

18 Aug 2026

AI for the Catalog, Catalog for the AI

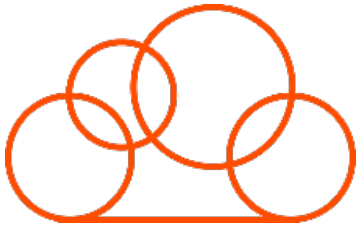
David Gaffaney, Senior Solution Architect, Professional Services

Ratish Neelakandan, Senior Technical Architect, Professional Services

Housekeeping Tips



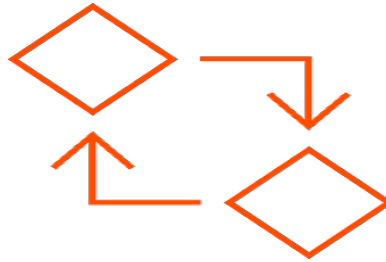
- Today's session is scheduled for **1 hour** and will feature a presentation followed by a Q&A session at the end.
- To ensure a smooth presentation without interruptions, all participants will be muted.
- Please submit your questions to "All Panelists" using the **Q&A panel**. We look forward to answering them during the live Q&A.
- This session is **being recorded**. The recording and slide deck will be available on our **Success Portal**, and a recorded link will be emailed to you after the event.
- We value your input! Please take a moment to complete the **post-webinar survey** to share your feedback and suggest future topics.



Bootstrap trial and POC
Customers



Enriched Customer
Onboarding
experience



Product Learning
Paths and Weekly
Expert Sessions

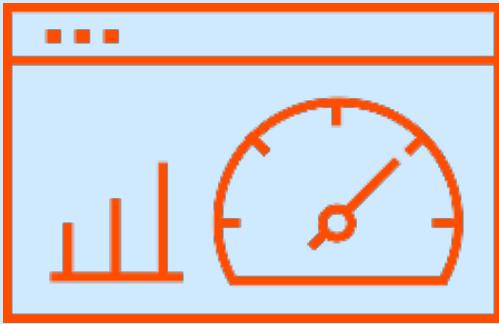


Informatica
Concierge



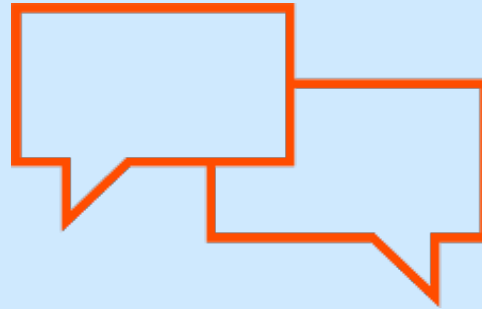
Tailored training and
content
recommendations

More Information



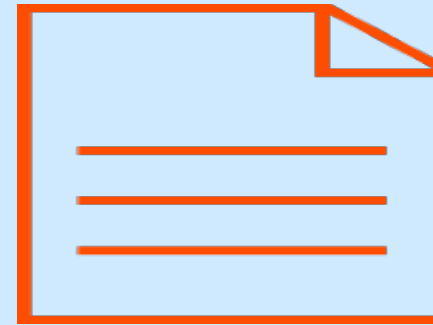
Success Portal

<https://success.informatica.com>



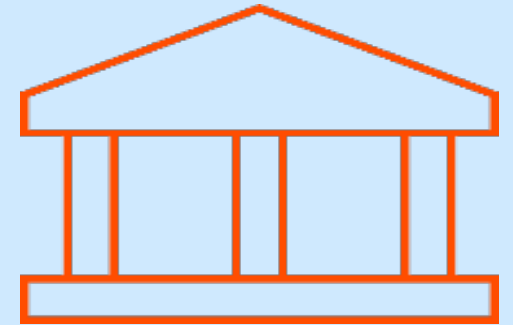
Communities & Support

<https://network.informatica.com>



Documentation

<https://docs.informatica.com>



University

<https://www.informatica.com/in/service-s-and-training/informatica-university.html>

Safe Harbor



Disclaimer: The information being provided herein is for informational purposes only. The development, release and timing of any Informatica product, service or functionality described herein remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made herein are based on information currently available, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products, services or functionality in the future.

Forward-looking statements



This presentation contains forward-looking statements about, among other things, trend analyses and statements regarding future events, anticipated growth and industry prospects, and our strategies, expectation or plans regarding product releases and enhancements. The achievement or success of the matters covered by such forward-looking statements involves risks, uncertainties and assumptions. If any such risks or uncertainties materialize or if any of the assumptions prove incorrect, results or outcomes could differ materially from those expressed or implied by these forward-looking statements. The risks and uncertainties referred to above include those factors discussed in Salesforce's reports filed from time to time with the Securities and Exchange Commission, including, but not limited to our ability to consummate the pending acquisition of Informatica on a timely basis or at all; our ability to meet the expectations of our customers; uncertainties regarding AI technologies and their integration into our product offerings; the effect of evolving domestic and foreign government regulations; regulatory developments and regulatory investigations involving us or affecting our industry; our ability to successfully introduce new services and product features, including related to AI and Agentforce; our ability to execute our business plans; the pace of change and innovation and our ability to compete in the markets in which we participate; and our ability to maintain and enhance our brands.

