



Enabling Successful AI Initiatives with Modern Data Management

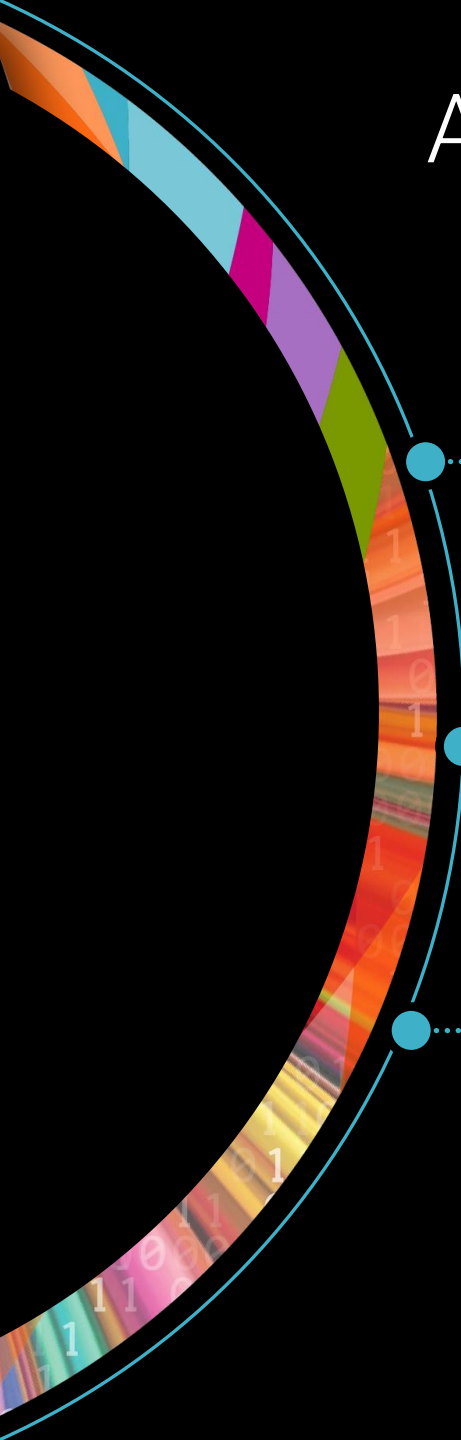
Innovative technical Approaches & Case Studies

Daniel Hein

Chief Architect, Field CTO Office - Asia, Pacific & Japan

**Where data
& AI come to** 

Agenda

- 
- Modern Data Management for AI
 - Emerging Best Practices for Architects
 - AI for Data Management & Data Democratisation

Modern Data Management for AI



"By 2026, more than 80% of enterprises will have used GenAI APIs/models and/or deployed GenAI-enabled applications in production environments, up from less than 5% in 2023."

Source: Gartner® Press Release, "Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026 "October 11, 2023. GARTNER® is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved. <https://www.gartner.com/en/newsroom/press-releases/2023-10-11-gartner-says-more-than-80-percent-of-enterprises-will-have-used-generative-ai-apis-or-deployed-generative-ai-enabled-applications-by-2026>

Data Management & AI

40%

of organizations using AI will invest more in AI overall because of advances in Generative AI

63%

said more than 5 percent of their digital budgets will go to AI use cases going forward

79%

had some exposure to gen AI at work or outside of work, and 22 percent routinely use it in their own work

72%

say that data management is a major obstacle in scaling AI use cases

Is your **data** ready to empower your teams and deliver trusted data for **AI**?

