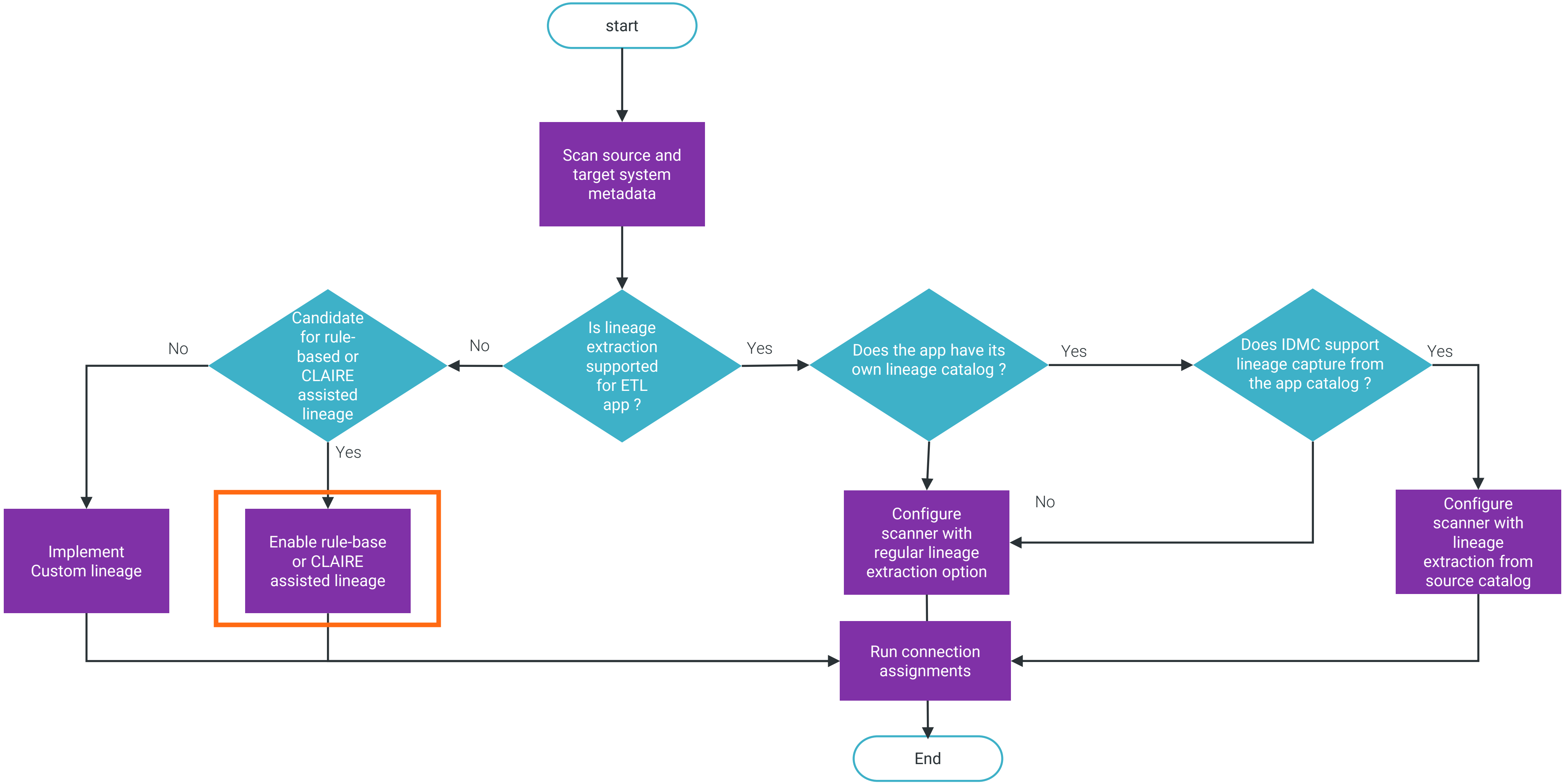



## Lineage Scans (from SQL Scripts)



# Scanner Decision Chart - Lineage

Lineage Scans ( from Data Integration Applications)



# CLAIRE Lineage Automation

**Challenge:** Technological or security limits may prevent lineage extraction after metadata extraction.

**Solution:** Use Metadata Command Center to link source and target catalog sources and construct detailed data lineage.

**Scope:** Link specific catalogs and restrict lineage to certain schemas or data object subsets.

**Lineage Generation:**

- **Automated:** Powered by CLAIRE, which recommends catalog source links with confidence scores. Links above a threshold are auto-accepted; others can be manually accepted.
- **Rule-Based:** Define custom rules using name-based matching or expression-based inclusion rules (with prefixes, suffixes, attributes, etc.) to generate lineage.