Session Agenda

1 The Enterprise Data Challenge

2 Introducing the CDGC & CLAIRE

3 AI for the Catalog — Building Trust at Scale

4 Catalog for the AI — Powering Reliable AI Solutions

5 Demo & Key Takeaways

The Enterprise Data Challenge

Enterprises manage data across dozens of disconnected sources – Snowflake, Oracle, SAP, Databricks and more

Volume



Manual metadata management is slow, inconsistent and impossible to scale

Risk



Without governed, enriched metadata – AI models hallucinate, compliance fails, and business users don't trust the data

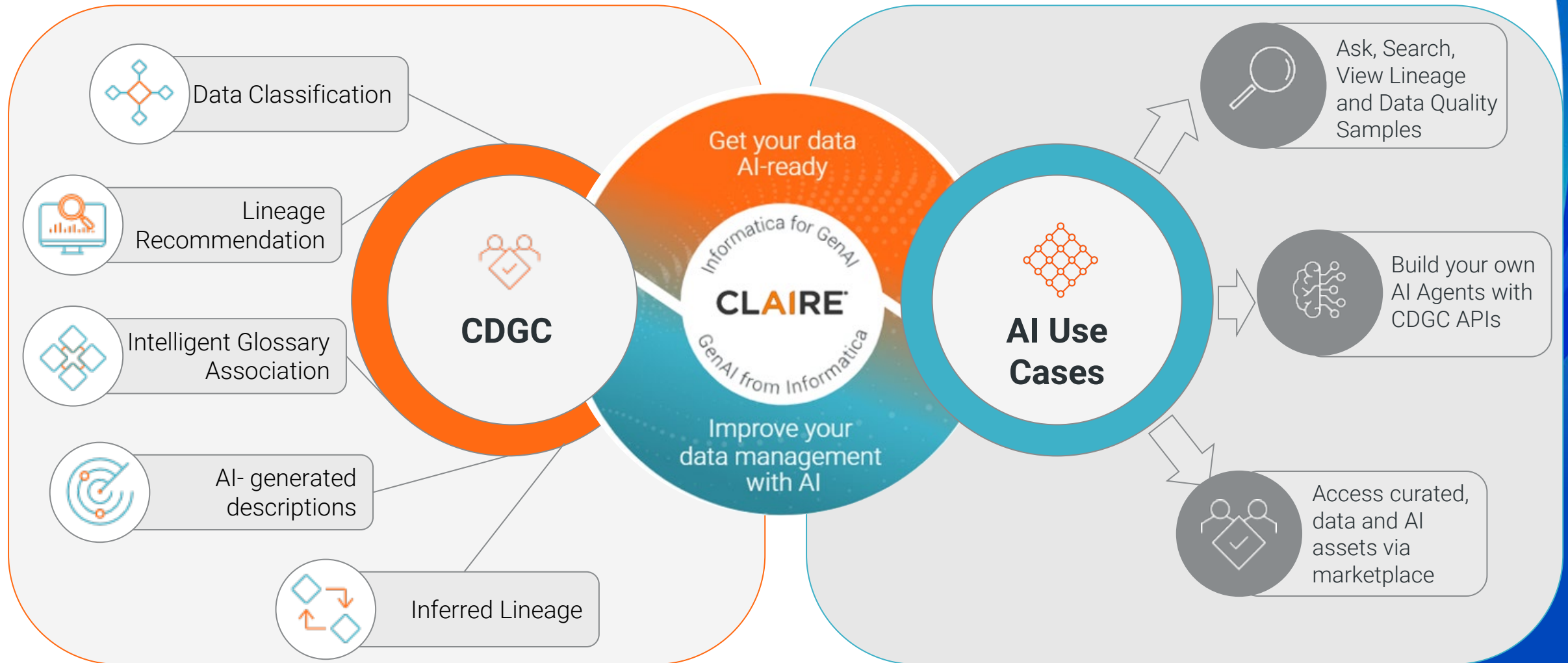
Impact



The cost of ungoverned data: regulatory fines, poor AI outcomes, and low catalog adoption

To unlock the full power of AI, enterprises need a governed, enriched and trusted catalog – built at speed and scale