Top Generative AI Challenge

42%

Quality of Data

40%

Data Privacy & Protection

38%

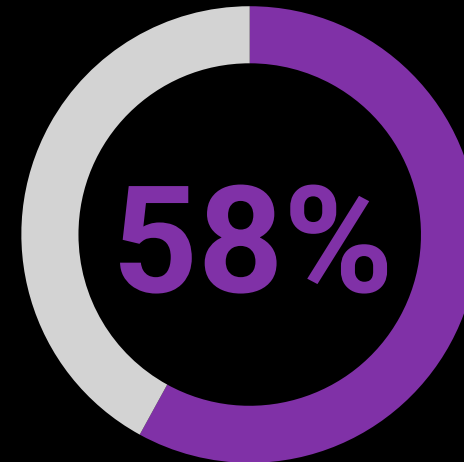
AI Ethics

38%

Quantity of domain-specific data for training & fine tuning of LLMs

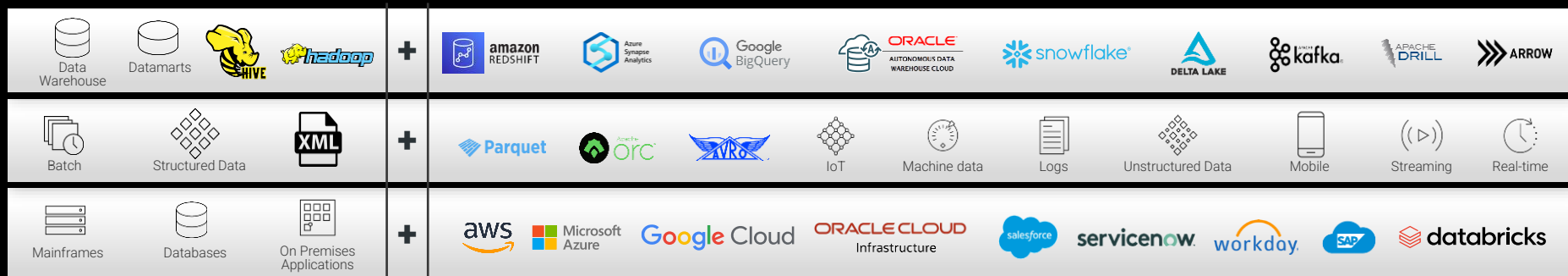
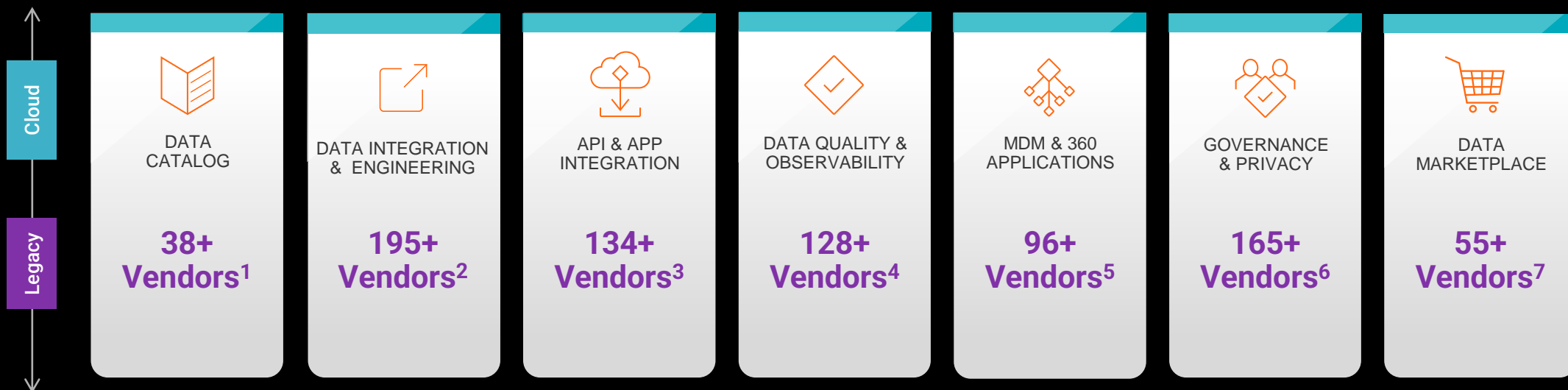
38%

AI Governance



58% of those who predict data management investment will need **five or more** data management tools to support their priorities

Source: [CDO Insights 2024: Charting a Course to AI Readiness \(Informatica\)](#)



¹Trust Radius, Data Catalog Software, May 2023

²Trust Radius, Data Integration Tools Software, May 2023

³Trust Radius, API Management Tools, May 2023

⁴Trust Radius, Data Quality Software, May 2023

⁵Trust Radius Master Data Management (MDM) Tools, May 2023

⁶Trust Radius, Data Governance Software, May 2023

⁷Trust Radius, Data Collaboration Tools, May 2023



Data Quality

Ensuring data accuracy, completeness, and relevance across various sources and systems



Data Security, Privacy, & Governance

Protecting sensitive information and adhering to data privacy regulations



Data Integration

Adapting existing infrastructure and tools to accommodate Gen AI requirements



The Challenges

Example: Data Quality and Reputation

AI is at risk of bias and poor-quality outcomes unless the data is of high quality



MS Chatbot Spew Racist Tweets²

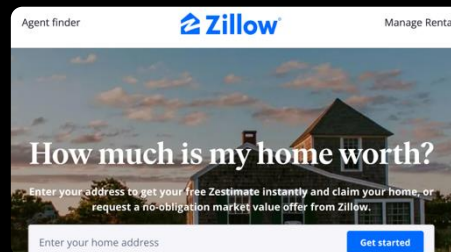
Used uncleansed X (Twitter) interactions for training ML models

Within 16 hrs, 95,000 tweets posted, and rapidly turned overtly racist, misogynist, and anti-Semitic



Bad AI → Attorney Sanctions²

GenAI used to support a new case, at least 6 cases did not exist



Zillow Home Values Corrupted²

Overpaid homes resulted in a \$304M inventory write-down Q3 2021



Data & AI bias

Amazon's AI recruiting tool was found to be biased against women, as it systematically downgraded resumes that included the word "women's," such as in "women's chess club captain."