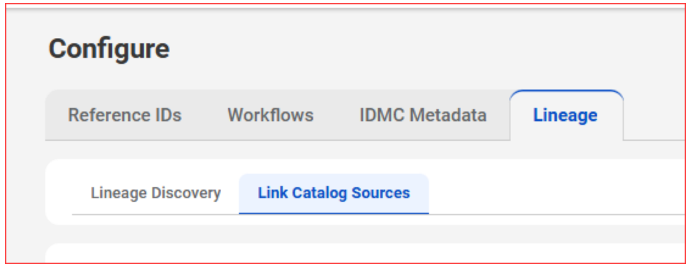
**Governance and Curation:**

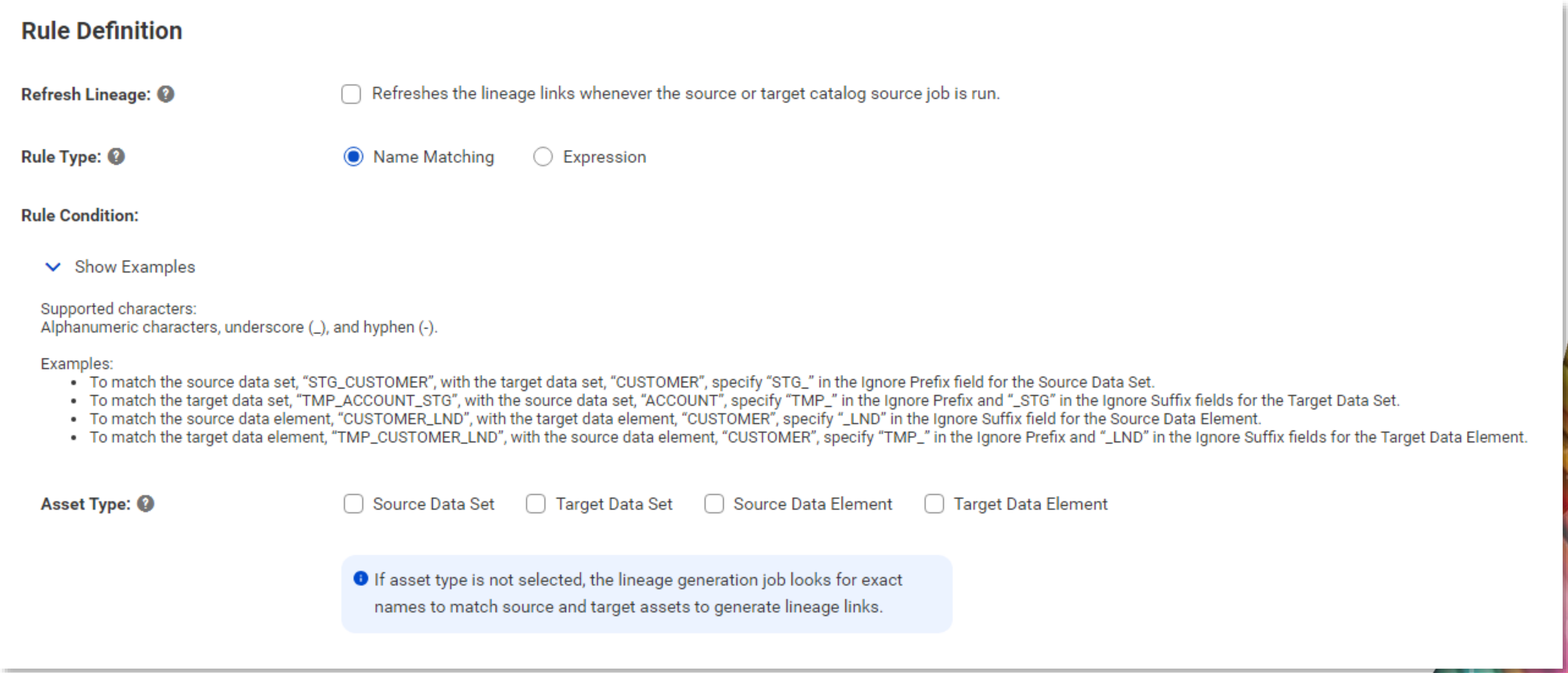
- Auto-accepted and manual links appear in the Catalog Source Links page.
- Stakeholders can accept or reject lineage links; rejected links can later be accepted.
- View lineage details on the asset's Lineage tab.

