



15 Oct 2024

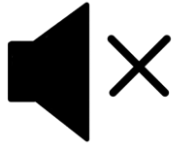
# Unlocking the Power of CLAIRE GPT: Next-Gen AI Solution for Intelligent Data Management

Samiran Karmakar, Sr. Principal Product Manager, R&D

Srinivasa Gopal, Sr. Principal Customer Success Architect, CSA

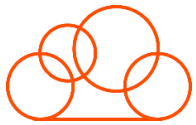
Where data  
& AI come to 

# 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.

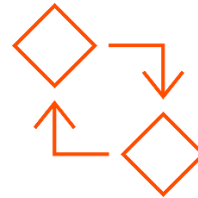
# Feature Rich Success Portal



**Bootstrap trial and  
POC Customers**



**Enriched Customer  
Onboarding  
experience**



**Product  
Learning Paths  
and Weekly  
Expert Sessions**



**Informatica  
Concierge**



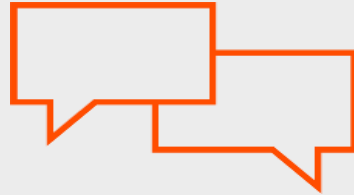
**Tailored training  
and content  
recommendations**

# More Information



## Success Portal

<https://success.informatica.com>



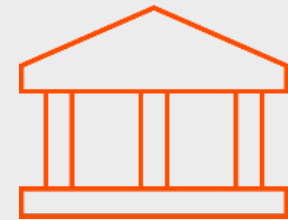
## Communities & Support

<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.

# Agenda

1 Understanding GPT & Business Value

2 CLAI**R**E GPT – Overview

3 Key Capabilities

4 Architecture Overview

5 Product Demo

6 Q&A

# Understanding GPT & Business Value

# What is a GPT?

Google's Pretty Talkative

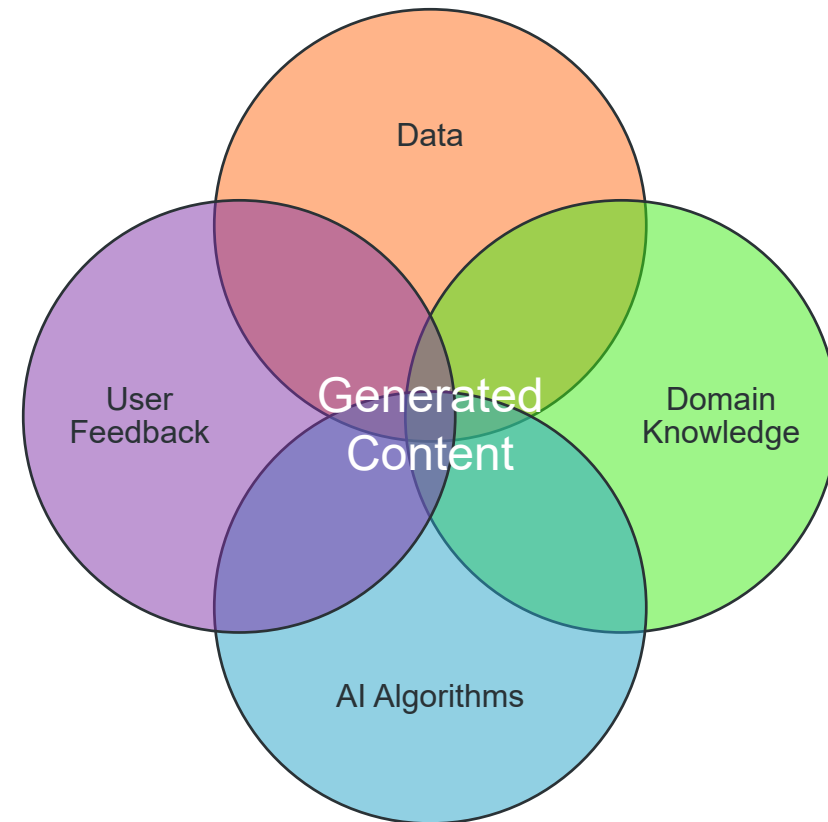
Gigantic Pile of Texts

Great Pizza Tracker

**Generative Pre-Trained Transformer**

## Generative

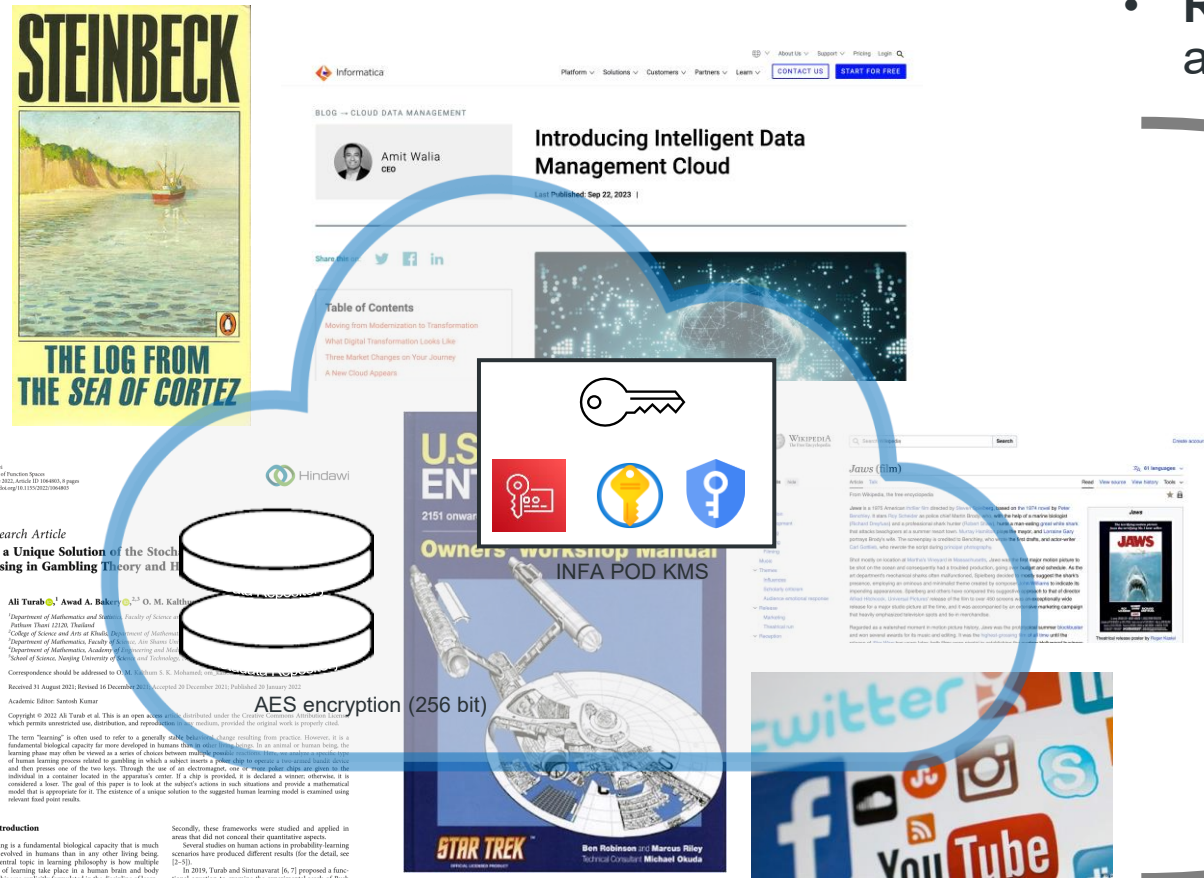
Creating Coherent, Contextually Relevant Content





