Informatica

22 Oct, 2024

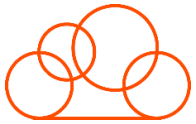
# Informatica CLAIRE: AI Co-Pilot in IDMC


- Akshada Sahebrao Sable, Solutions Architect, CSA
- Rashmi P, Senior Solutions Architect, CSA


Where data & AI come to LIFE

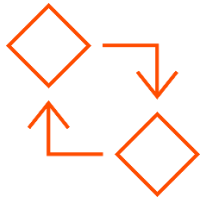
# Housekeeping Tips

- 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.



**Bootstrap trial and POC Customers**

**Enriched Customer Onboarding experience**



**Product Learning Paths and Weekly Expert Sessions**

**Informatica Concierge**

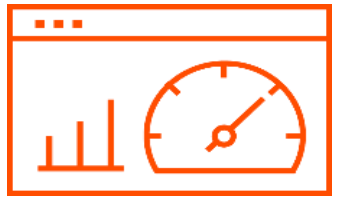
**Tailored training and content recommendations**





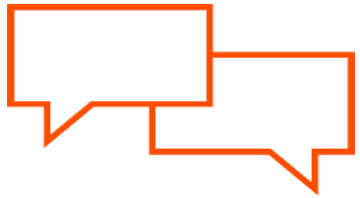
Informatica

# More Information



## Success Portal

https://success.informatica.com

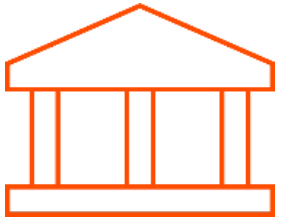


## Communities & Support

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

22 Oct, 2024

# Informatica CLAIRE: AI Co-Pilot in IDMC


- Akshada Sahebrao Sable, Solutions Architect, CSA
- Rashmi P, Senior Solutions Architect, CSA


Where data & AI come to LIFE

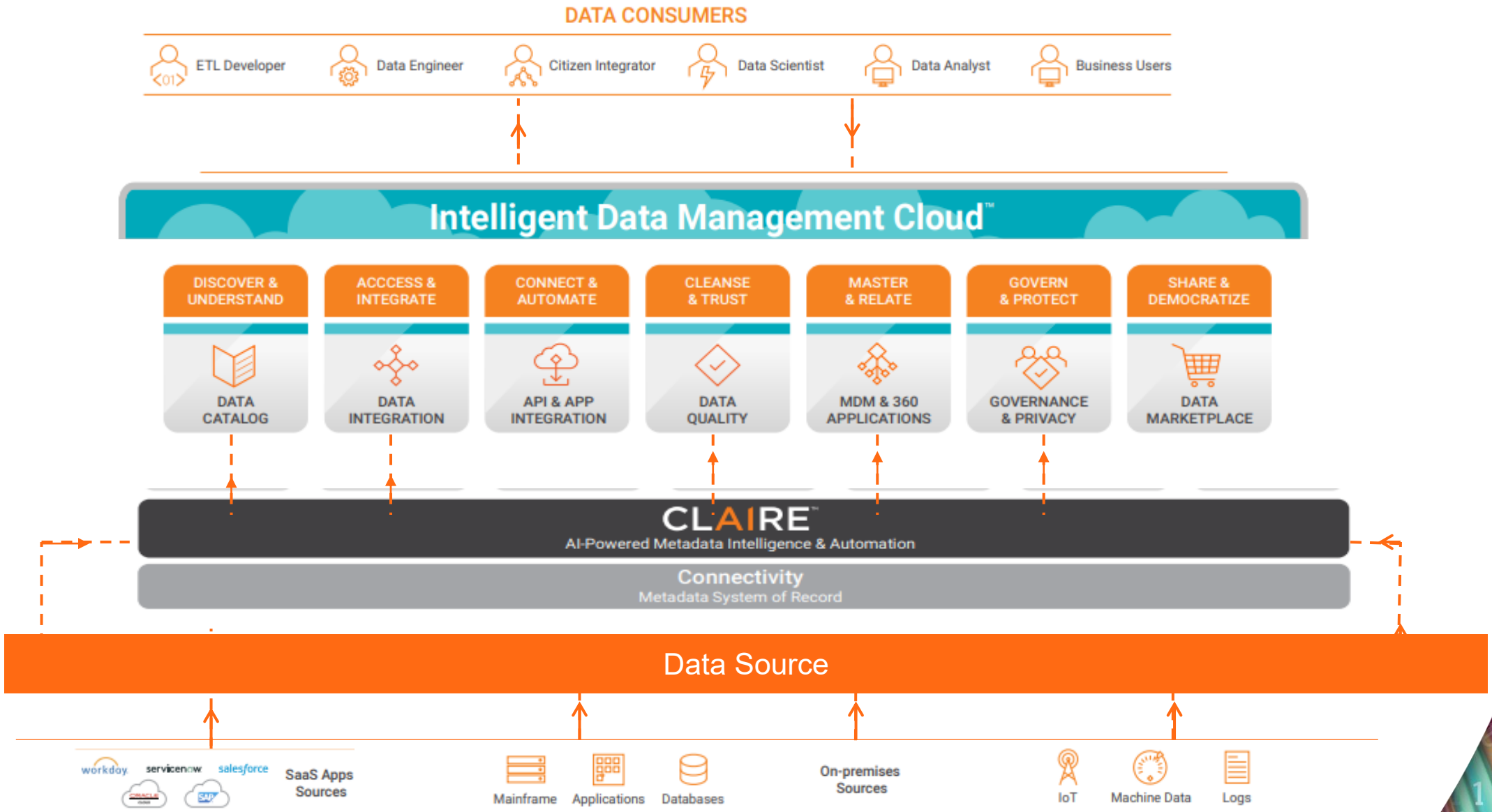
# Agenda

- Overview of Informatica CLAIRE: The AI Copilot in the Intelligent Data Management Cloud
- CLAIRE as Data Cataloging Copilot
- CLAIRE as DataOps Copilot
- CLAIRE as Analytics Copilot
- CLAIRE as Data Governance and Data Quality Copilot
- CLAIRE GPT
- Q&A

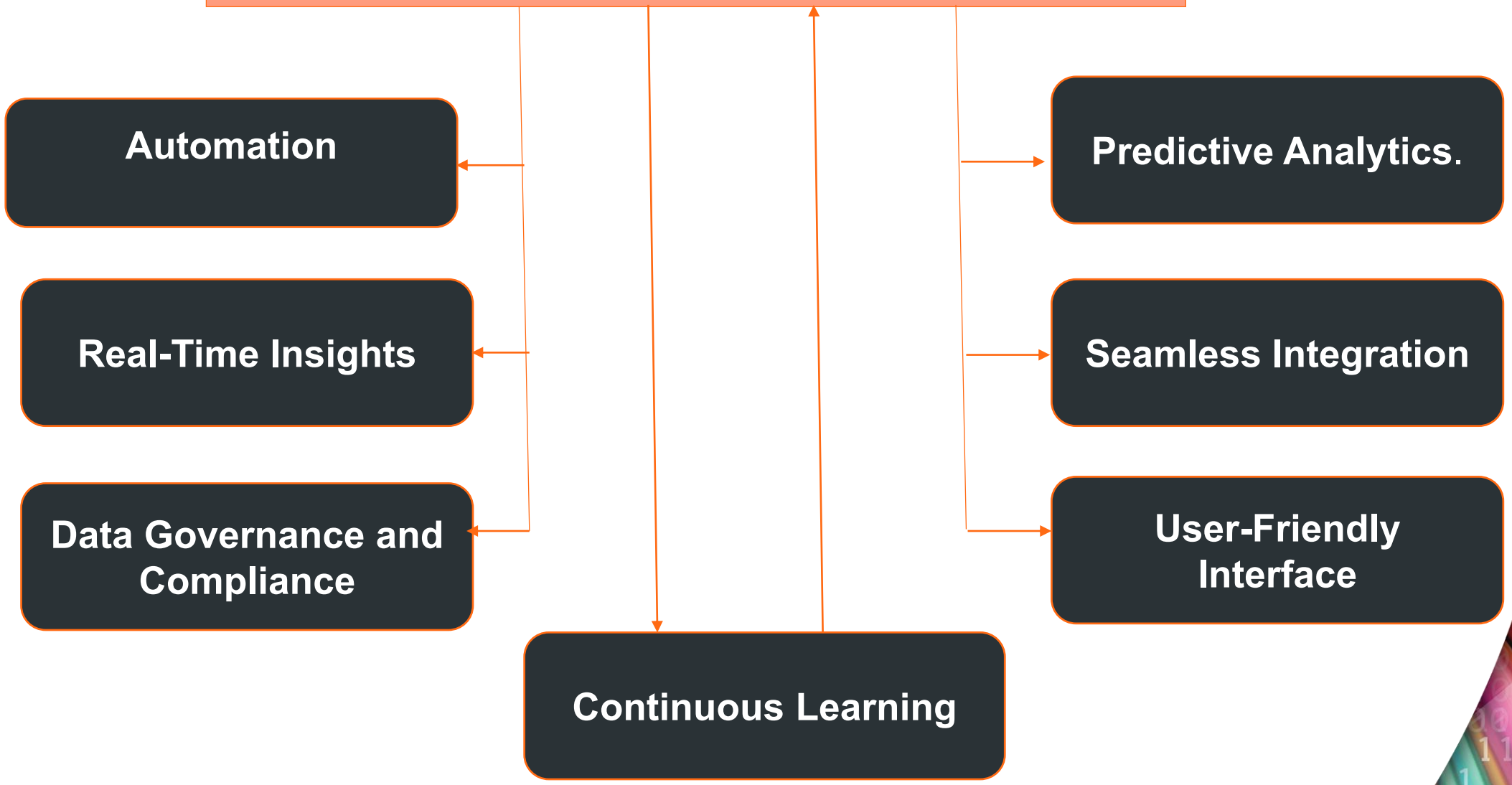
# Overview on AI Co-Pilot in IDMC

The AI Co-Pilot in IDMC serves as an intelligent assistant designed to enhance data management processes by automating routine tasks, providing insightful analytics, and facilitating seamless integration across platforms. This innovative tool empowers users to make data-driven decisions more efficiently, ultimately leading to improved productivity and operational effectiveness within organizations. By leveraging advanced machine learning algorithms, the AI Co-Pilot transforms the way businesses interact with their data, ensuring a more agile and responsive approach to data management challenges.