# CLAIRE Lineage Automation

## Rule-Based linking

- Link catalog sources to generate data lineage using specified rules. Incomplete lineage due to source data constraints can be enhanced in Metadata Command Center by linking sources with criteria like name-based matching or inclusion rules



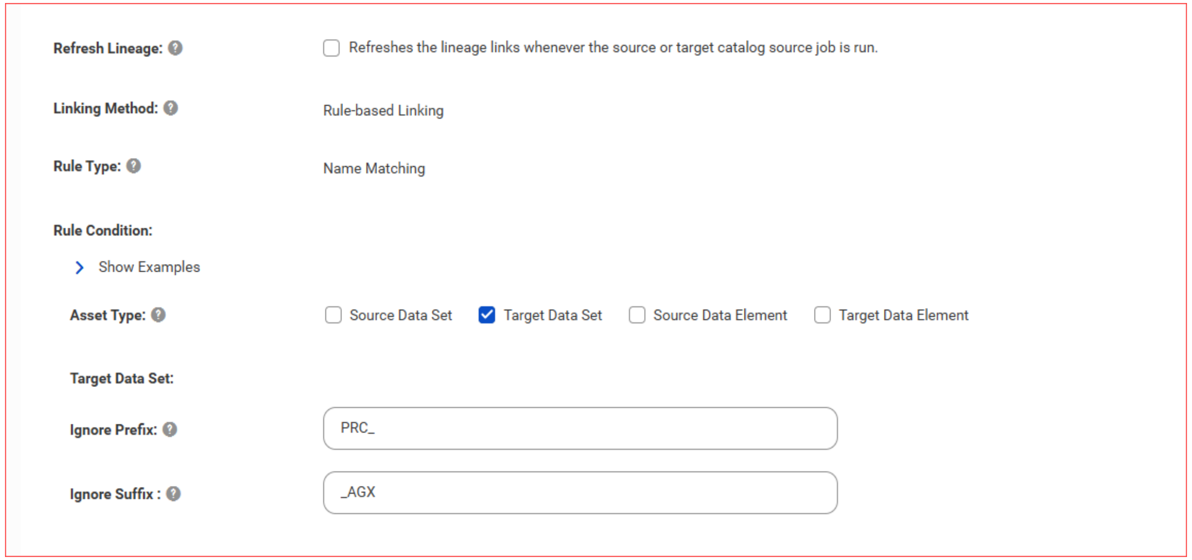




Informatica

# Rules

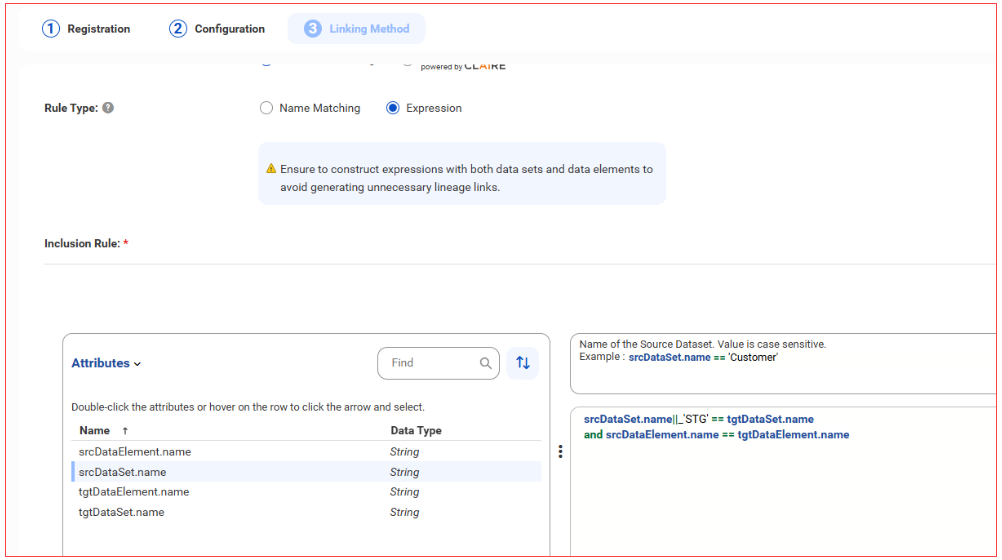
**Name Matching** : Ignores specified prefixes and suffixes of an asset name and matches the rest of the asset name to generate catalog source links.