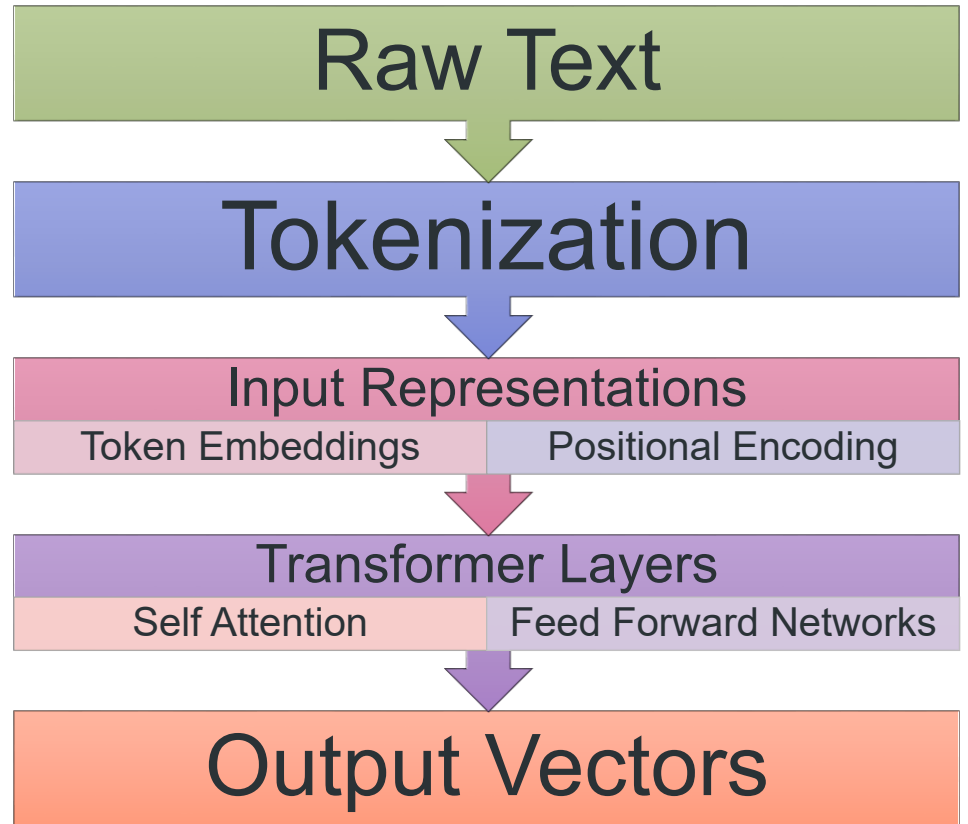
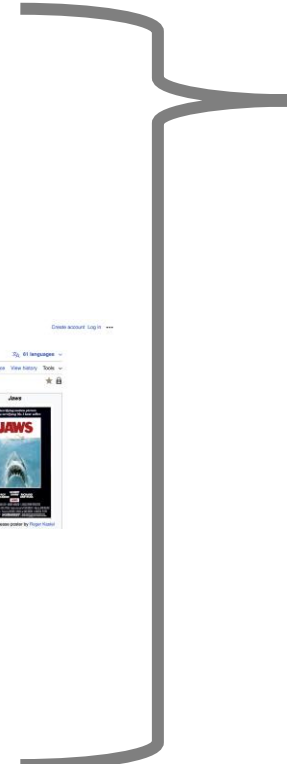
# What is a GPT?