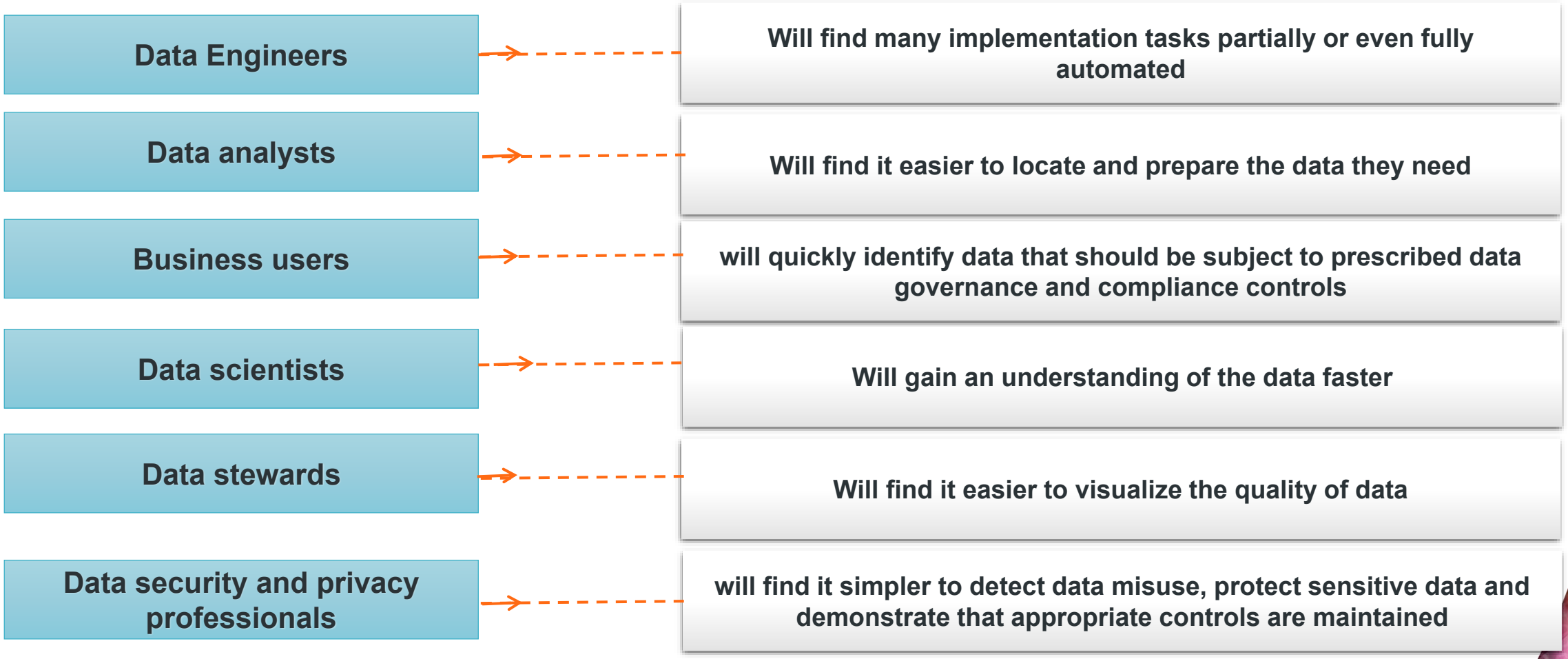
# CLAIRE in IDMC



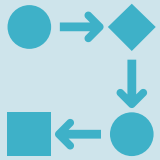
# Key Features of Co-Pilot



# CLAIRE acts as a copilot for a broad spectrum of users:



# CLAIRE as Data Cataloging Copilot



Discovering and understanding your data is the first step in any data-driven initiative.

**CLAIRE** copilots an ML-based discovery engine to scan and catalog data assets across the enterprise. An intelligent data catalog copiloted by **CLAIRE** can help data scientists, analysts and engineers find and recommend the necessary data, significantly reducing the time spent in data discovery and preparation. Moreover, it facilitates the integration of business context with technical metadata, enhancing the relevance and usability of data within organizational workflows. By continuously learning from user interactions and feedback, Copilot refines its algorithms to provide more precise recommendations and insights.


**Organizations struggle with:**

1. **Finding relevant data** across siloed systems, wasting time and effort.

2. **Inconsistent data quality and governance**, leading to compliance risks.

3. **Poor collaboration**, with teams unable to access or share data efficiently.

4. **Manual, time-consuming processes**, reducing focus on strategic work.

5. **Delayed decision-making**, missing opportunities to act quickly on insights.

1

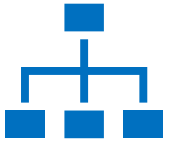
## Smart Tagging

Streamlines data field discovery and labelling with automated tagging

2

## Enhanced Relationship Discovery

Automates dataset relationships, improving efficiency and user interaction



3

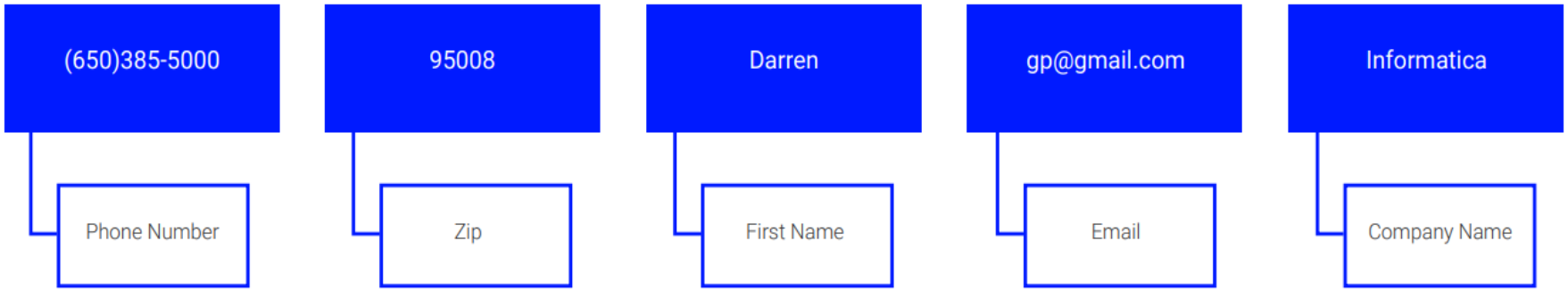
## Intelligent Data Similarity

Leverage ML for data similarity , aiding efficiency and integration

# Intelligent Domain Discovery with Tags : Smart Tagging

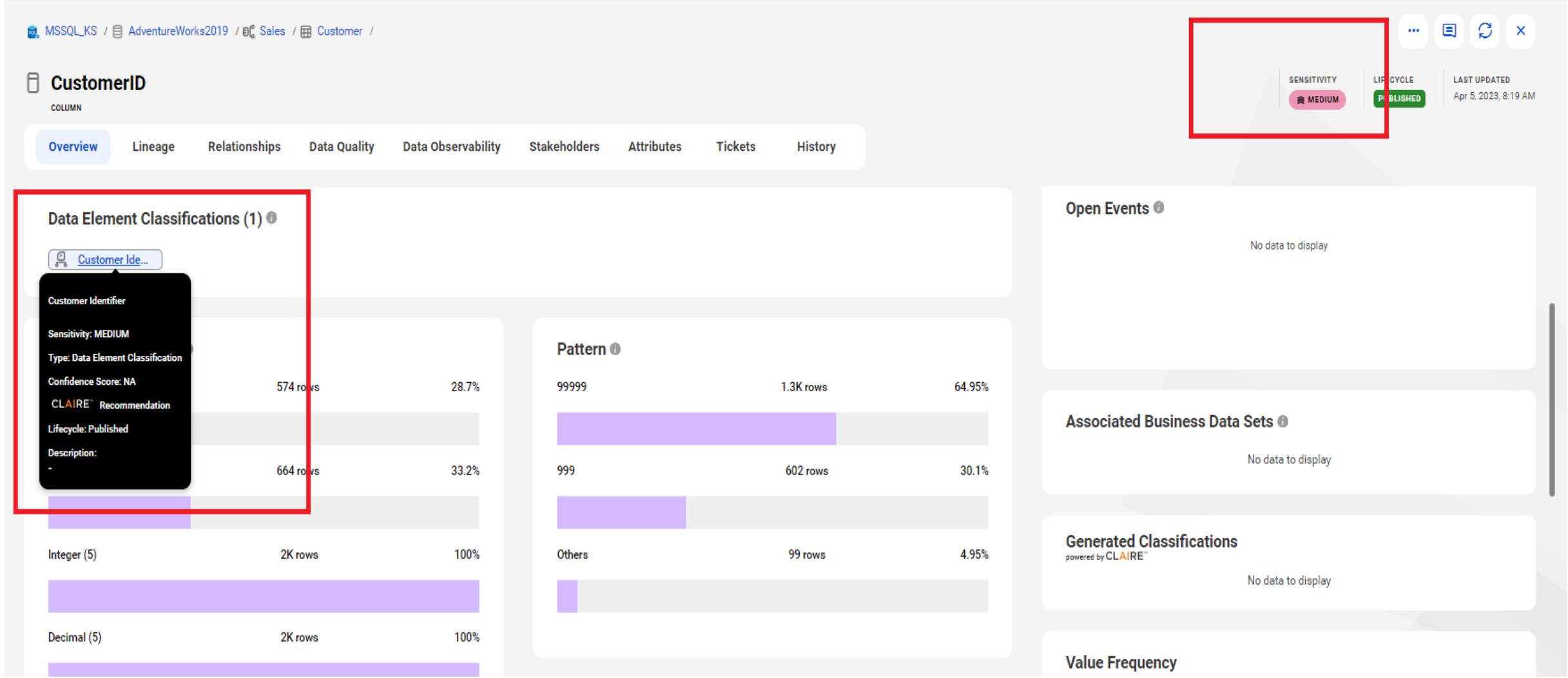
Smart tagging is supported by CDGC's ability to auto-tag columns and fields using inference rules and machine learning algorithms. Users can further customize classifications by renaming or adding conditions to the default rules, refining tags to suit specific business contexts.

With CLAIRE as your copilot, you can automatically classify data fields by applying semantic labels to each column. These semantic labels are called data domains.