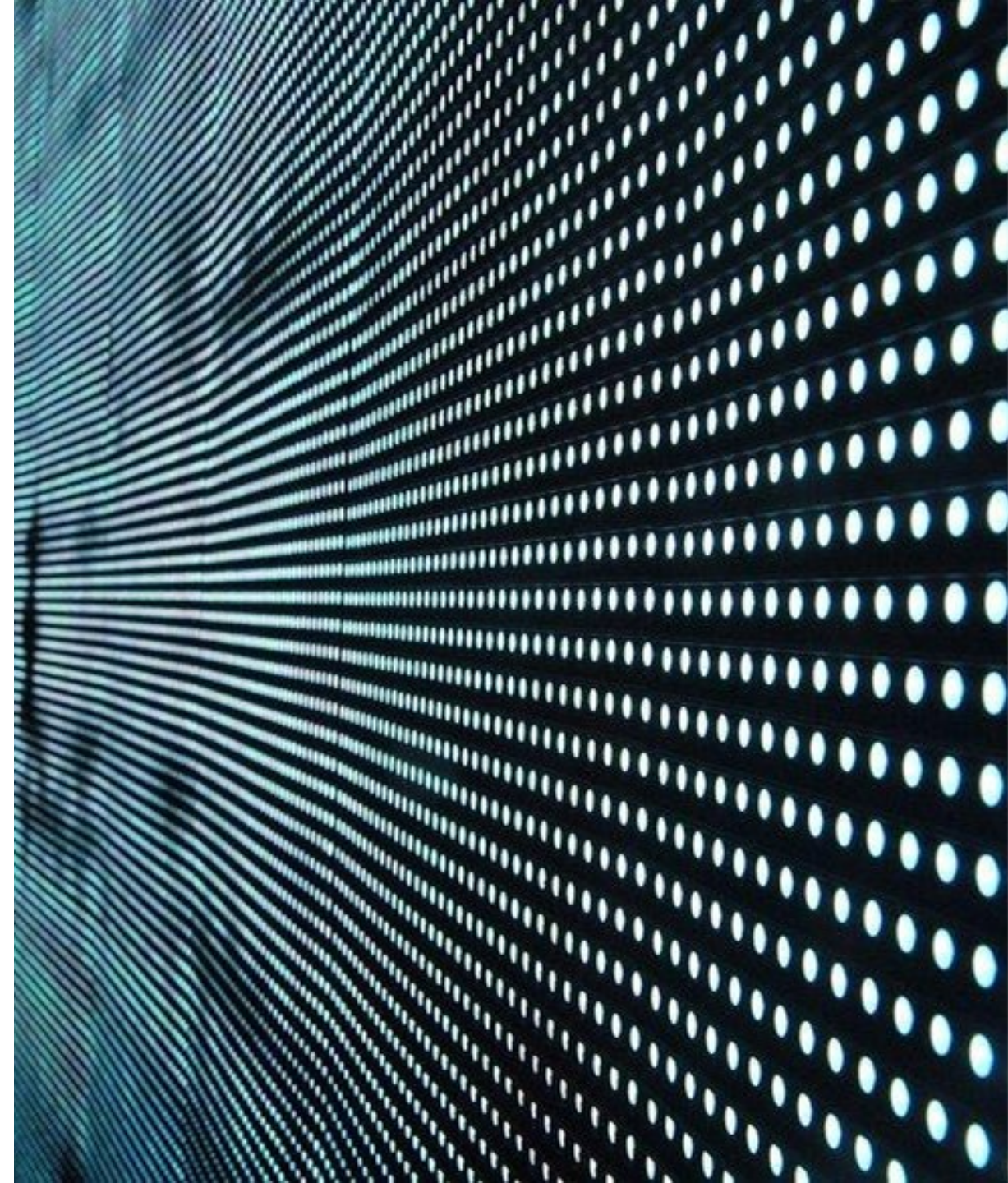
AI Builds the Catalog + Catalog Powers AI



Cycle: AI Builds the Catalog



Let's See How
CLAIRE Builds the Catalog



AI for the Catalog

Automatically Build Lineage - From Source to Target

- **Enabled during metadata scan** on any source – Snowflake, Informatica, Oracle, Databricks, SAP etc.
- **Claire AI analyzes metadata** and intelligently recommends connection assignments between source & target
- **One-click acceptance** triggers automatic lineage creation – no manual mapping required

Business Value: End-to-end data flow visibility at enterprise scale, from day one. Eliminates weeks of documentation effort with instant, accurate lineage.

Registration Configuration Associations Schedule

Metadata Extraction Lineage Discovery Data Profiling and Quality Data Observability Data Classification Relationship Discovery Glossary Association

Enable Lineage Discovery: ☒

Filters

Specify lineage discovery filters: ☒ No ☐ Yes

1 Enable Lineage Discovery in MCC

Related Catalog Sources Assign Connections

Related Catalog Sources (1)

Find

CLAIRe™ recommendations can become outdated over time. Run the catalog source job again with the lineage discovery capability to refresh the recommendations.