## Generative Pre-Trained Transformer



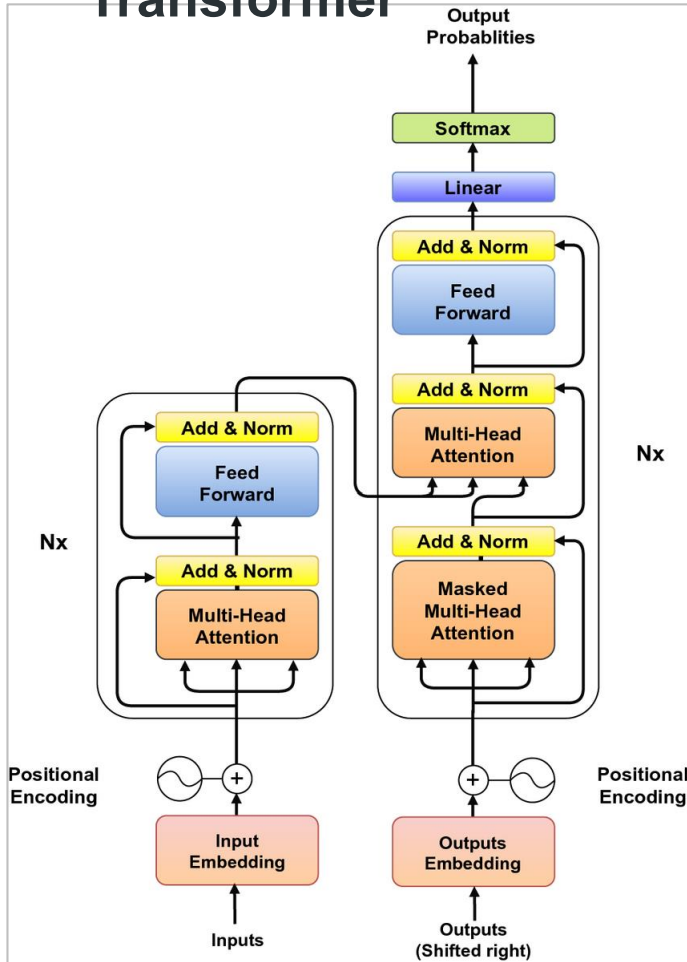
# Pre-Trained

- **Essential for On-Demand Responses:** Impossible to respond to on-demand questions without pre-training
- **Generic LLMs:** Pre-trained on vast amounts of internet data
- **RAG Applications:** Pre-trained on proprietary, authoritative data not available to generic models



# What is a GPT?

## Generative Pre-Trained Transformer



# Transformer

- **Focuses on Relationships:** Transformers analyze sequences to understand connections between tokens
- **Attention Mechanisms:** “Pays attention” to relevant parts of the sequence, not just neighbors
- **Powerful & Efficient:** Handles large amounts of data while processing sequences faster than older models



Key Vector

Everything I've seen so far



Query Vector

What have I seen that's most relevant to me?



Value Vector

The details of what's relevant

Encoder Understands the Input

Decoder Generates the Output

Clever Math

Clever Math

Input Prompt

Output Response

I threw a delicious apple at Srin

I had a delicious apple for lunch

**This Matters When You're Sending Prompts to Claire GPT**



CLAIRE<sup>®</sup> copilot + CLAIRE<sup>®</sup> GPT

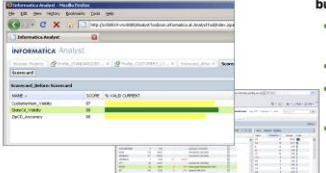
# Why We Built It

## CLAIRE GPT for Data Democratization and Efficiency

### Bridging the Data Literacy Gap

- Data Sprawl & Hidden Insights

**Business Empowerment**  
*Simple-to-use browser-based tools*



Designed for the tasks and skills of business data stewards and analysts

- Purpose-built, web-based UI for fast ramp-up
- Scorecarding & trending
- View business, not technical, representations
- Interact with data directly through profiling, rule validation, and scorecarding

Work with relevant data to meet business needs while reducing reliance on IT

Business Manager ↔ Analyst & Steward

INFORMatica

**Unified role-specific tools for all stakeholders**  
Productive development environment with mid-stream profiling for IT developers



- Full palette of data quality transformations



**“We are hoping it will help us unlock the secrets of our data as we have a lot of hypothesis but lack a clear way to test/prove them.”**

- C-suite, software industry, 1,000 - 5,000 employees



**“I believe [GenAI] will allow us to make better use of the data that we currently have. We have too much data and that impedes meaningful analytics. It's also a drain on resources as well.”**

- Director, software industry, 10,000+ employees

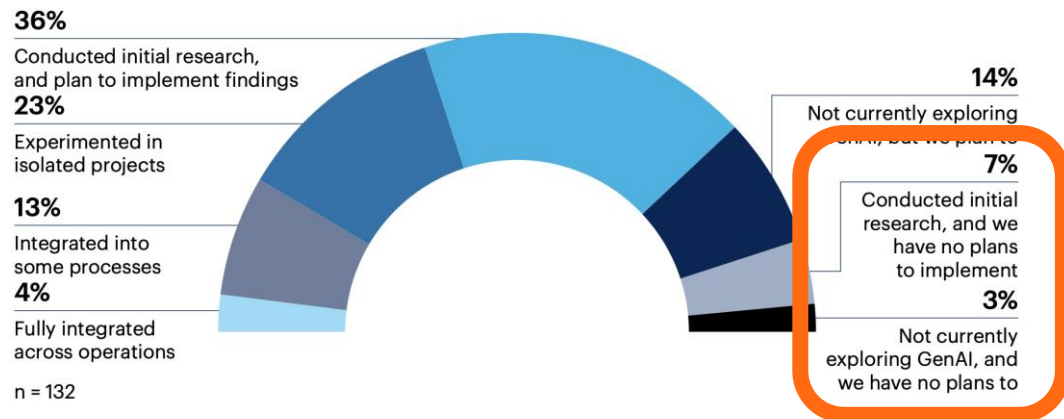
# Why We Built It

## CLAIRE GPT for Data Democratization and Efficiency

### Most respondent D&A functions are in the early stages of GenAI adoption

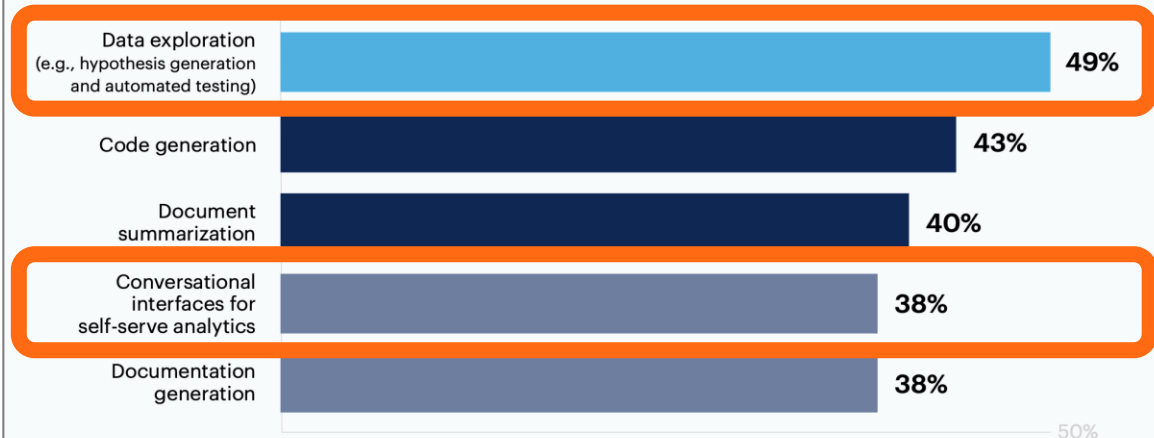
Only **10%** of respondents are **not planning to explore or implement GenAI** for their D&A function. **36%** have conducted initial research and plan to implement their findings, and **40%** are already using GenAI to some extent.