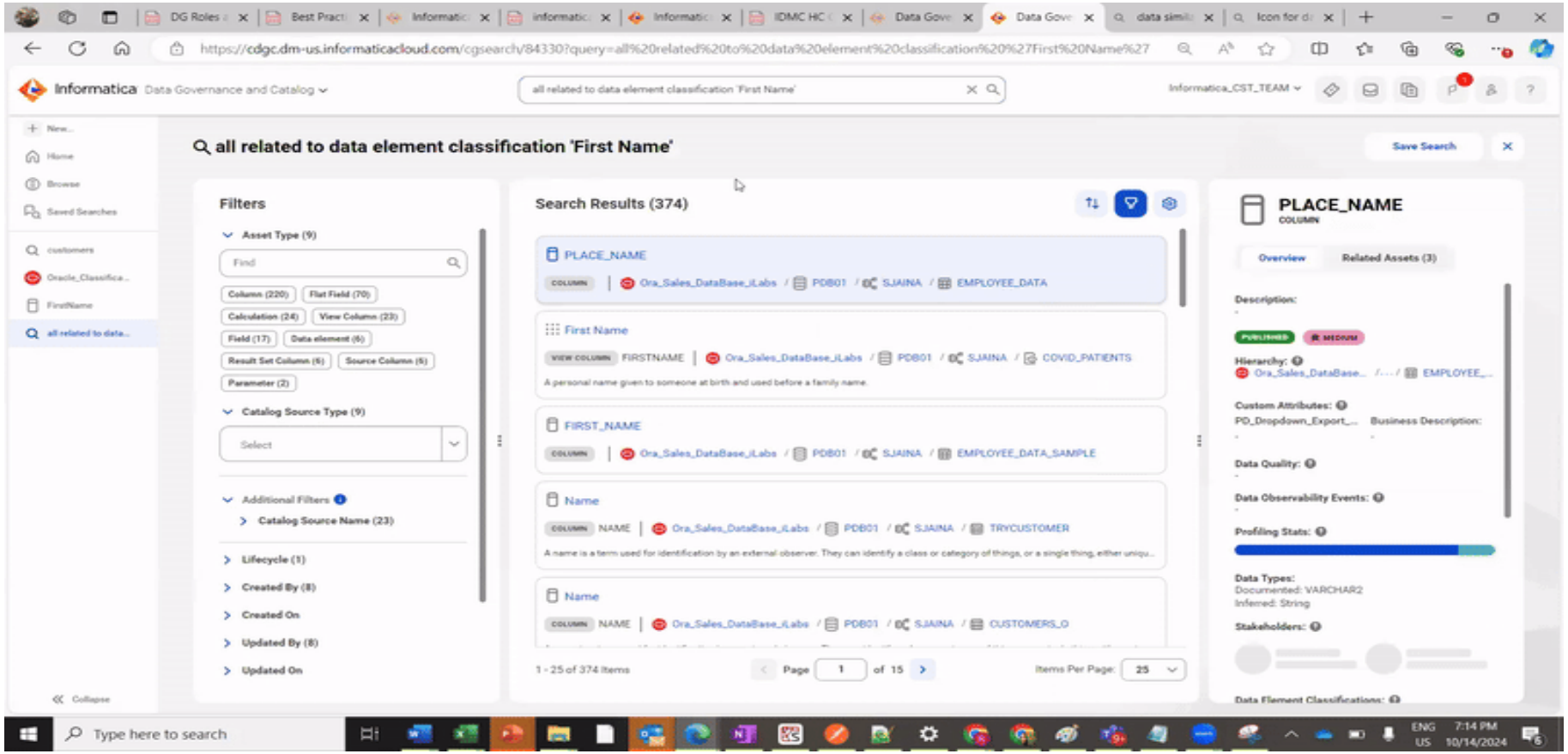


# Intelligent Data Element Classification

Data element classification labels and categorizes information contained in data elements based on the metadata extracted from source systems and the facts collected as the result of data

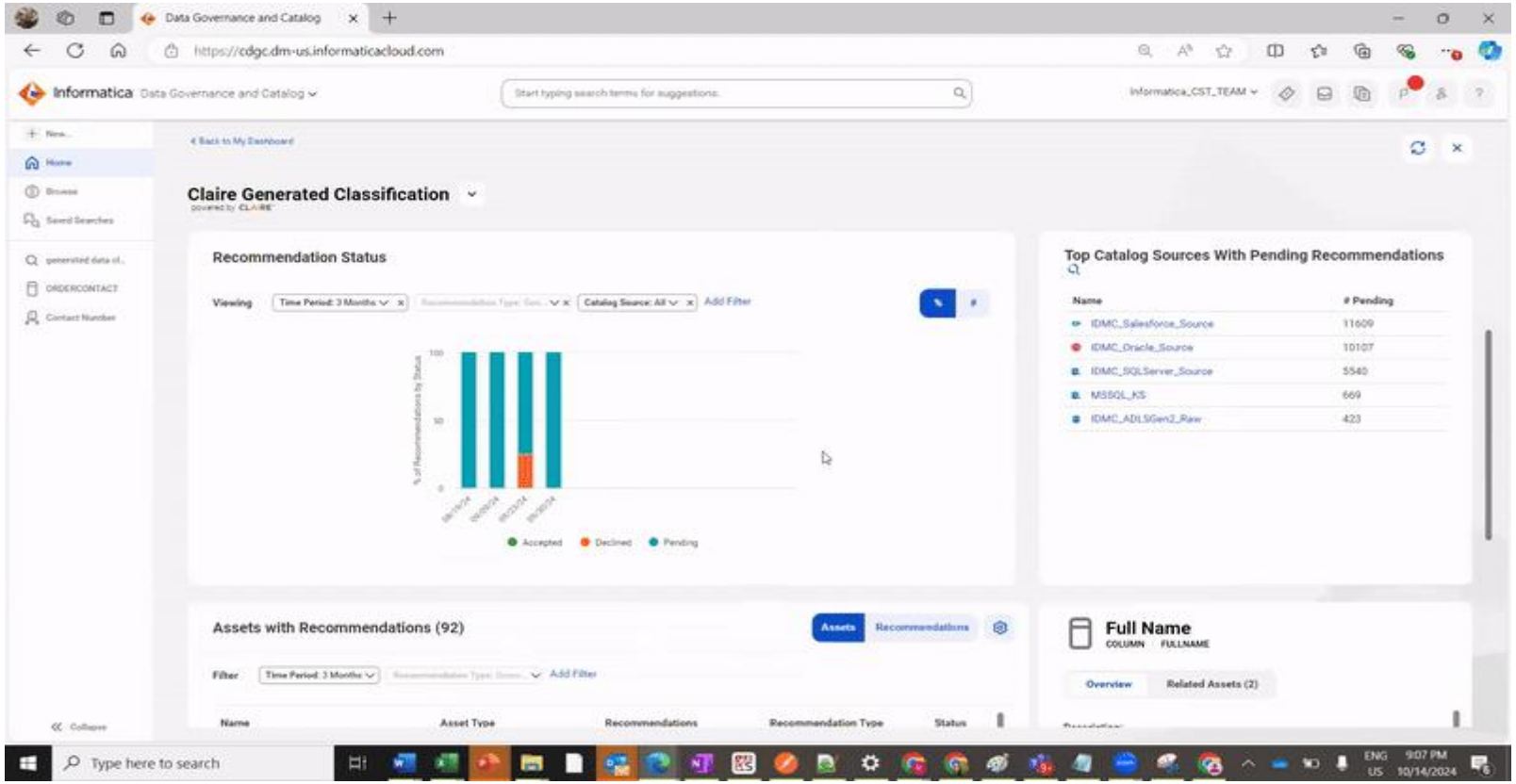


# Intelligent Search - Data Classification



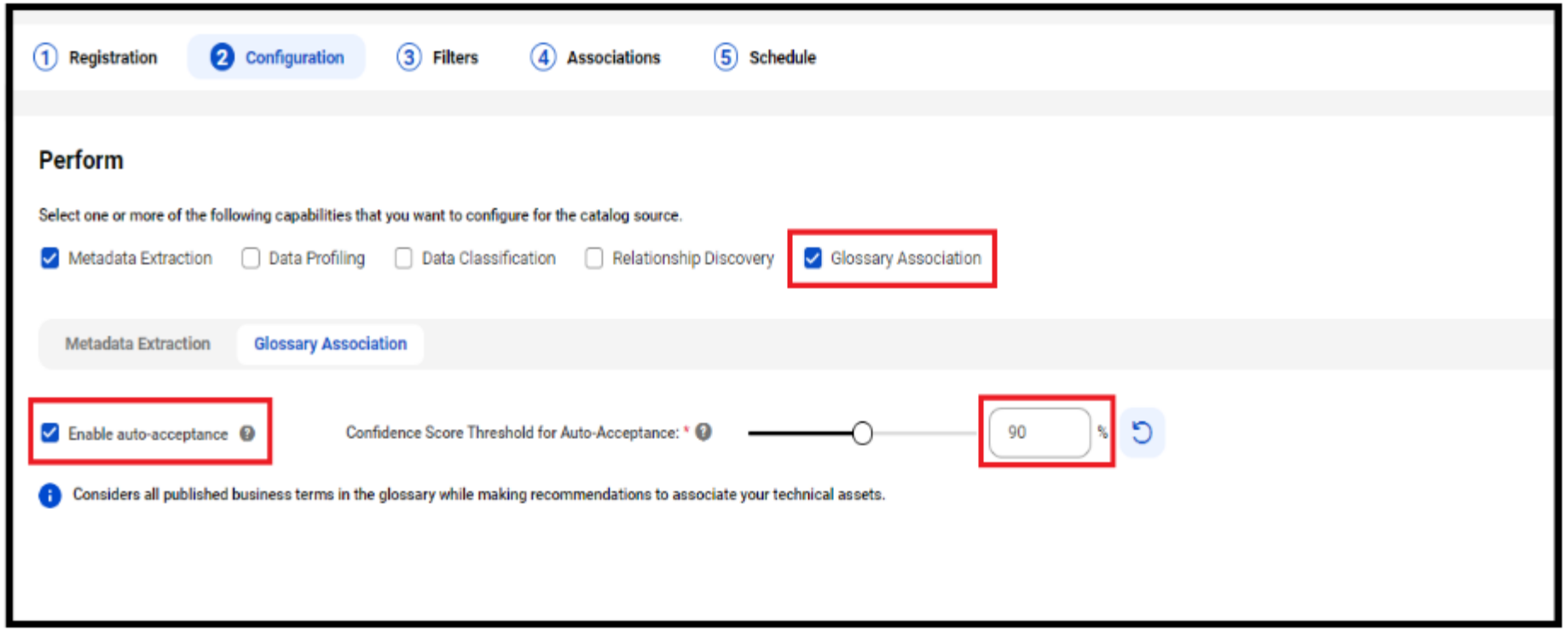
# Claire - Generated Classification

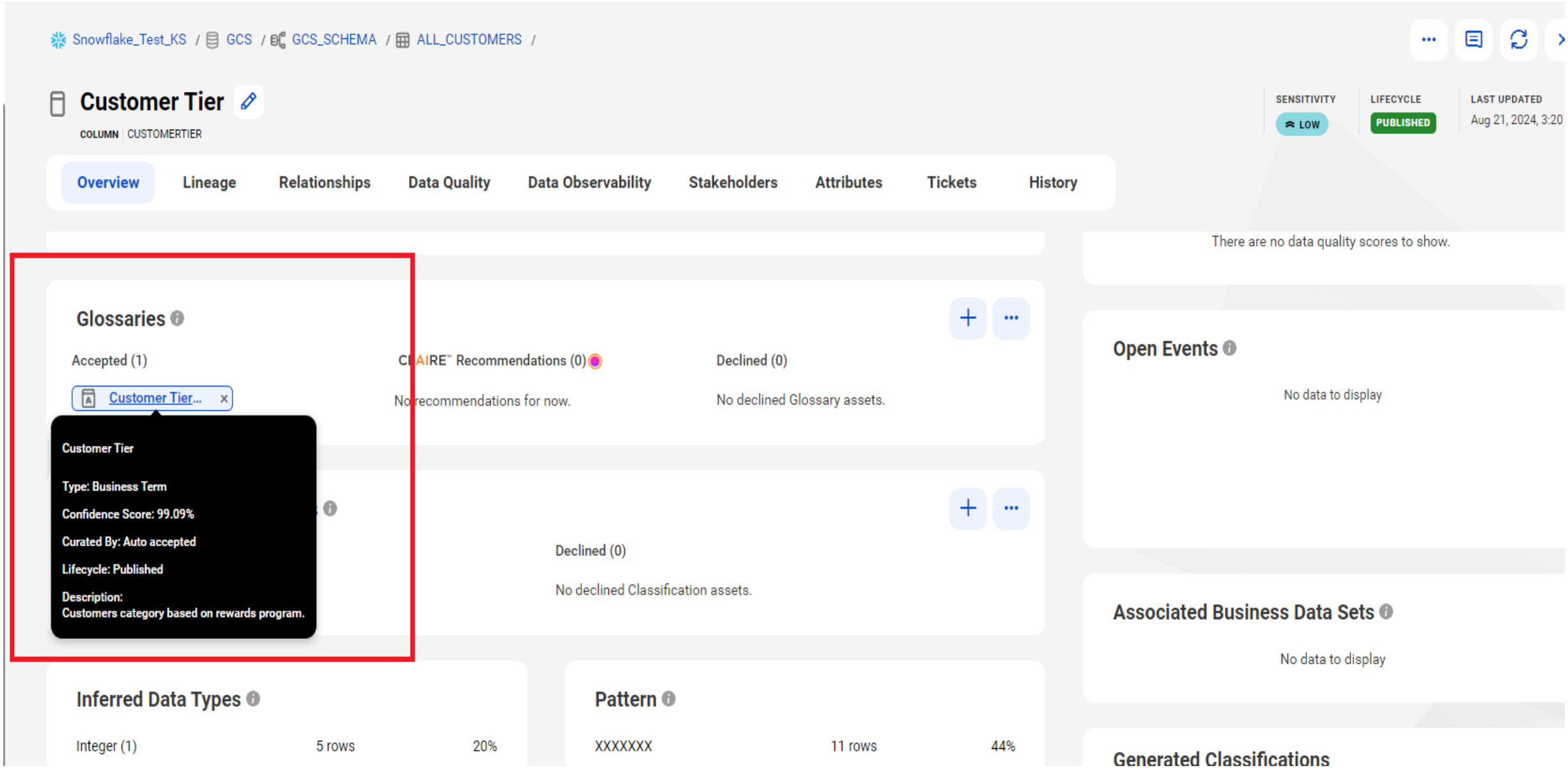
This classification is powered by CLAIRE. When you select it, CLAIRE automatically generates data classifications for the data elements. When you use data classification rules, you choose from predefined or custom data classifications. When you use generated data classifications, CLAIRE uses the nomenclature of technical data assets to generate classifications. To generate potential classification labels based on asset names, CLAIRE uses an embedded dictionary. Once you promote a generated data classification, Data Governance and Catalog remembers the promoted data classification and automatically applies it in future runs.