Select All

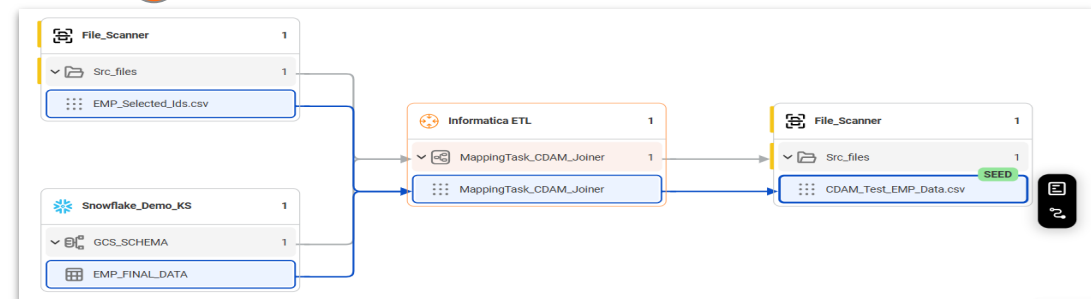
☐	Oracle_Nus_test	0 Assigned - 1 Pending	100 %	CLAIRe™ Recommendation
--------------------------	-----------------	------------------------	-------	------------------------

Recommended On: Aug 21, 2025, 12:42 AM

Accept

Reject

2 Accept the Recommendation in MCC



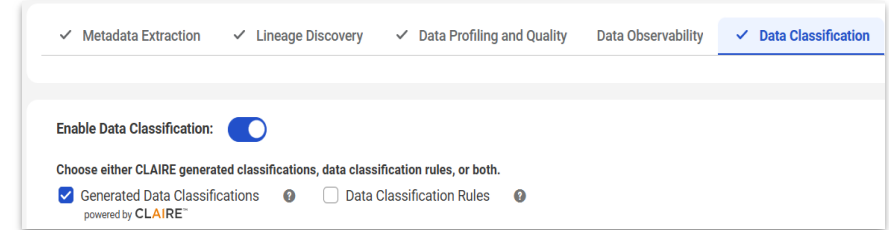
3 End to End Lineage in CDGC

AI for the Catalog

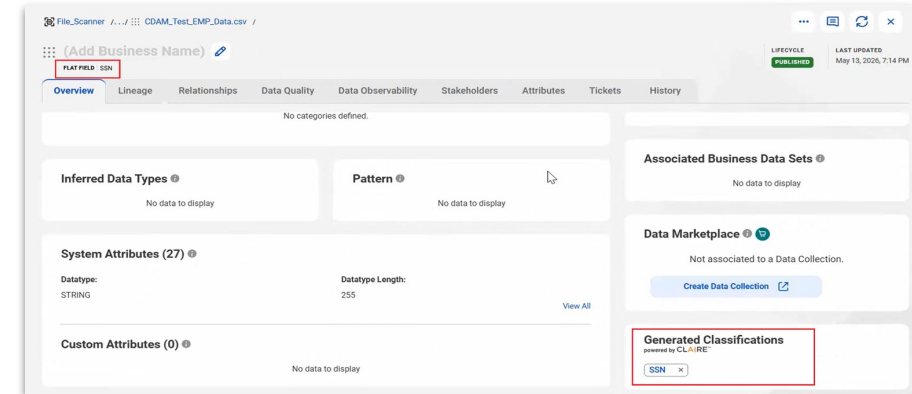
AI-Driven Data Classification — Emphasize Data's Sensitivity

- **Enabled via Generated Data Classifications** during the metadata scan
- **CLAIRE reads column name metadata** and intelligently generates Classification tag for every data element.
- **No manual rules**, no data profiling required — purely metadata-driven intelligence
- Example: A healthcare org scans 500+ Databricks tables — SSN, DATE_OF_BIRTH, DIAGNOSIS_CODE instantly flagged as **High/Medium Sensitivity**

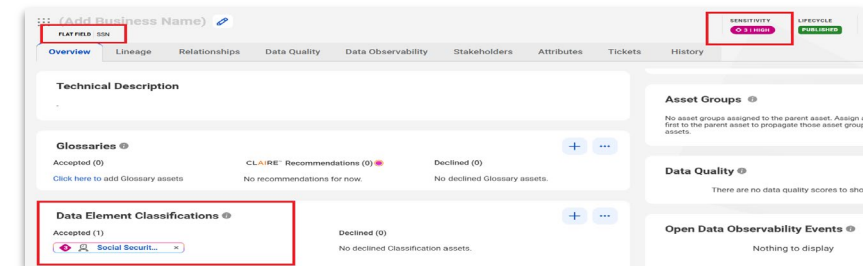
Business Value: Accelerates compliance readiness for GDPR, HIPAA, CCPA and other regulations. Consistent, repeatable, scalable classification across thousands of tables — in minutes