Which of the following best describes your data and analytics (D&A) function's engagement with generative AI (GenAI)?



Among the respondents whose D&A functions are **currently using GenAI** in some capacity (n = 53), the most common applications are for **data exploration (49%)**, **code generation (43%)** and **document summarization (40%)**.

How is your D&A function currently using GenAI? Select all that apply.



Conversational interfaces for data professionals **36%** | Productivity (e.g., generating to-do lists, meeting summaries, etc.) **34%** | Data narrative generation (i.e., contextualized explanations of datasets) **32%** | Synthetic data generation **19%** | Personalized and interactive training for D&A **11%** | I don't know **0%** | Other **0%**

Question shown only to respondents who answered "Experimented in isolated projects," "Integrated into some processes," or "Fully integrated across operations" to "Which of the following best describes your data and analytics (D&A) function's engagement with generative AI (GenAI)?"

Source: Gartner Peer Community, Unleashing the Power of Generative AI for Data and Analytics survey

# CLAIRE<sup>®</sup> GPT

Revolutionize Data Management with the Power of Generative AI



# CLAIRE<sup>®</sup> GPT Business Value

A New Conversational Interface to Data Management for All Data Consumers

## Data Discovery

- Democratize data access
- Conversational Interface for Data Discovery
- Enables Data Accessibility
- Drives Adoption
- Increases Enterprise Data Literacy

## Data Exploration

- Single place to easily understand data that's fit of use
- View data
- Comprehend Data profile
- Explore Data
- Ask Data Questions
- Data Query in Natural Language
- Instant Answer to Data Questions
- Exploratory Data Analysis

## Data Transformation (ELT)

- Transform data within the same Data Lake
- Data Pipeline Quick Start
- NL as the Transformation Language
- Draft ELT Data Pipelines Creation
- Edit and modify as needed in Data Integration

## Product Help

- Increased literacy of IDMC products for data management use cases
- Ask documentation questions
- Product Help
- Usage and best practices

# CLAIRE® GPT– Key Capabilities



# Data Discovery

## Key Highlights

- **Easily find** out relevant data assets, tables, columns, processes, policies etc. in your Catalog using Natural Language prompts
- **Discover** data with specific characteristics such as PII/**sensitive data**, GDPR compliant data, and data marked as critical in Catalog
- Click the asset name to **view** more **details** of the asset in DGC (Data Governance and Catalog)
- Examples:
  - Find catalog sources of type Snowflake, Databricks, and Oracle
  - List business terms marked as critical data elements

**JD** show me the datasets to analyze customer engagement

Absolutely, searching for datasets to analyze customer engagement

Here is the dataset that can be used to analyze customer engagement.

1. **CUSTOMER\_ANALYTICS**: This dataset contains information about customer profile, transactions, and demographics. It is part of the **AnalyticsReadyDW** resource in **Snowflake**. The table has all the data attributes required to analyze customer engagement.

Here are the datasets that can be used to analyze customer engagement. They are spread across different resources and databases.

1. Datasets in **Tableau**:
  - **Customer Retention**: This dataset is part of the **Tableau\_sales\_kpis\_Tableau** resource.
  - **Customer Churn**: This dataset is also part of the **Tableau\_sales\_kpis\_Tableau** resource.
2. Datasets in **Azure Blob Storage**:
  - **CustomersEast.csv**: This dataset is part of the **Customer Transactions Lake** resource.
3. Datasets in **Azure Data Factory**:
  - **Customers**: This dataset is part of the **Customer Transactions Data Factory** resource.
4. Datasets in **SqlServer**:
  - **customers**: This dataset is part of the **Customer Transactions DWH** resource.
  - **Customer**: This dataset is part of the **Adventure Works OLTP** resource.
  - **Customer**: This dataset is part of the **Banking Cloud Sales Data** resource.
5. Datasets in **Oracle**:
  - **CUSTOMERS\_VIEW**: This dataset is part of the **ora\_retail\_till** resource.
  - **CUSTOMER**: This dataset is part of the **ora\_retail\_till** resource.

👍 🗨

**DT** List business terms marked as critical data elements

Sure, listing business terms marked as critical data elements

Here are the business terms marked as critical data elements.

1. **Marital Status**: This term refers to the marital status of an individual, such as married, single, or divorced.
2. **Age**: This term refers to the age of an individual.
3. **Contract Type**: This term refers to the contract type, such as Fixed Term, Temporary, or Permanent.
4. **Promotion Key**: This term refers to the primary key of the Promotion table.
5. **Total Cost**: This term refers to the total cost.

👍 🗨

# Metadata Explore

## Key Highlights

- **Search** and explore the **metadata** of data assets that you discover
- **Explore**, view **and understand** business terms, definitions, DQ scores, data owners/stakeholder
- Quickly understand the lineage and dependencies of data assets
- Click the asset name to **view** more **details** of the asset in DGC (Data Governance and Catalog)
- **Examples:**
  - Show me the overview of FACT\_SALES
  - Show me the DQ Score of Employee
  - Find the data lineage of CUSTOMER\_ANALYTICS

**DT** Overview of @Sales KPIs

Sure, providing an overview of Sales KPIs

### Sales KPIs Overview

The Sales KPIs is a **Dashboard** type asset, housed in a **Tableau** environment. Unfortunately, it has not undergone profiling to assess its quality and structure comprehensively.

### Key Components Overview

#### Stakeholdership

- **No specific stakeholdership information is available:** It's unclear who owns or uses this asset, which could impact its management and strategic planning.

#### Data Characteristics

- Resource Type: Tableau
- Profiled: No
  - This indicates that the asset has not undergone a comprehensive assessment of its structure and quality.