# Intelligent Glossary Association

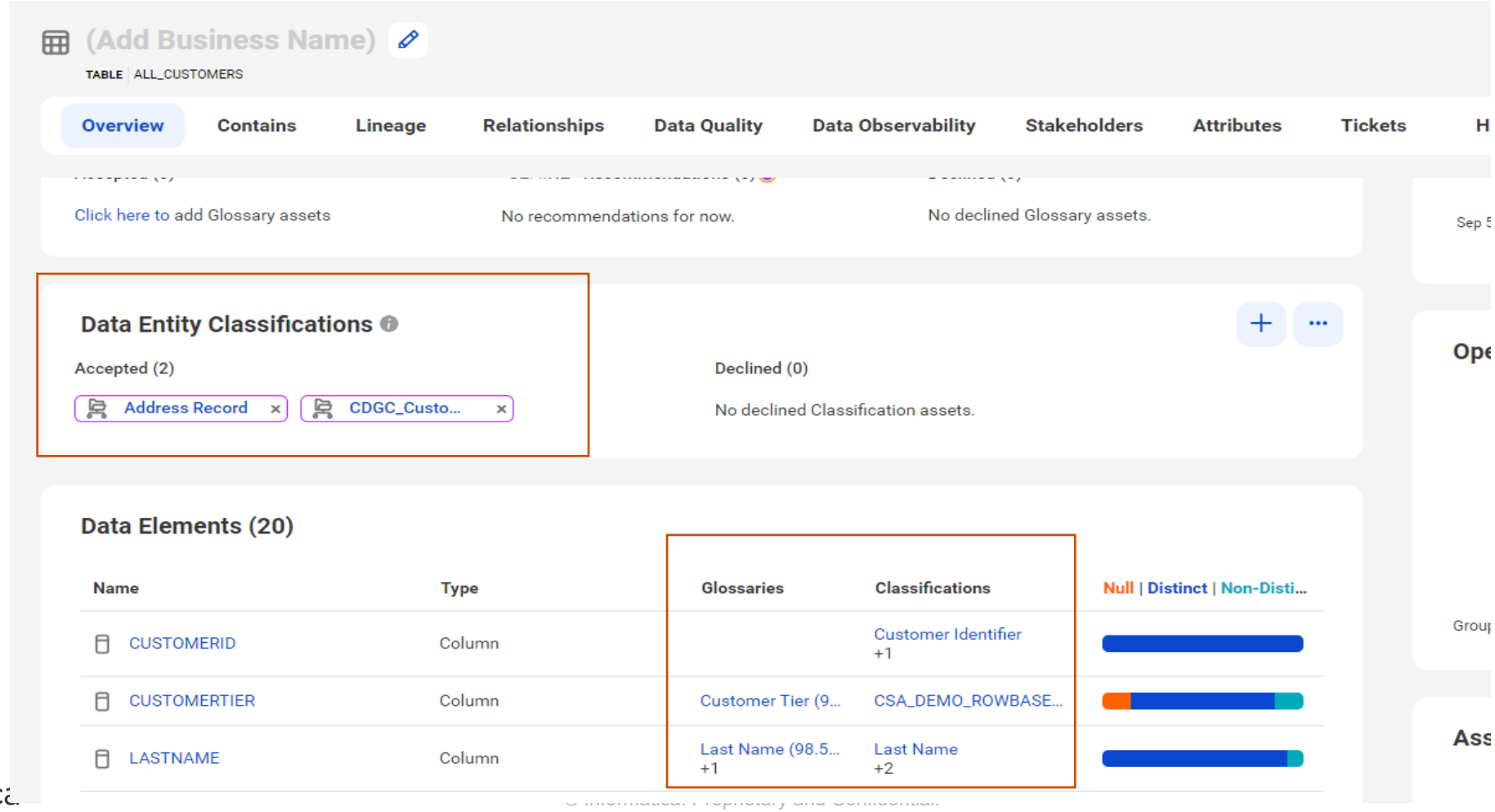
**CLAIRE engine** within Informatica's Cloud Data Governance and Catalog (CDGC) plays a critical role in automating the process of associating glossary terms with metadata elements.




Snowflake_Test_KS / GCS / GCS_SCHEMA / ALL_CUSTOMERS /
Customer Tier
COLUMN CUSTOMERTIER
SENSITIVITY
LOW
LIFECYCLE
PUBLISHED
LAST UPDATED
Aug 21, 2024, 3:20
Overview
Lineage
Relationships
Data Quality
Data Observability
Stakeholders
Attributes
Tickets
History
There are no data quality scores to show.
Glossaries
Accepted (1)
Customer Tier...
CLAIRE™ Recommendations (0)
No recommendations for now.
Declined (0)
No declined Glossary assets.
Customer Tier
Type: Business Term
Confidence Score: 99.09%
Curated By: Auto accepted
Lifecycle: Published
Description:
Customers category based on rewards program.
Declined (0)
No declined Classification assets.
Open Events
No data to display
Associated Business Data Sets
No data to display
Inferred Data Types
Integer (1) 5 rows 20%
Pattern
XXXXXXX 11 rows 44%
Generated Classifications

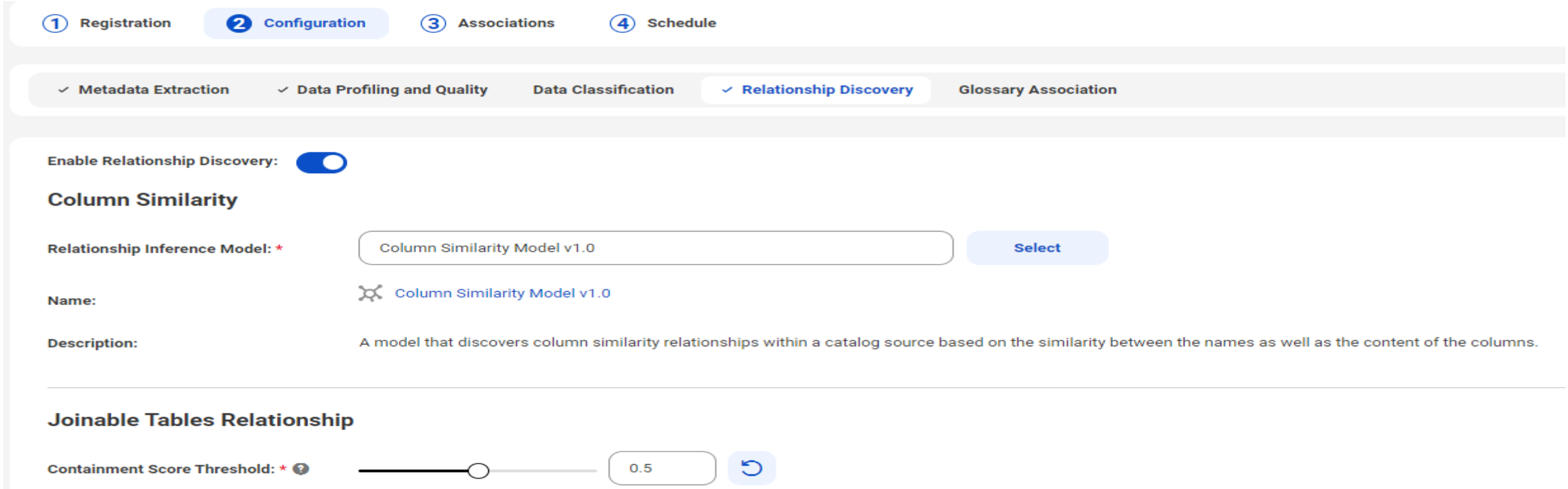
# Intelligent Data Entity Classification

Once domains for columns have been identified, CLAIRE acts as a copilot by assembling these individual fields into higher-level business entities.



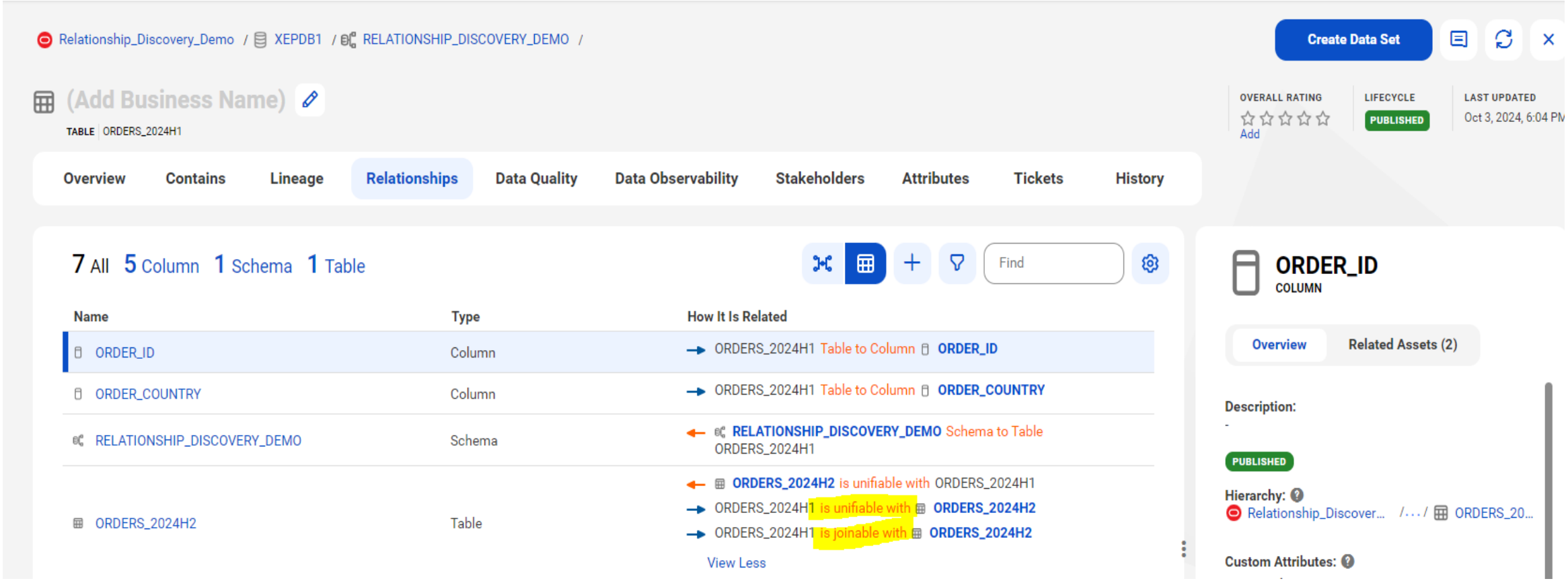
# Enhance Relationship Discovery

- **CLAIRE** uses machine-learning techniques to automatically identify primary and unique keys and joins across structured datasets.
- **CLAIRE** continuously improves its ability to identify relationships by including humans in the data-curation process.