e.g., source CUSTOMER, Target: PRC_CUSTOMER_AGX




**Expression** : Constructs an inclusion rule using expressions. Use a combination of attributes, operators, functions, or comments to build an inclusion rule.

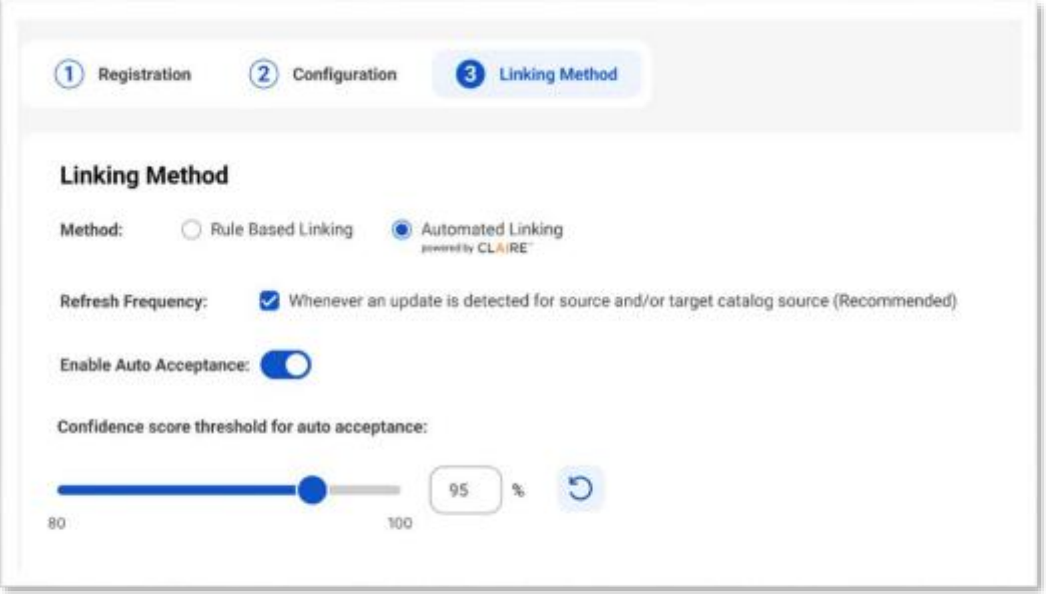
e.g., source: CUSTOMER, Target: CUSTOMER_STG



# CLAIRE Lineage Automation

## CLAIRE powered linking

- Added capability for "Linked Lineage"
- Uses trained AI models to automatically detect and build lineage between user provided source and target
- Lineage is recommended or auto-accepted based on user defined option / threshold
- Allows user curation (Accept/Reject) of generated lineages
- Supported for Relational Sources



# CLAIRE Lineage – Use Case

## Lineage for Data Replication or Mass Ingestion

**A major medical devices customer consolidated data from multiple databases into their BigQuery storage using data replication software. The customer needed to document the complete data lineage—from the original source through every stage of processing—up to its final consumption in Power BI.**

The customer utilized Informatica’s native scanning capabilities to perform scans across BigQuery and PowerBI environments.