#### Data Quality

- **No data quality metrics are available:** This suggests that the asset has not undergone a data quality process, which could affect its reliability and usefulness.

Please note that the absence of data profiling and data quality information indicates a potential area for improvement in the management and use of this asset.

TABLEAU DASHBOARD | Sales KPIs  
There is no description.  
Tableau\_sales\_kpis\_Tableau / ... / Cloud DG demo - Sales KPIs Stakeholders: - Last Updated: 01/04/2024, 09:28 PM

**JD** show me the lineage of @DIM\_CUSTOMER

Absolutely, showing the lineage of DIM\_CUSTOMER

- The lineage of **DIM\_CUSTOMER** involves a total of **5 catalog sources**.
- The data flow begins from two sources: **iics\_retail\_dataFlow** and **snowflake\_retail\_dwh**, which collectively include **4 unique assets** such as **mt\_m\_onlineSales\_02\_s3\_landing\_to\_enrichment** and **DIM\_CURRENCY**.
- This data then flows into **DIM\_CUSTOMER** and proceeds towards two target resources: **Tableau\_sales\_kpis\_Tableau** and **snowflake\_retail\_dwh**, covering another **8 assets** including **Product List** and **V\_TEMP\_FACTSTORESALES\_DETAILED\_EUROPE**.

### Lineage of DIM\_CUSTOMER Table

Download

# Data Explore

## Key Highlights

- Explore data assets with natural language questions to **understand** KPI, spot **trends**, find out top **customers**, high-performing **products etc.** without the knowledge of SQL
- **Empower** non-tech **business** users to find out and understand data **insights** on their own by abstracting the SQL layer
- Examples:
  - Show me data sample of 100 rows from @CUSTOMER\_ANALYTICS
  - Show me the total number of orders across cities
  - Show me the email, first name, last name where discount is 15%

ST Show data preview of @ORDERS

Absolutely, previewing data for ORDERS

TABLE | ORDERS  
SnowflakeScan\_Copy\_New\_Sales / PLATFORMQA / SALES Columns: 8

Privacy guidelines prevent us from storing sample data, but you can download the data for your reference.

#	REQUIRED_D...	ORDER_DATE	# ORDER_ID	# SHIPPED_DA...	CUSTOMER_...	STAFF_ID	#
6		Oct 05, 2021	1	1	Oct 10, 2021	Oct 08, 2021	2
23		Oct 06, 2021	2	1	Oct 11, 2021	Oct 09, 2021	3
11		Oct 07, 2021	3	1	Oct 12, 2021	Oct 10, 2021	5
14		Oct 08, 2021	4	1	Oct 13, 2021	Oct 11, 2021	6
30		Oct 09, 2021	5	1	Oct 14, 2021	Oct 12, 2021	8
7		Oct 10, 2021	6	1	Oct 15, 2021	Oct 13, 2021	9
18		Oct 11, 2021	7	1	Oct 16, 2021	Oct 14, 2021	1
9		Oct 12, 2021	8	1	Oct 17, 2021	Oct 15, 2021	2
26		Oct 13, 2021	9	1	Oct 18, 2021	Oct 16, 2021	3
2		Oct 14, 2021	10	1	Oct 19, 2021	Oct 17, 2021	5

Viewing sample data: 30 rows and 2 data types

Show Code Download

ST Identify the staff members who managed stores located in 'London' or 'Santa Cruz'

Sure, identifying the staff members who managed stores located in 'London' or 'Santa Cruz' using STAFFS

TABLE | STAFFS  
SnowflakeScan\_Copy\_New\_Sales / PLATFORMQA / SALES Columns: 2

Privacy guidelines prevent us from storing sample data, but you can download the data for your reference.

A FIRST_NAME	A LAST_NAME
Fabiola	Jackson
Mireya	Copeland
Genna	Serrano
Virgie	Wiggins

Viewing sample data: 4 rows and 1 data types

Show Code Download

ST Can you provide me the shipping date for the order ID 5?

Absolutely, finding the shipping date for the order ID 5

Oct 12, 2021

Show Code

# Create ELT Pipeline

## Key Highlights

- **Generate** new **data products** by creating ELT pipelines using natural language on data warehouses like Snowflake, AWS Redshift, Azure Synapse, Google Big Query, Databricks Delta Lake and more
- Empower **Citizen Data Integrators** to quickly build **first drafts** of ELT mapping
- Open the mapping in Data Integration apps and execute the mapping
- Examples:
  - Create a mapping to combine Customer and Products Datasets

show me the costumers from the state CA, NY, FL and TX

Fetching the customers from the states ca, ny, fl, and tx from customer

TABLE | CUSTOMER  
Snowflake\_RetailScan / PLATFORMQA / SNOW\_NEXGEN\_RETAIL Columns: 10

Privacy guidelines prevent us from storing sample data, but you can download the data for your reference.

A	Address	A	City	A	Country	A	Party_ID	A	Email_Addre...	A	First_Name
56579	Annamar...	San Diego	USA	C984	tshuqgrb@theq...	Trumann					
77086	Warbler C...	Cleanwater	USA	C972	tmourqueqz@a...	Tiena					
04	Rutledge Jun...	Modesto	USA	C928	trickardspr@fee...	Timothee					
831	Wayridqe P...	Miami	USA	C936	zreappz@china....	Zeke					
3538	Hallows Pa...	Sarasota	USA	C911	zmowsdellpa@L...	Zacharias					
567	Pine St	Miami	USA	C1005	william.gonzalez...	William					
933	Susan Place	Lehigh Acres	USA	C848	whanmoreni@q...	Waneta					
9	Ongard Lane	Tallahassee	USA	C988	rwolstencrotrf...	Rahel					
9	David Alley	Fort Worth	USA	C947	rkaaskooperaa...	Ralph					
9	Northridge Str...	Jamaica	USA	C931	rmbaypu@trip...	Rebecka					

Combine CUSTOMER with ORDERS Create a mapping for the above

Create a mapping for the above

Creating a mapping for CUSTOMER

If you want to save the proposed mapping, click Open in Data Integration and then save the mapping.

m\_Snowflake\_src\_PLATFORMQA\_SNOW\_NEXGEN\_RETAIL\_ORDERS\_1\_1\_tgt\_TARGET\_TABLE\_CUSTOMER\_ORDE...

```
graph LR; src1[src_PLATFORMQA_SN...] --> jnr[jnr_src_PLATFORMQA...]; src2[src_PLATFORMQA_SN...] --> jnr; jnr --> tgt[tgt_TARGET_TABLE_CU...];
```

Open in Data Integration

# Ask Product Help Questions

## Key Highlights

- **Ask for help** on any products or services offered by Informatica
- Find out responses to your questions by **leveraging** info **available** in the INFA **documentation** portal, KB articles, and How-to-articles
- Click on the citations provided to **find out more** details
- Examples:
  - How do I create a user in IICS?
  - How do I create a business term in Data Governance and Catalog?