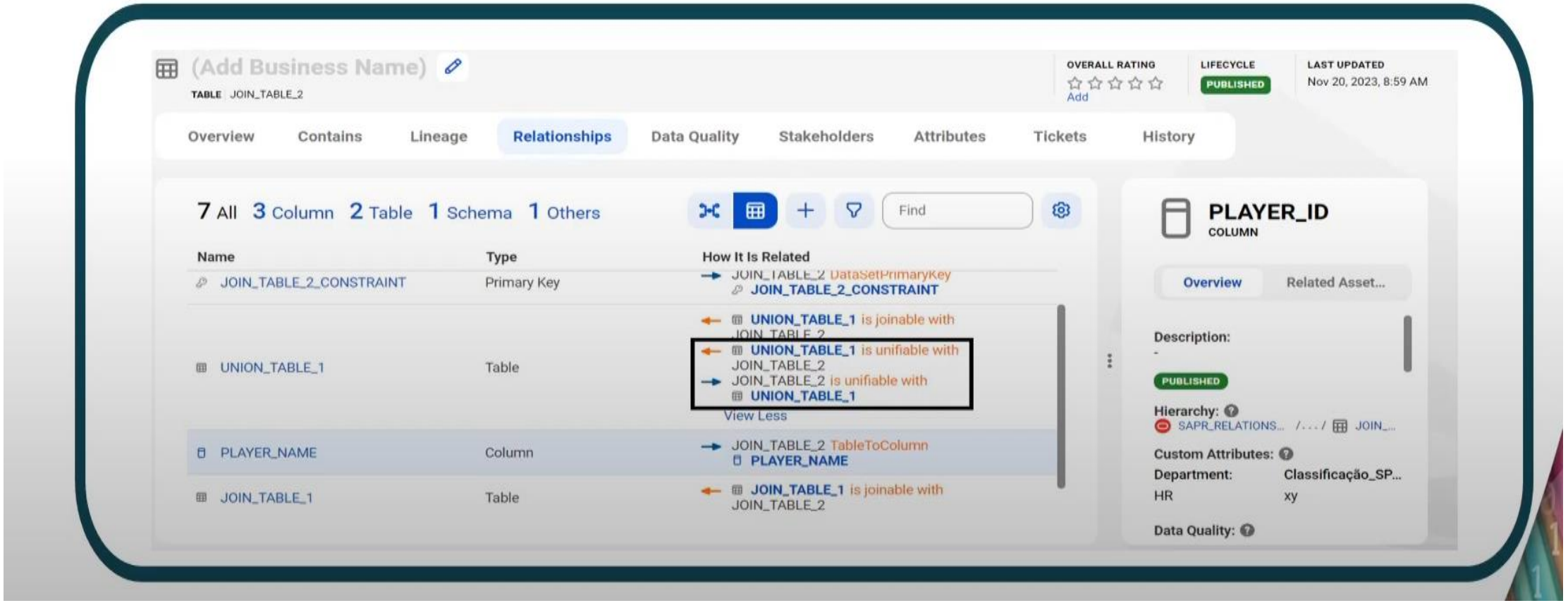
# Joinable Inference

This feature discovers and visualizes relationships between tables based on the join conditions between them.



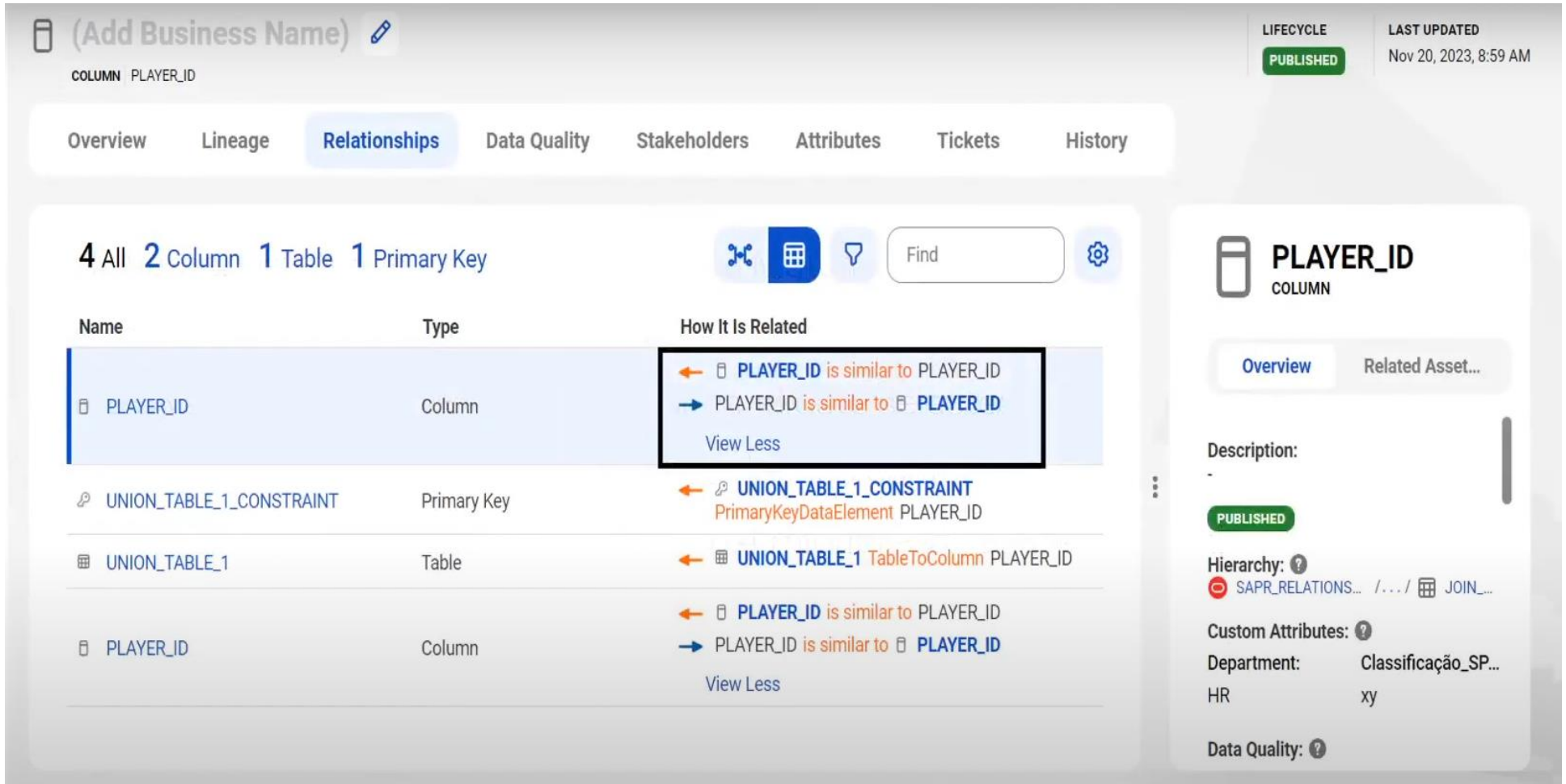
# Union Inference Predictions

This feature discovers and visualizes relationships between tables based on the union conditions between them.



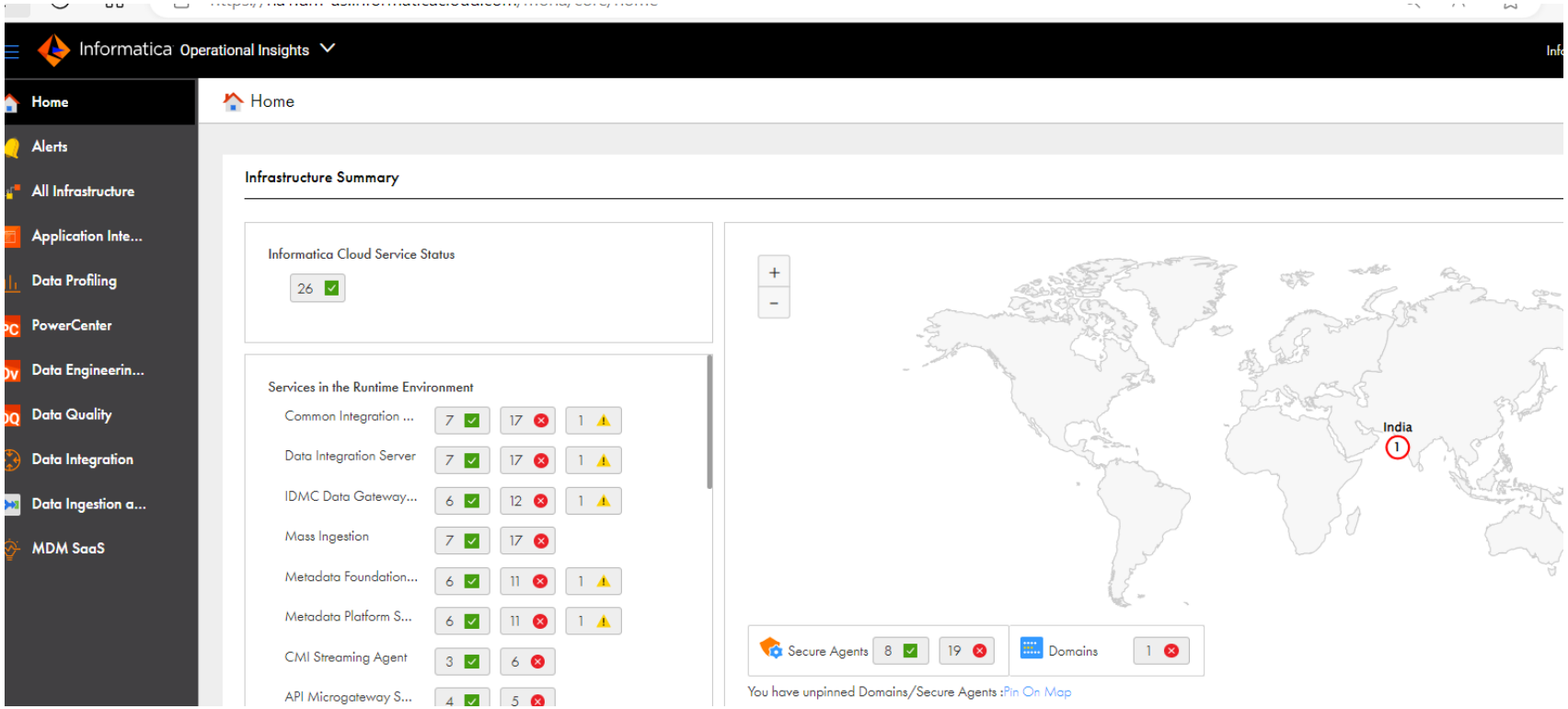
# Data Similarity

CLAIRE acts as a copilot using ML techniques like clustering to detect similar data across thousands of databases and file sets. Intelligent data similarity is one of the critical capabilities used for multiple purposes, including identifying data, detecting duplicates, combining individual data fields into business entities, propagating tags across datasets and recommending datasets as a trusted user copilot.