Enable CLAIRE Generated Classifications in MCC



CLAIRE Generated Classification on data element in CDGC



Promoted Classification on Data Element along with the Sensitivity

AI for the Catalog

Intelligent Glossary Association – Business Language at Scale

- **Automatically associates** technical data elements with Business Glossary terms available in CDGC
- **Assigns meaningful business names** to cryptic technical column names
- **Uses synonyms, keywords, and abbreviations** to intelligently match glossary terms to data elements
- Example: LCR_TRUST_TYPE_CD → matched via LCR = Lender Credit Rating, CD = Code → assigned business name "*Lender Credit Rating Trust Type Code*"

Business Value: Higher catalog adoption, faster onboarding, and a shared business vocabulary across the enterprise. Business users can now find, understand, and trust data – without ever asking IT

The screenshot displays the Informatica Intelligent Glossary Association interface. The top navigation bar includes tabs for Registration, Configuration, Associations, and Schedule. The Configuration tab is active, showing settings for Metadata Extraction, Lineage Discovery, Data Profiling and Quality, Data Observability, Data Classification, Relationship Discovery, and Glossary Association. The Glossary Association section includes checkboxes for 'Enable auto-acceptance' and 'Enable Below-threshold Recommendations', both of which are checked. Below these are sliders for 'Confidence Score Threshold for Auto-Acceptance' (set to 90%) and 'Confidence Score Threshold for Recommendations' (set to 80%). There are also sections for 'Assign Business Names and Descriptions' (set to Yes) and 'Ignore Keywords' (set to No). The bottom section shows the 'Glossary Association Scope' for the 'Lender Credit Rating Trust Type Code' column. This section includes tabs for Overview, Lineage, Relationships, Data Quality, Data Observability, Reference Data, and Stakeholders. The Overview tab is active, showing a description of the Lender Credit Rating Trust Type Code, a technical description, and a list of glossaries. The glossaries list shows one accepted glossary, 'Lender Credit ...', and no recommendations or declined glossaries.

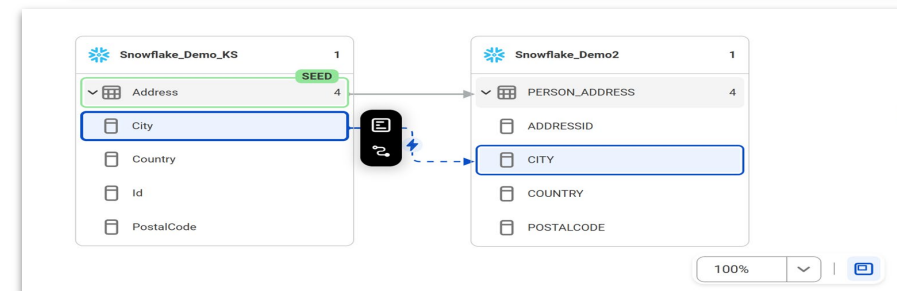
AI for the Catalog

AI Lineage Across Catalog Sources – No ETL\ELT Code Needed

- Configured between **Source and Target** catalog sources – e.g., Oracle ↔ Google BigQuery
- Claire analyzes **column and table name** metadata across platforms and auto-generates lineage links
- No ETL pipeline metadata** or shared code required – works purely on metadata intelligence

Business Value: Unified lineage view across heterogeneous environments – without a single line of code. Solves the "disconnected systems" problem in complex, multi-platform data estates

The screenshot shows the 'AI Address Data Load' configuration window. It has tabs for 'Registration', 'Configuration', and 'Linking Method'. The 'Configuration' tab is active, showing 'Catalog Source' as 'Snowflake_Demo_KS' and 'Schema' as 'GCS_SCHEMA'. Below this is a 'Filters' section with a description and a button to 'Add Criteria'. The 'Target Catalog Source' section shows 'Catalog Source' as 'Snowflake_Demo2' and 'Schema' as 'GCS_SCHEMA', also with a 'Filters' section. Below these are three settings: 'Refresh Lineage' (a checkbox), 'Linking Method' (radio buttons for 'Rule-based Linking' and 'Automated Linking powered by CLAIRe', with the latter selected and highlighted by an orange box), and 'Enable auto-acceptance' (a toggle switch). At the bottom is a 'Confidence Score Threshold for Auto-Acceptance' slider set to 95%.