JD how do I create a user in IICS

Showing how to create a user in IICS

To create a user in Informatica Intelligent Cloud Services (IICS), follow these steps:

1. Log into your IICS account as an administrator.
2. On the left side of the screen, click on "Administrator," then select "Users" from the dropdown menu.
3. Click on the "Add User" button located at the top right corner of the page.
4. Fill out the required fields, such as user name, first name, last name, email, authentication method, and roles or user groups you want to assign to the new user.
5. Click on "Save" to complete the process and create the new user.

Remember that depending on the authentication method you choose, the user may need to verify their email address or complete other steps before they can access IICS.

Citations:

1. [Data Integration/Users-Users-Creating a user](#)
2. [Administrator/Users-Users-Creating a user](#)

Like and Share icons

JD how do I configure a Catalog Source in MCC

Showing how to configure a Catalog Source in MCC

To configure a catalog source in Metadata Command Center (MCC), follow these steps:

1. Log in to Informatica Intelligent Cloud Services and select Metadata Command Center from the My Services page.
2. Click on the "New" button located at the top right corner of the page.
3. Select a source from the list of predefined source systems available in Metadata Command Center.
4. Click on "Create" after selecting the source system.
5. On the Registration page, enter a name and an optional description for the catalog source.
6. In the Connection Information area, select a connection for the catalog source from the Connection list. This list will contain all the connections that your organization administrator has created in Administrator.
7. Test the connection to ensure it's successful.
8. Click on "Next" to proceed to the Configuration page.
9. Configure the catalog source as per your requirements by choosing options available on the Configuration page.
10. Assign connections to reference catalog sources if necessary, so that you can link reference assets to actual objects in the source system and view complete lineage.
11. Optionally, set a schedule to run your catalog source.
12. After completing all details, click on "Save."

By following these steps, you will successfully configure a catalog source in Metadata Command Center.

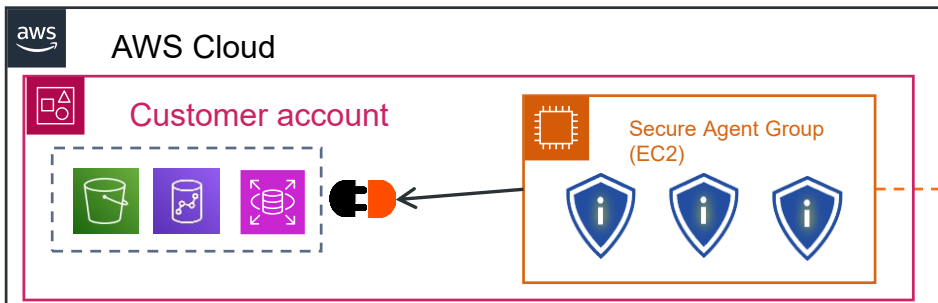
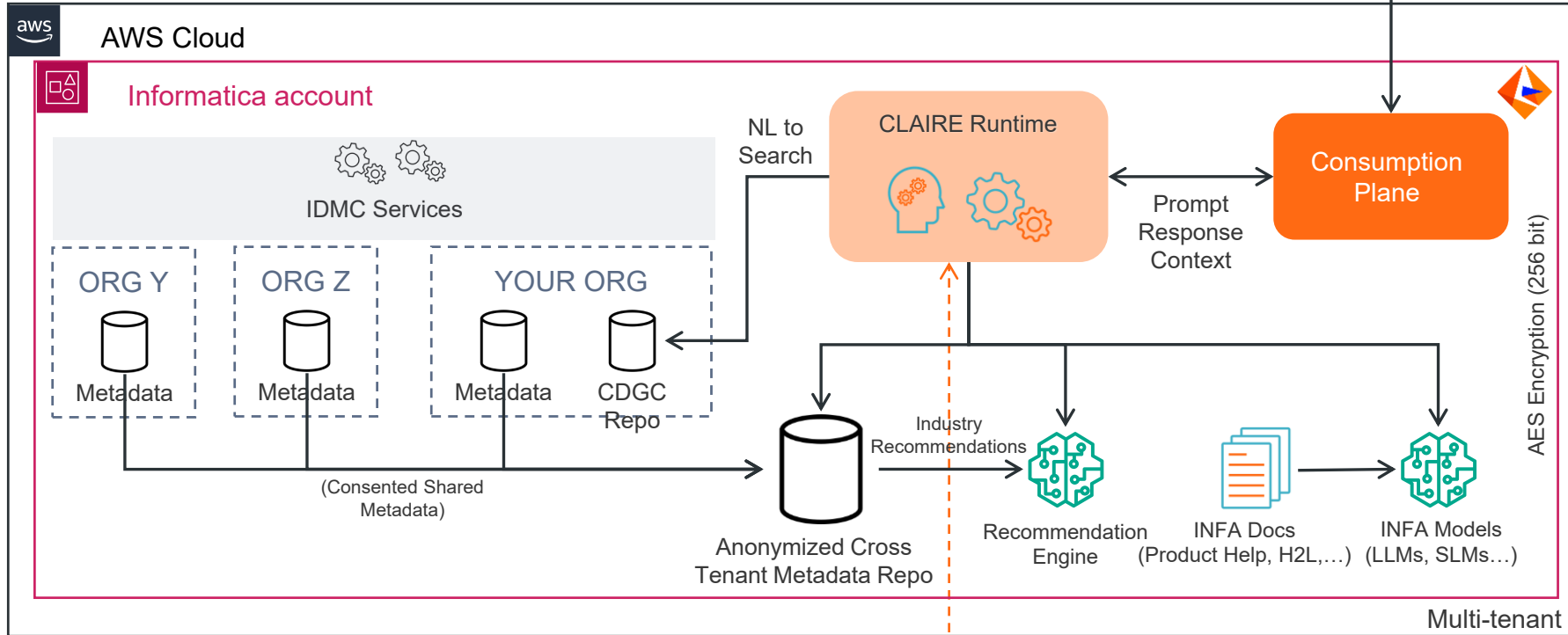
Citations:

1. [Metadata Command Center/Creating A Catalog Source-Creating A Catalog Source-Step 1. Register a catalog source](#)
2. [Metadata Command Center/Referenced Catalog Sources-Referenced Catalog Sources-Assign connections to reference catalog sources](#)
3. [Metadata Command Center/Manage Catalog Sources-Manage Catalog Sources-Creating a catalog source](#)

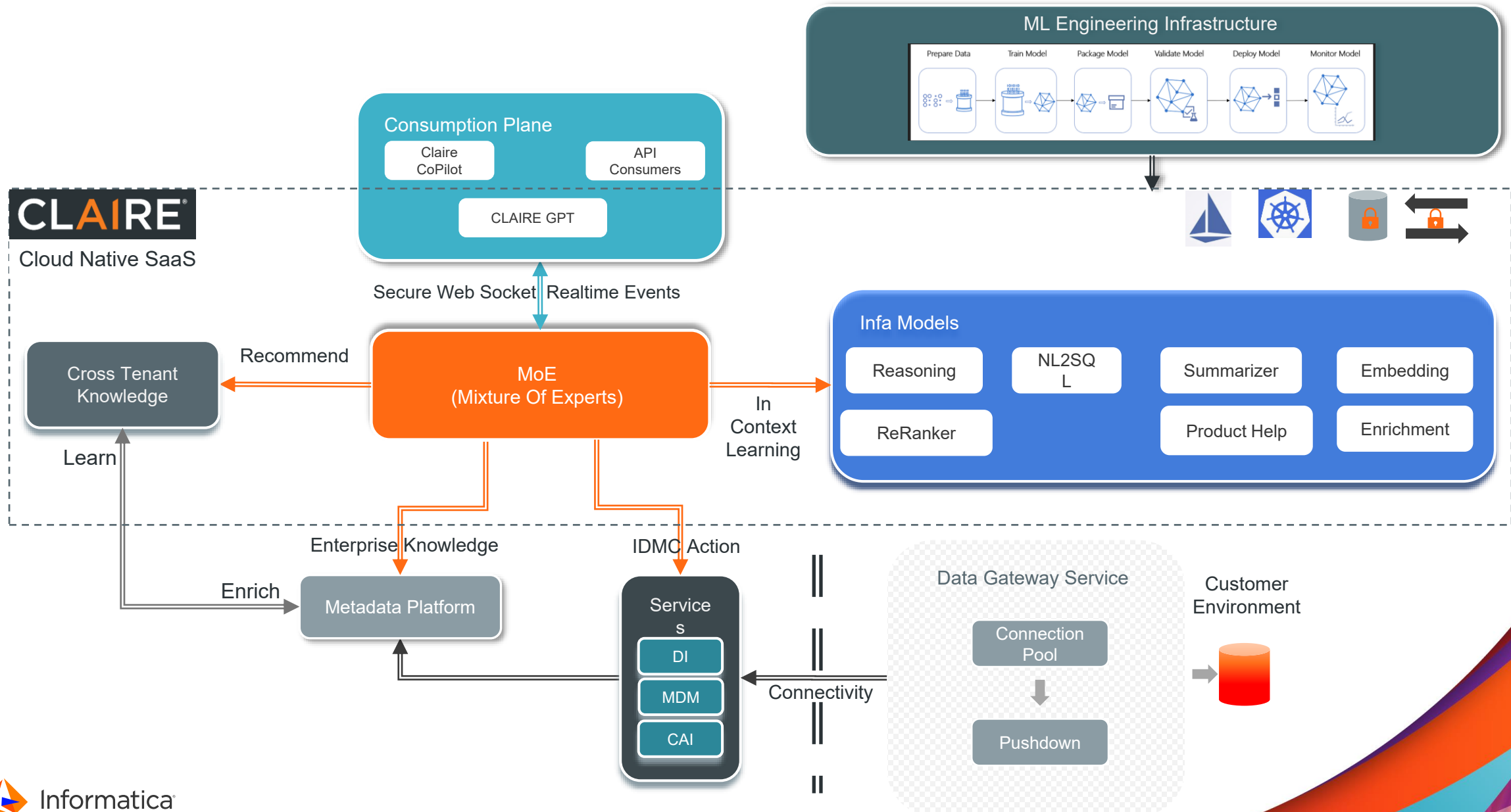
Like and Share icons

# CLAIRE® GPT–Architecture Overview

# CLAIRE GPT Service Architecture

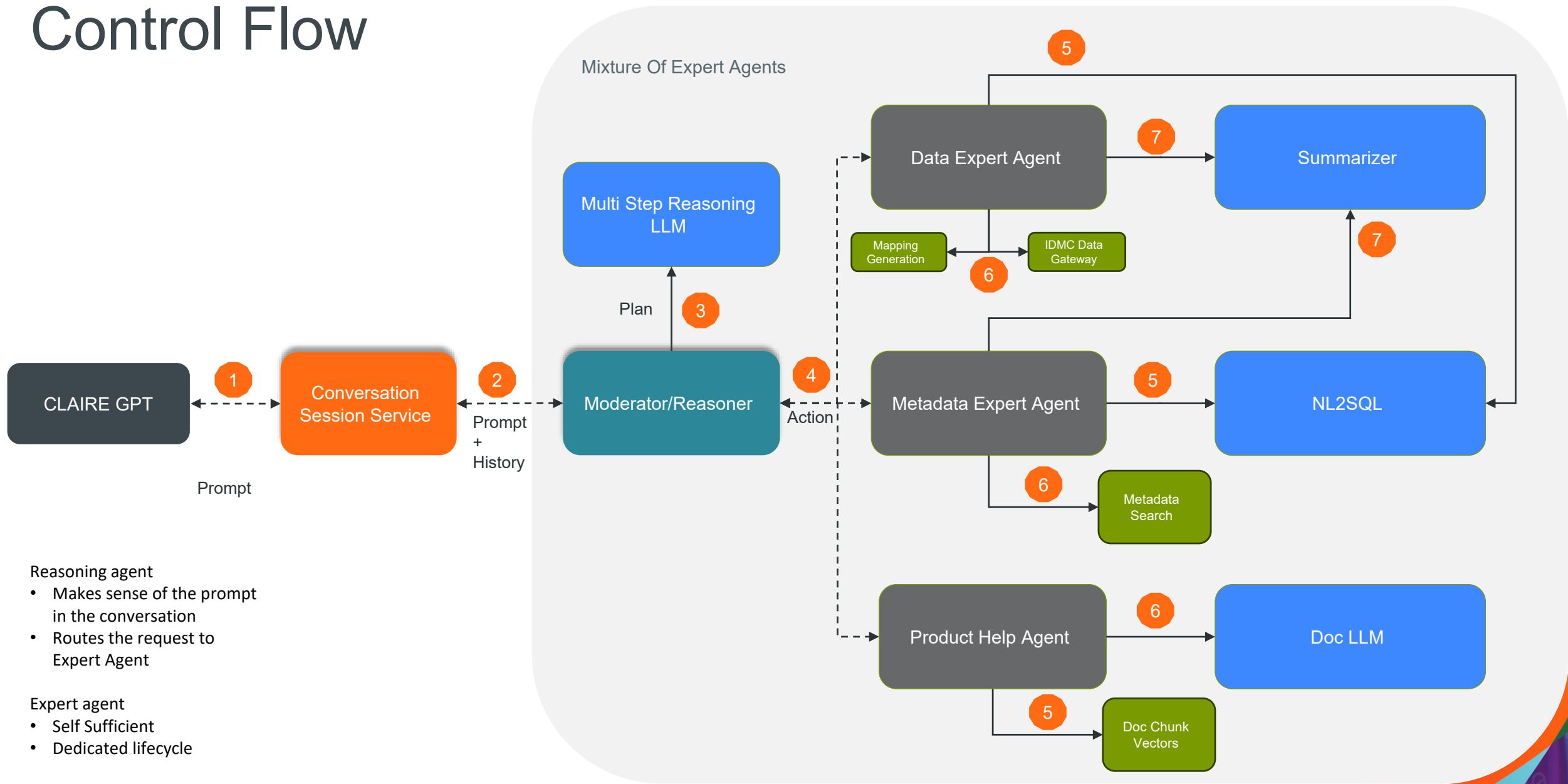


# Architecture





# Control Flow



## Reasoning agent

- Makes sense of the prompt in the conversation
- Routes the request to Expert Agent

## Expert agent

- Self Sufficient
- Dedicated lifecycle

# CLAIRE GPT Considerations

- CDGC Repository is a mandatory pre-requisite to enable CLAIRE GPT service, since it works as the main tenant-only metadata input and foundation
- CLAIRE GPT service runs entirely on Informatica PODs, which means it will not have any impact on customers' secure agents. Only when users perform data exploration tasks, these will be directed to Secure Agents for execution (since only Secure Agents do have access to customer data endpoints)
- Data exploration tasks will only fetch data in real-time (as described), but this data will not be persisted in any repository, only used in the user session on CLAIRE GPT while this is open
- CLAIRE GPT service will honor platform RBAC model and permissions, i.e. user will only be able to perform data exploration over connections they have permissions for
- Users can provide feedback back to the CLAIRE GPT application on the responses they are getting from it. This feedback will be used for learning areas of improvement (if you opted-in for metadata sharing)

# DEMO

CLAIRE GPT: Talk to Your Data



# Business Analysis & Revenue Improvement

## Scenario

- Role: Business Analyst at XYZ Retail Ltd.
- Objective: Understand Customer Behaviour, Analyze Past Orders, and Recommend how to Improve Revenue