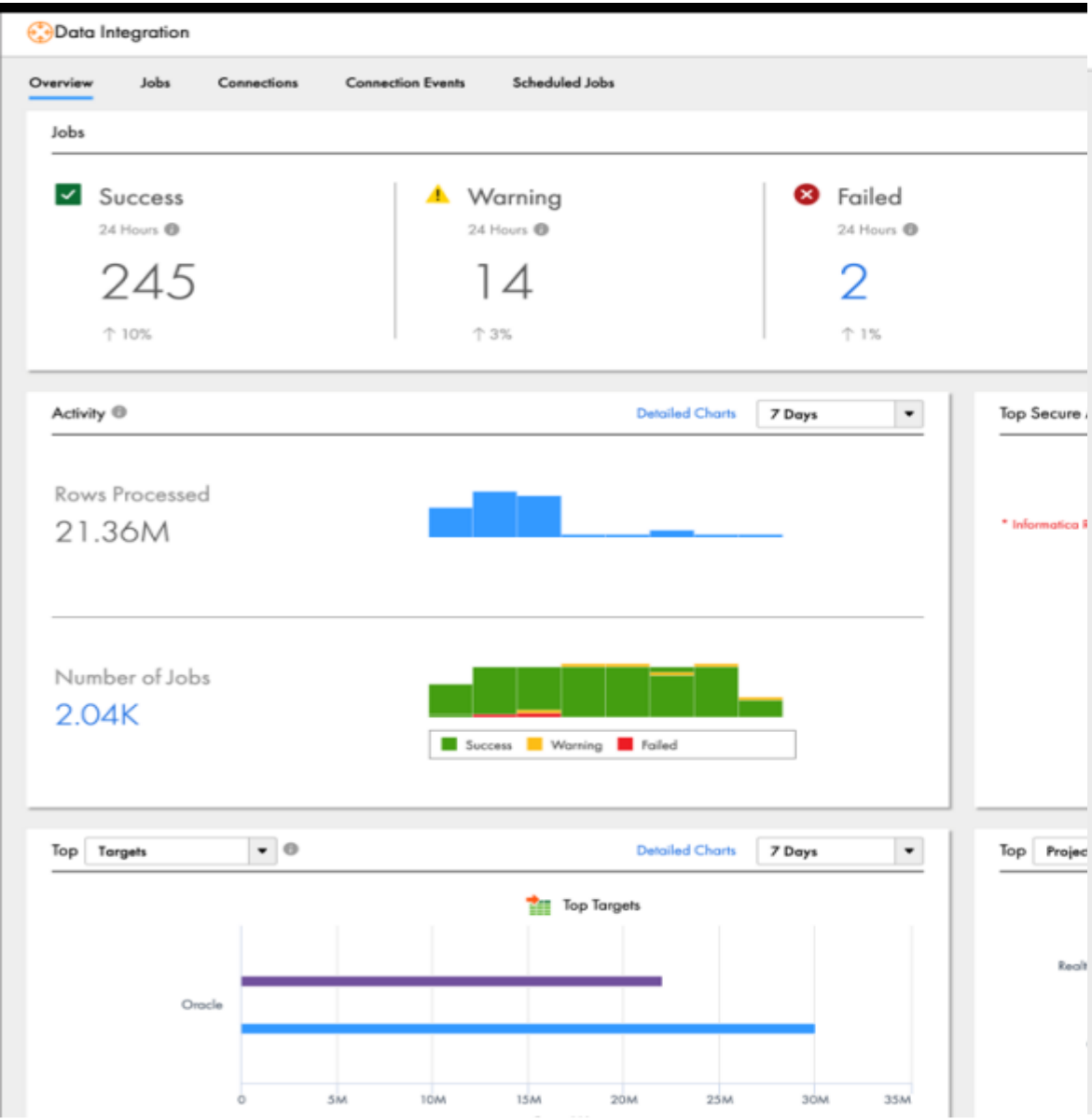


# CLAIRE as a DataOps Copilot

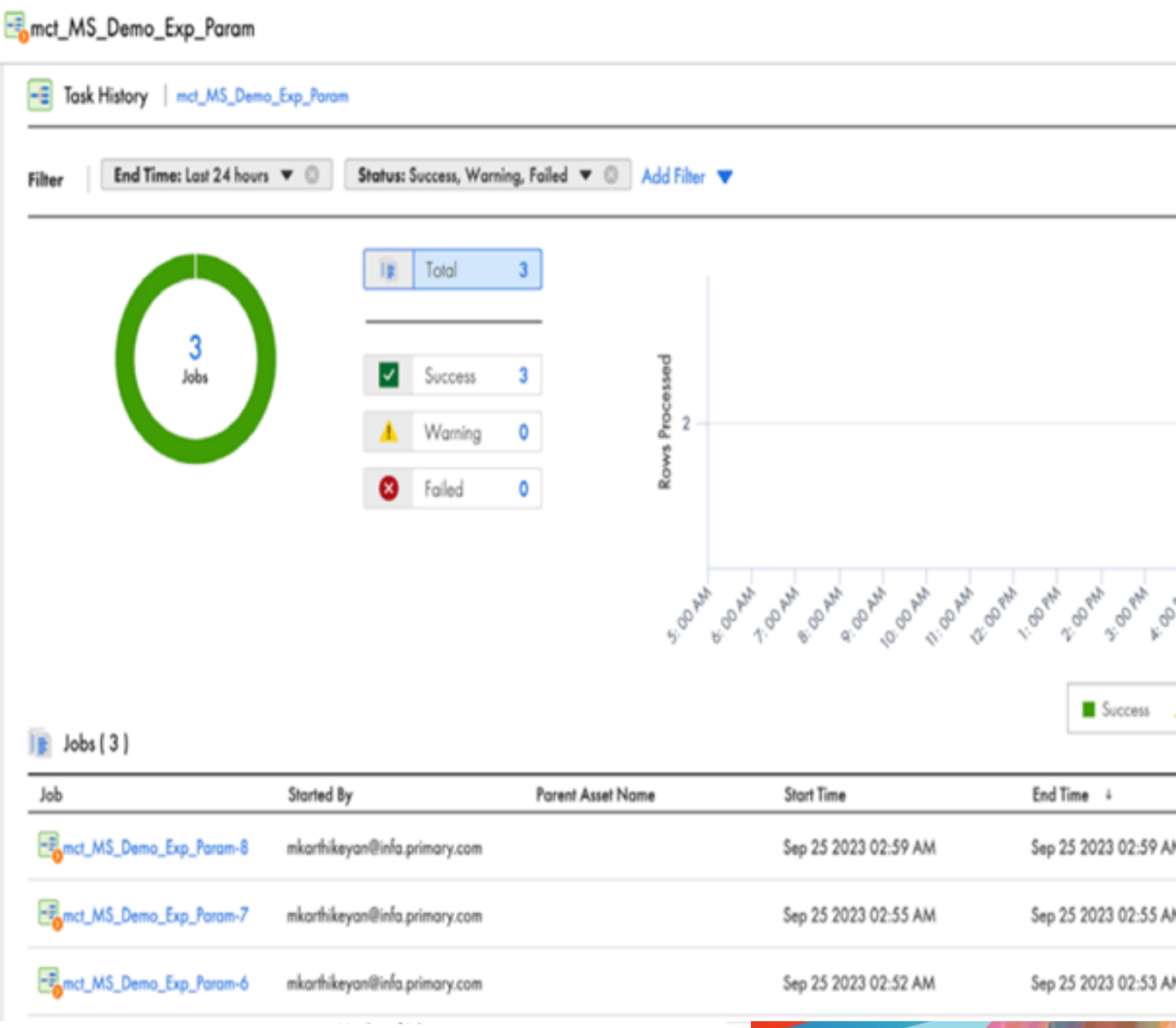
CLAIRE serves as an innovative DataOps Copilot, streamlining data workflows and enhancing collaboration across teams.This intelligent assistant not only simplifies complex data tasks but also fosters a culture of data-driven insights, ultimately driving business success in a competitive landscape.