AI for the Catalog

AI-Generated Descriptions & Business Names – Enrich at Scale

- **CLAIRE AI generates rich, meaningful descriptions** and business names for technical assets within CDGC
- **Industry-aware** – understands context across Banking, Healthcare, Telecom, Retail, Manufacturing and more
- Example: Employees_CST_SP → “Cost of Service per Employee”

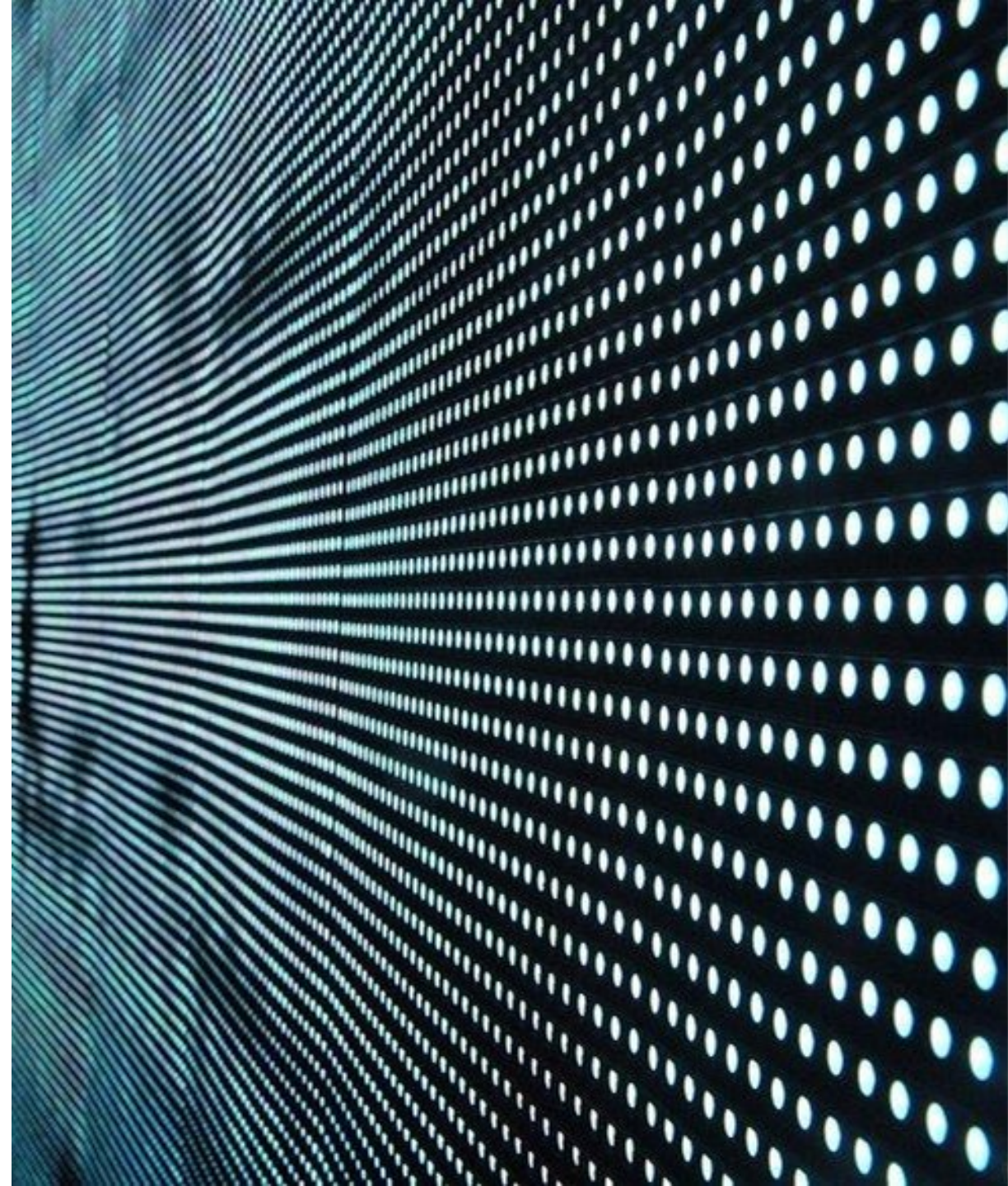
Business Value: Catalog enrichment at scale – no human effort, no delays, no inconsistency. Removes the burden of manual documentation from data stewards and catalog owners.

The screenshot displays the Informatica Data Governance and Catalog interface. The top navigation bar includes 'Data Governance and Catalog', 'Adaptive', and a search bar. The main content area shows a table of data elements for 'Employees_CST_SP'. A modal window titled 'Edit Business Name' is open, allowing users to 'Type or use CLAIRE to generate business name'. A 'Generate with CLAIRE' button is visible. Below this, a 'Generate Text' modal is shown, where users can select 'Business Name' and 'Description' for generation. The 'Industry' dropdown is set to 'Generic', and the 'Text length' is set to 'Generic'. A green notification banner at the bottom states 'Generated business names and description for 16 assets'. The bottom section of the interface shows a table of data elements with columns for '#', 'Name', 'Description', and 'Business Name'. The table lists 16 data elements, including BusinessEntityID, NationalIDNumber, LoginID, OrganizationNode, OrganizationLevel, JobTitle, BirthDate, and MaritalStatus, each with a generated description and business name.