- Strategy: Use data to segment customers, identify top performing products, and underperforming categories
- Data Sources: [Amazon S3](#): Customer Demographics, [Oracle](#): Customer Portfolio, [Snowflake](#): Customer Master, Products, Orders

## Approach



# SSM Health

Large Not-for-profit Healthcare Organization

“CLAIRE GPT can **strategically** be deployed to **help** our clinical teams find information that would historically require IT ticket requests. For example, our Chief Medical Information Officers will be able to **easily find** how many orthopedic **providers** are in network to ensure that the health system is **meeting** patient needs while having the **appropriate coverage**.”

~Katie Germano, Director - Healthcare Analytics at SSM Health



# Deloitte Consulting

Innovating Customers with Generative AI

"With CLAI<sup>R</sup>E GPT, our joint customers can benefit from navigating through vast pools of information, rapidly find trusted data, unveil hidden insights, and illuminate the intricate relationships between data points. We believe that our joint customers can accelerate their journey to knowledge and driving business value with the easy-to-use natural language interface for all aspects of data enablement and management on a single platform."

~Abhiman Matlapudi, Specialist Leader | Informatica AI Lead at Deloitte Consulting LLP"

# Deloitte.

**MAKING AN  
IMPACT THAT  
MATTERS**



# Thank You

Where data & AI come to **LIFE**



# Want To Learn More?

## Helpful Links

### Videos

- [CLAIRE GPT 2-minute Demo](#)
- [CLAIRE GPT Overview Video](#)
- [Getting Started with CLAIRE GPT Video](#)
- [CLAIRE GPT Click-Thru Demo](#)

### Content

- [EU AI Act Blog](#)
- [CLAIRE GPT Blog](#)
- [CLAIRE GPT Data Sheet](#)
- [Informatica AI for Data Management Blog](#)
- [Getting Started with CLAIRE GPT PDF \(Ungated\)](#)
- [Getting Started with CLAIRE GPT PDF \(Gated\)](#)
- [CLAIRE Privacy & Security Technical Whitepaper](#)



# Where data & AI come to