## ❖ Insightful and Predictive Analytics

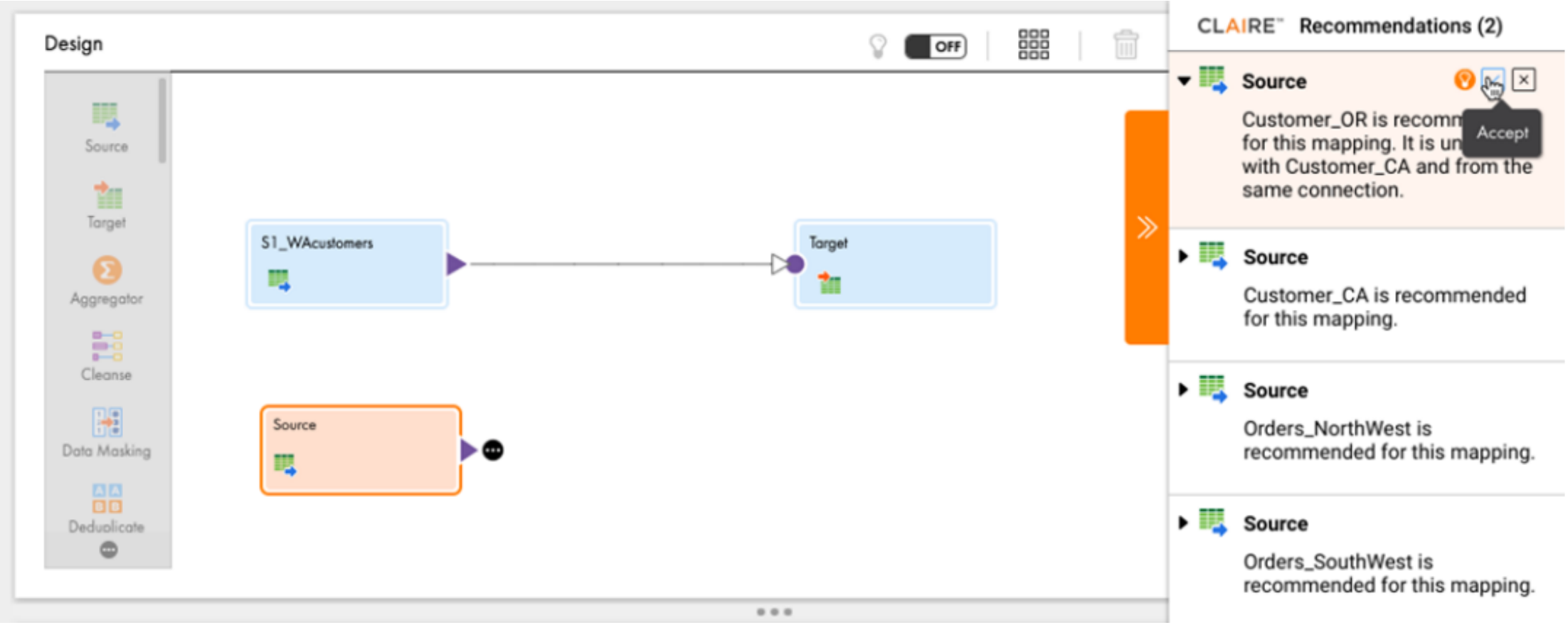



## ❖ Anomaly Detection in Job Runs



# CLAIRE As Analytics Co-pilot

## Source Recommendations







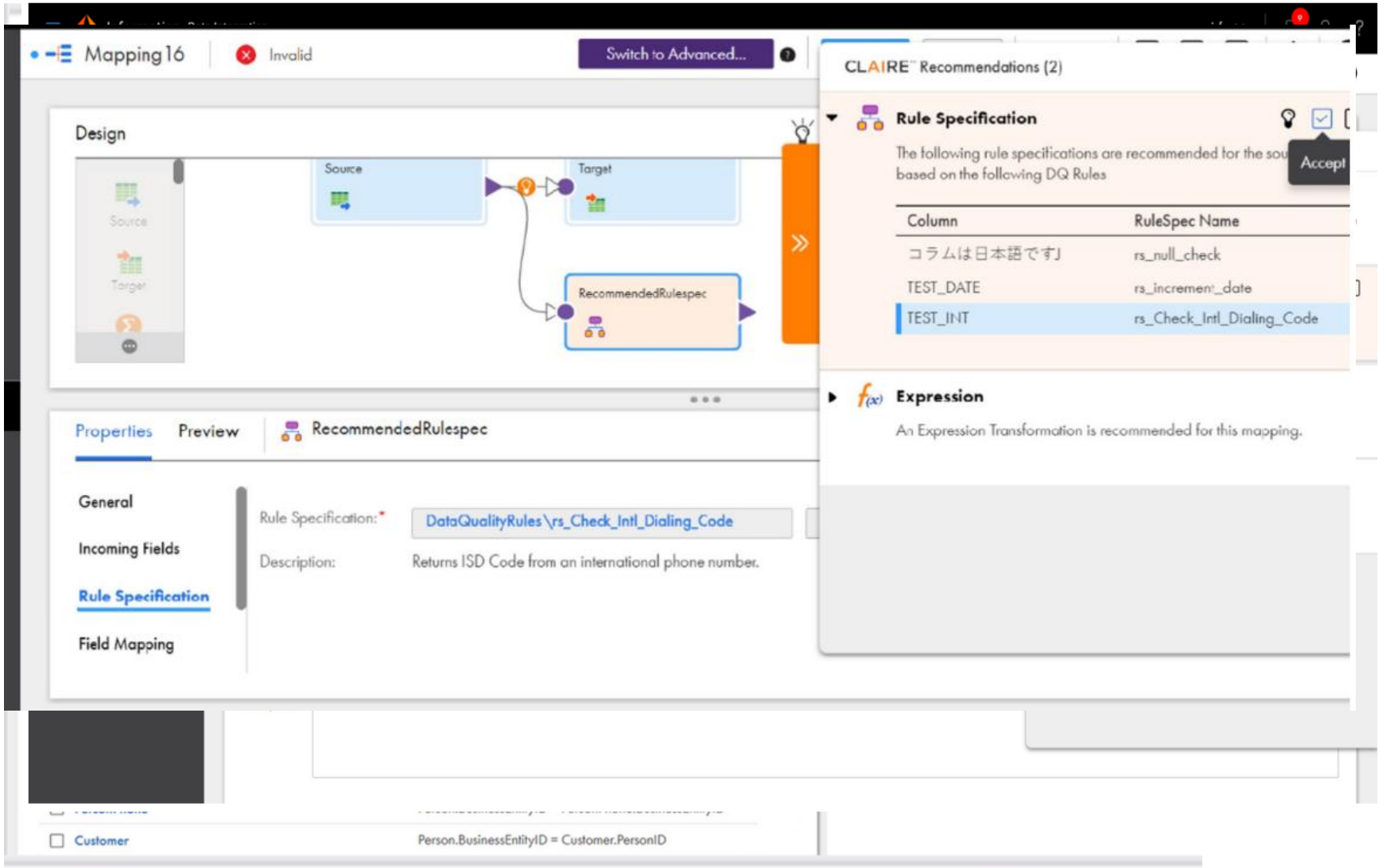

Informatica

# CLAIRE As Analytics Co-pilot

Transformation Recommendations

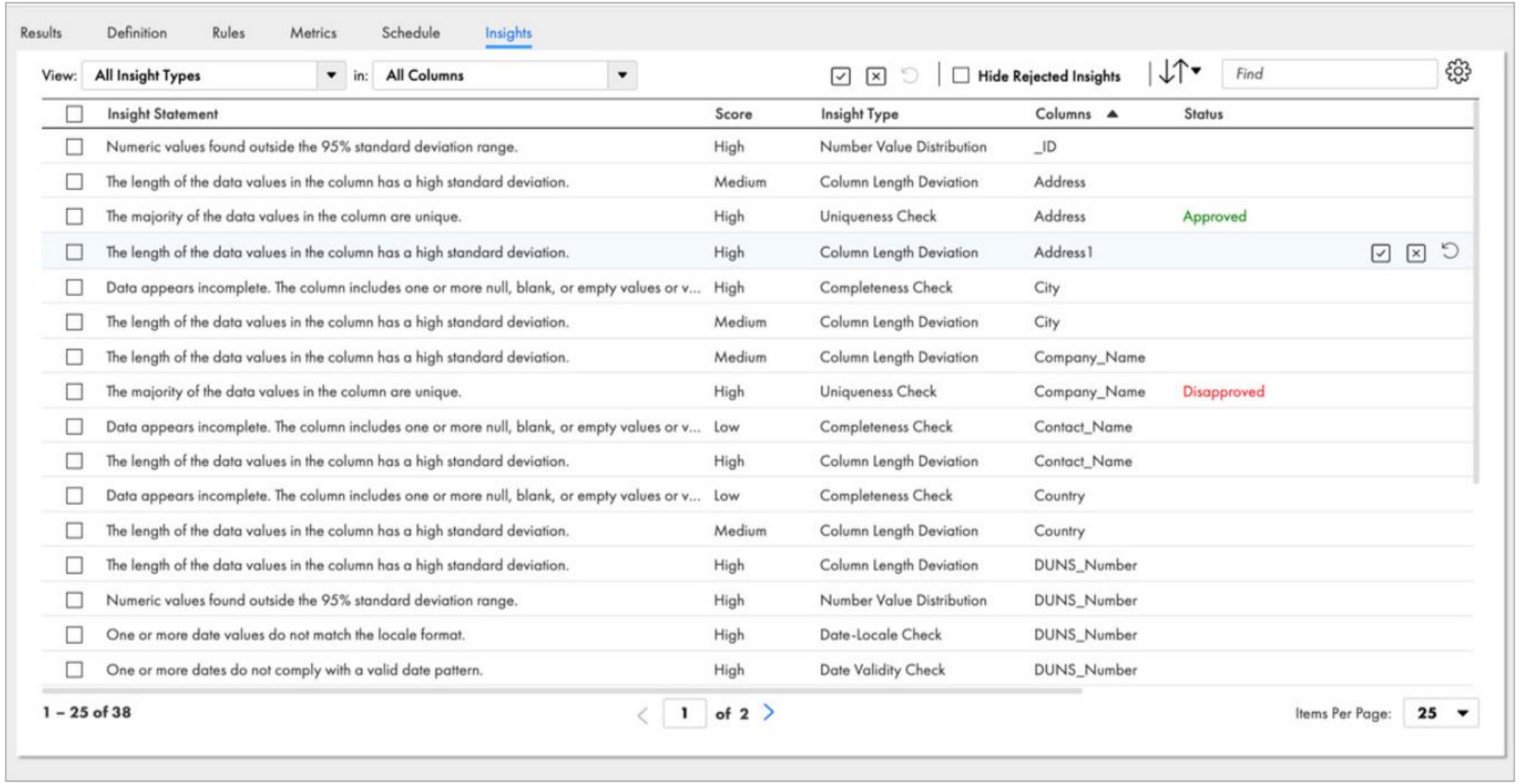
# CLAIRE As Analytics Co-pilot

- Join-Column Recommendations
- Auto mapping Recommendations
- Mapplet/UDF Recommendations
- DQ Rule Recommendations



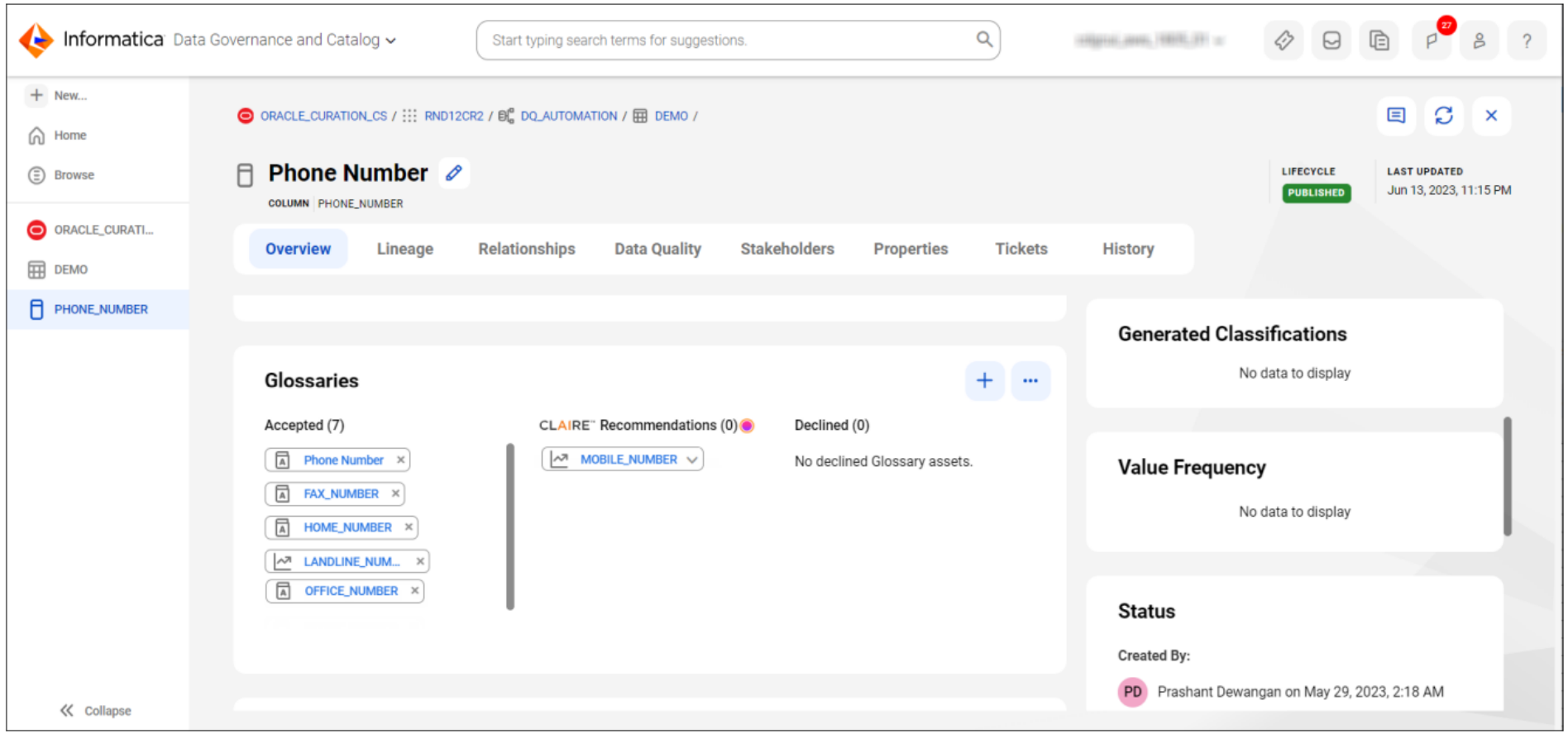
# CLAIRE as Data Governance and DQ Copilot