#	Name	Description	Business Name
1	BusinessEntityID	A Business Entity Identifier is a unique code assigned to individuals or entities within ba...	Business Entity Identifier
2	NationalIDNumber	A National ID Number is a unique identifier assigned to individuals by a government or ...	National ID Number
3	LoginID	A Login Identifier is a unique code assigned to employees in the banking industry to aut...	Login Identifier
4	OrganizationNode	An Organization Node represents a unique identifier or reference point within a hierarch...	Organization Node
5	OrganizationLevel	Organizational Unit refers to a hierarchical structure within an organization that defines ...	Organizational Unit
6	JobTitle	Job Title represents the designated position or role within an organization, typically def...	Job Title
7	BirthDate	Birth Date represents the specific calendar date on which an individual was born. It is a ...	Birth Date
8	MaritalStatus	Marital Status refers to the legal and social status of an individual's relationship, typica...	Marital Status

Cycle: Catalog Powers AI

It is time to see how
CDGC Powers AI



Catalog for AI

The Core Idea

AI and RAG systems are only as good as their data context. Raw tables provide nothing to reason with; metadata enrichment delivers the critical understanding and trust required for enterprise intelligence.

an

sl

Definitions (semantic enrichment)

at

e

1. Definitions

Semantic Enrichment

- Business glossary terms, descriptions, and synonyms tell the AI what a field actually represents (e.g. "Customer_ID" vs "Client_Key" mean the same thing).
- This lets an agent correctly map a user's natural-language question ("show me customer churn") to the right physical fields, instead of guessing or hallucinating a join.

fa

ct

_c

Data quality metadata

he

ck

2. Data Quality

Quality Metadata

- DQ scores, rule results, and trend metadata let AI weight or filter sources — e.g. deprioritize a table with a 60% completeness score, or flag an answer as "low confidence" because it drew from unvetted data .
- Without this, an AI treats a stale or dirty table exactly the same as a certified one — the answer looks equally confident either way, which is dangerous.

ac

co

un

t

Lineage

tr

ee

3. Data Lineage

Traceability

- Lineage tells the AI (and the user) *where a number came from* and what transformations touched it.
- For AI, lineage enables traceability/explainability — an agent can show its work ("this KPI derives from source X via transform Y"), which is critical for trust in regulated industries like pharma or insurance

Catalog for AI

CLAIRE GPT - Ask Anything. Find Everything. Instantly



Business Value: The CDGC catalog becomes the brain behind every AI conversation — delivering instant, governed intelligence that makes AI responses trustworthy, explainable, and enterprise-ready

Three screenshots of the Informatica CLAIRE GPT interface. The top screenshot shows a table titled "Business Assets in Data Governance and Catalog" with columns for #, Assets, Description, Created On, and Created By. The middle screenshot shows a "Lineage of EMP_FINAL_DATA Table" with a flow diagram of data sources and transformations. The bottom screenshot shows "Data Quality Scores" for the IDMCACCOUNTMASTER table, with three donut charts for Accuracy (100%), Completeness (43.9%), and Uniqueness (100%). Each screenshot includes a chat interface on the right with a query, a plan, reasoning, and key insights.

"Your catalog speaks your language — powered by Claire AI"

MCP Servers

The Core Idea

Informatica now offers headless capabilities, so that users can interact with the Catalog from their system of work and embed this in their workflow.

MCP Servers for Catalog

search

CDGC Metadata Search server

Execute high-performance metadata search queries, retrieve catalog assets, and integrate discovery features directly into external enterprise systems and workflows.