Since the data replication software was not natively scannable by Cloud Data Governance and Catalog (CDGC), the customer leveraged CDGC’s rule-based inferred lineage functionality to establish and execute source-to-target lineage rules. Consequently, Informatica automatically inferred one-to-one lineage relationships between source and target tables, both at the dataset and data element levels, to capture end to end lineage

# CLAIRE Lineage – Use Case

## Lineage for Snowflake Data Vault

**A prominent financial institution established a Snowflake data warehouse utilizing scripts for data movement across layers. The organization required comprehensive data lineage documentation spanning multiple vault layers, tracing the data flow from the landing zone through to the data marts.**

The customer leveraged Informatica's native scanning capabilities to capture Snowflake metadata through the Snowflake native scanner.

Data movement across layers was orchestrated by reading from source-layer views and writing to corresponding target tables in the subsequent layer. Informatica's CLAIRE-assisted lineage feature automatically inferred the data lineage by analyzing the naming conventions of the views and their related target tables, thereby generating comprehensive end-to-end lineage.

Informatica

DEMO

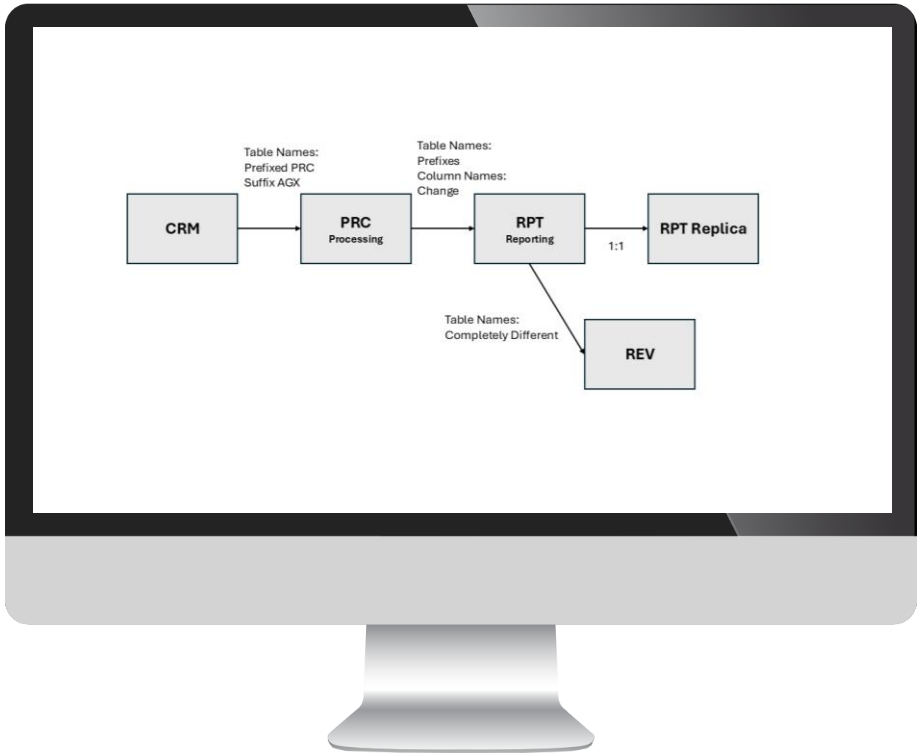


Table Names:
Prefixed PRC
Suffix AGX
Table Names:
Prefixes
Column Names:
Change
CRM
PRC
Processing
RPT
Reporting
1:1
RPT Replica
Table Names:
Completely Different
REV

Informatica

# Thank You

Where data & AI come to LIFE

# Lineage in Cloud Data Governance

## Connection-aware scanners (Source systems)

- AWS Glue
- Databricks
- dbt
- Google BigQuery SQL Script
- IBM Db2 for LUW Script
- IBM Infosphere DataStage
- Informatica Intelligent Cloud Services
- Informatica PowerCenter
- Microsoft Azure Analysis Services
- Microsoft Azure Data Factory
- Microsoft Azure SQL Server Script
- Microsoft Azure Synapse Script
- Microsoft Power BI
- Microsoft SQL Server Analysis Services
- Microsoft SQL Server Integration Services
- Microsoft SQL Server Reporting Services
- Microsoft SQL Server Script
- Oracle Cloud Infrastructure GoldenGate
- Oracle PL/SQL Script
- Snowflake SQL Script
- Tableau
- Teradata BTEQ Script
- Teradata FastExport Script
- Teradata FastLoad Script
- Teradata MultiLoad Script