## Insights from profiling results



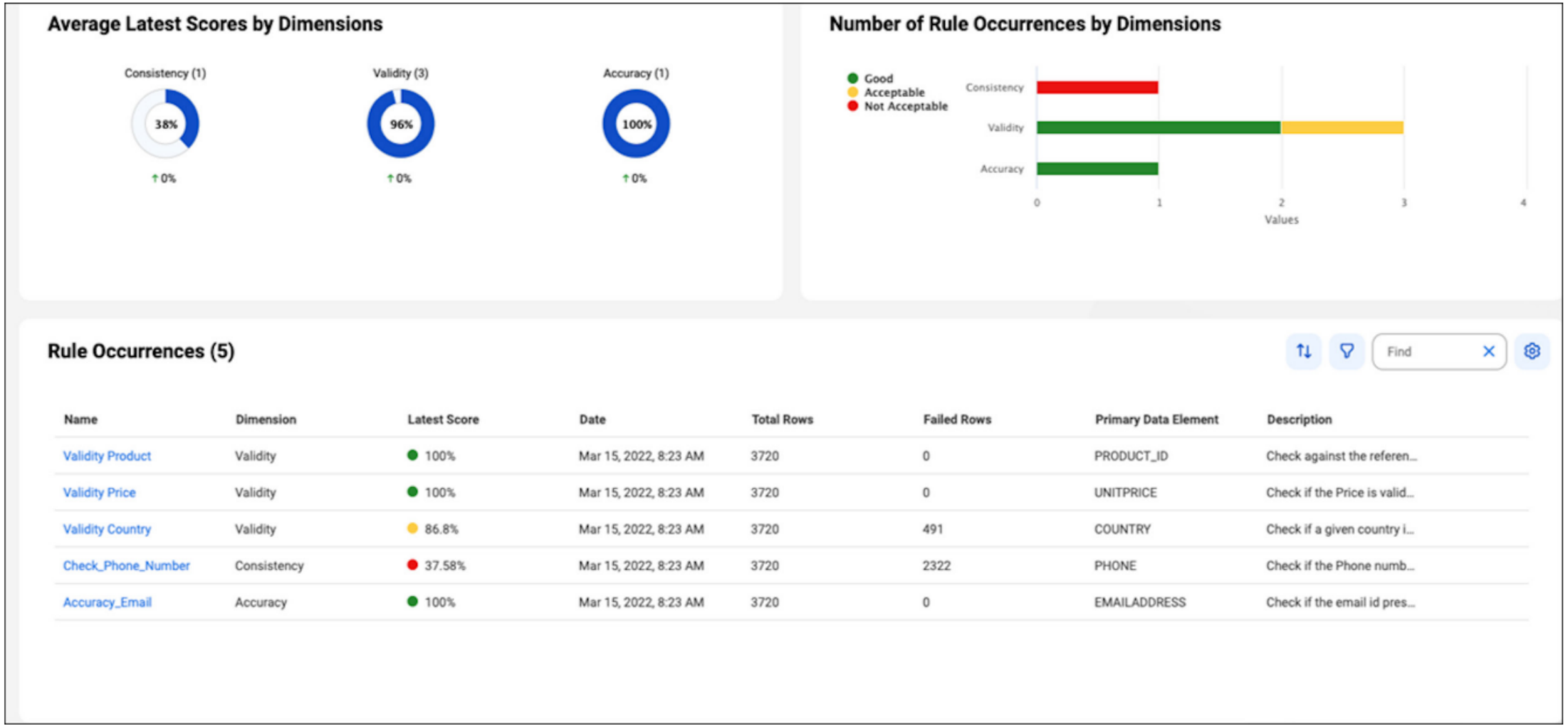
# CLAIRE as Data Governance and DQ Copilot

## Automatically Associate Business Terms With Physical Datasets



# CLAIRE as Data Governance and DQ Copilot

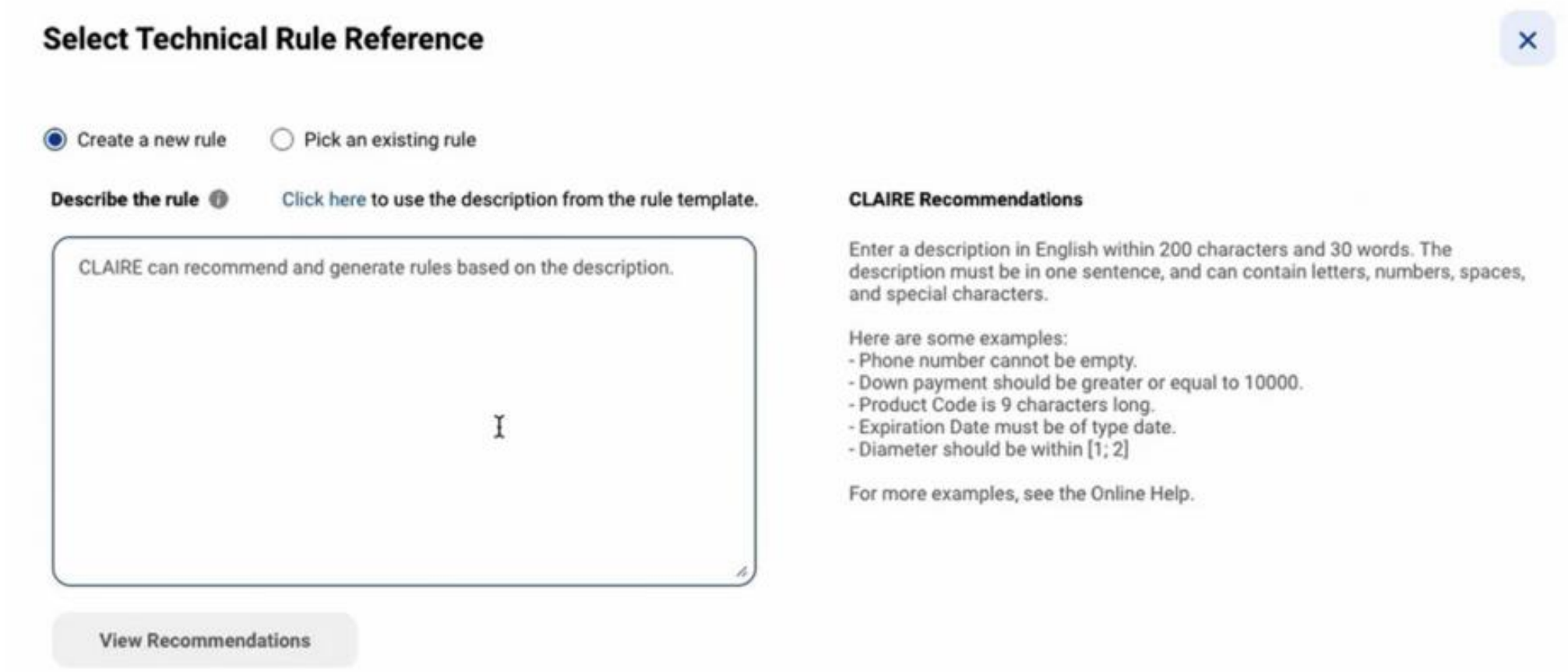
## Automatically Assess Data Quality




| Name | Dimension | Latest Score | Date | Total Rows | Failed Rows | Primary Data Element | Description |
|---|---|---|---|---|---|---|---|
| Validity Product | Validity | 100% | Mar 15, 2022, 8:23 AM | 3720 | 0 | PRODUCT_ID | Check against the referen... |
| Validity Price | Validity | 100% | Mar 15, 2022, 8:23 AM | 3720 | 0 | UNITPRICE | Check if the Price is valid... |
| Validity Country | Validity | 86.8% | Mar 15, 2022, 8:23 AM | 3720 | 491 | COUNTRY | Check if a given country i... |
| Check_Phone_Number | Consistency | 37.58% | Mar 15, 2022, 8:23 AM | 3720 | 2322 | PHONE | Check if the Phone numb... |
| Accuracy_Email | Accuracy | 100% | Mar 15, 2022, 8:23 AM | 3720 | 0 | EMAILADDRESS | Check if the email id pres... |

# CLAIRE as Data Governance and DQ Copilot

## ML/NLP Assisted Data Quality Rule Creation



# Informatica's AL/ML Solutions


CLAIRE® copilot + CLAIRE® GPT

# CLAIRE as GPT

## CLAIRE® GPT Business Value

### A New Conversational Interface to Data Management for All Data Consumers

| Data Discovery | Data Exploration | Data Transformation (ELT) | Product Help |
|---|---|---|---|
| • Democratize data access<br>• Conversational Interface for Data Discovery<br>• Enables Data Accessibility<br>• Drives Adoption<br>• Increases Enterprise Data Literacy | • Single place to easily understand data that's fit of use<br>• View data<br>• Comprehend Data profile<br>• Explore Data<br>• Ask Data Questions<br>• Data Query in Natural Language<br>• Instant Answer to Data Questions<br>• Exploratory Data Analysis | • Transform data within the same Data Lake<br>• Data Pipeline Quick Start<br>• NL as the Transformation Language<br>• Draft ELT Data Pipelines Creation<br>• Edit and modify as needed in Data Integration | • Increased literacy of IDMC products for data management use cases<br>• Ask documentation questions<br>• Product Help<br>• Usage and best practices |

Informatica

# DEMO

CLAIRE Co-pilot


Where data & AI come to LIFE

# References

| | |
|---|---|
| **CLAIRE AI Engine – Intelligent Automation** | https://www.informatica.com/in/about-us/claire.html |
| **Informatica Intelligent Data Management Cloud CLAIRE®Security & Privacy Overview** | https://docs.informatica.com/content/dam/source/resources/non-cms-pdfs/claire-gpt/IDMC%20CLAIRE%20security%20and%20privacy%20whitepaper.pdf |
| **Artificial Intelligence for the Data-Driven Enterprise** | https://www.informatica.com/content/dam/informatica-com/en/collateral/white-paper/artificial-intelligence-for-data-driven-disruption_white-paper_3328en.pdf |
| **Relationship Discovery in CDGC** | https://www.youtube.com/watch?v=tTlz4x6sni8 |


# Thank You

Where data & AI come to LIFE