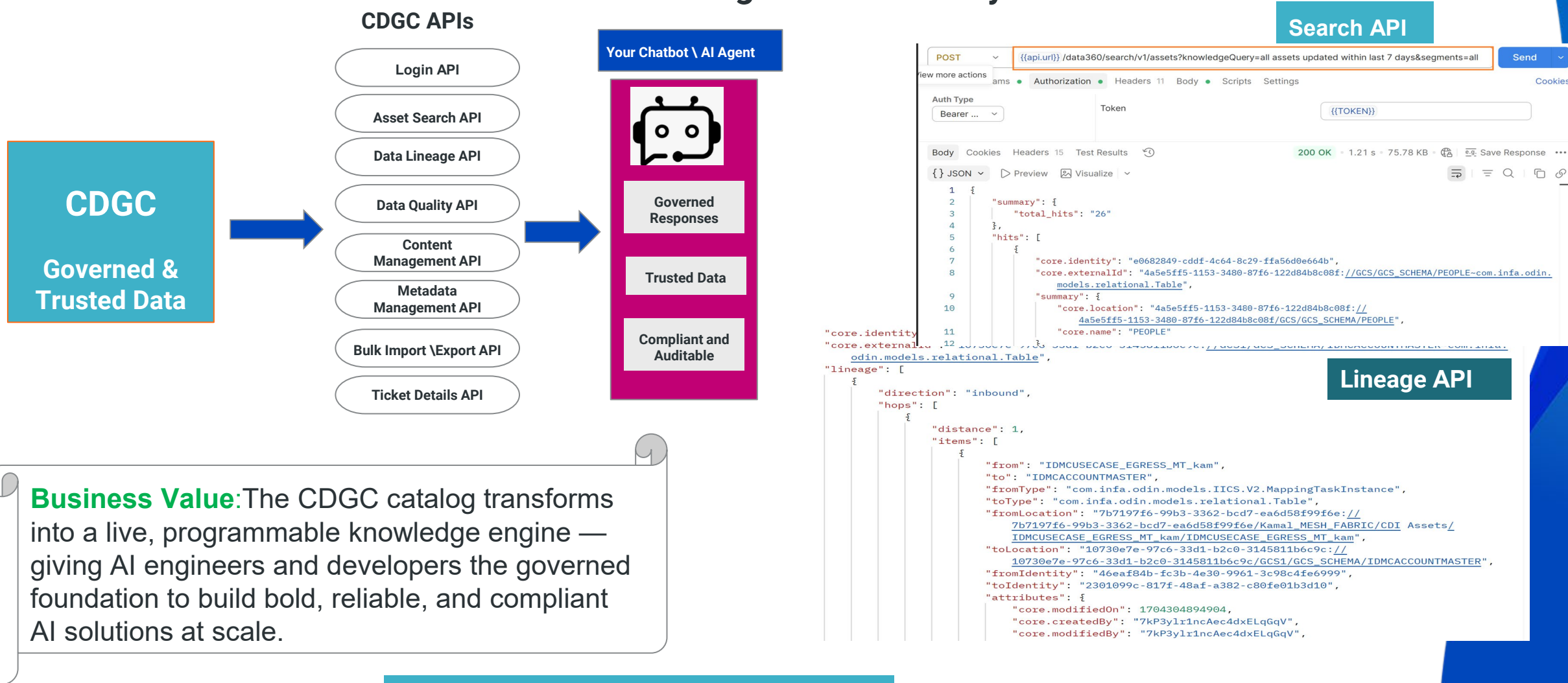
auto_awesome

Catalog Enrichment server

Programmatically enrich metadata, apply custom attributes, manage business glossary terms, and automate catalog documentation headless within your system of work.

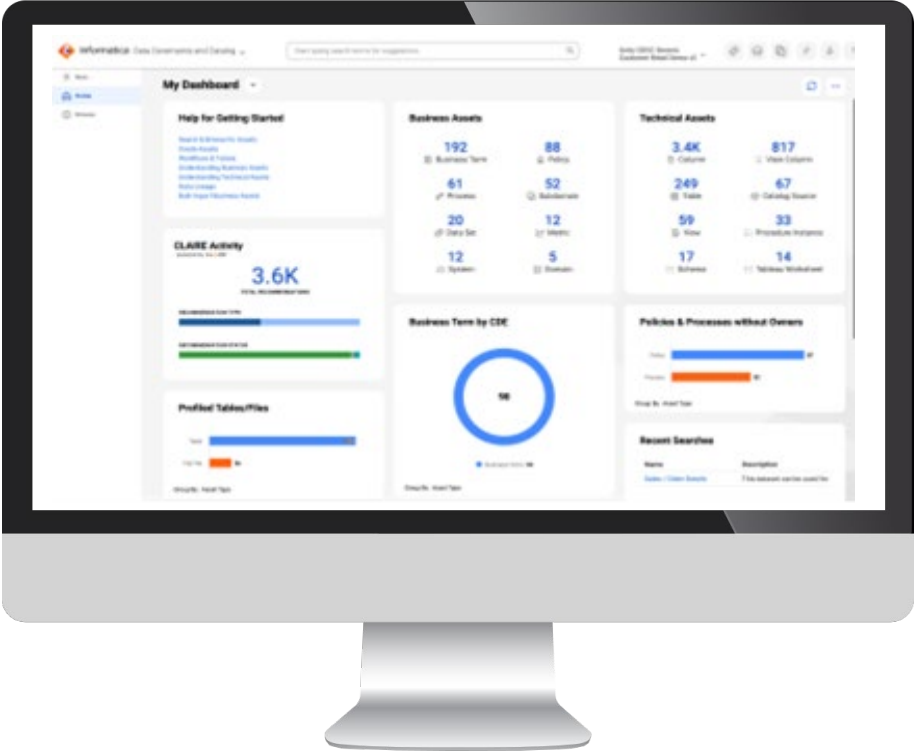
Catalog for AI

CDGC API's - Build Your Own ChatBOT or Agent. Powered by Your Own Data



“ Your catalog. Your AI. Your rules “

DEMO





Informatica®

Thank you