DPD caused chatbot to swear at customer³

AI used in online chat to answer queries, asked to ruin business reputation

NOTE: [1] "Car Buyer Hilariously Tricks Chevy AI Bot Into Selling A Tahoe For \$1, 'No Takesies Backsies'." (2023), [HotHardware.com](https://www.hothardware.com)

NOTE: [2] "Famous Analytics and AI Disasters." (2023), [CIO.com](https://www.cio.com)

NOTE: [3] "DPD error caused chatbot to swear at customer." (2023), [BBC.com](https://www.bbc.com)

Example: AI Laws in a State of Flux - Overview of the EU AI Act



First EU regulatory framework for AI proposed by European Commission



Risk Based Classification for all AI/ML Algorithms in EU Organisations



Priority that AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly

Businesses must stay abreast of relevant laws and regulations, such as intellectual property rights and data protection, to ensure compliance and mitigate legal risks.

Source – European Parliament News

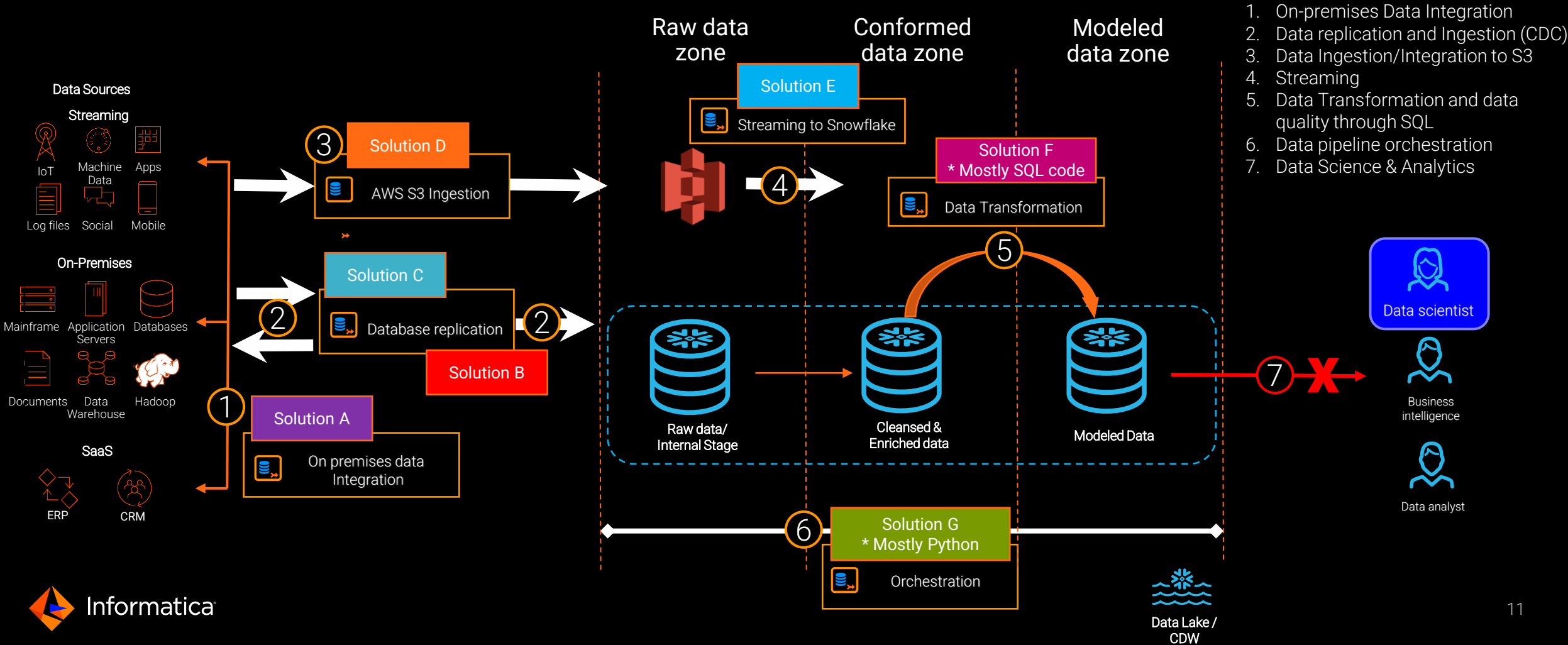
Example: Data Integration at Large Retailer in Australia

7 solutions just to move data around

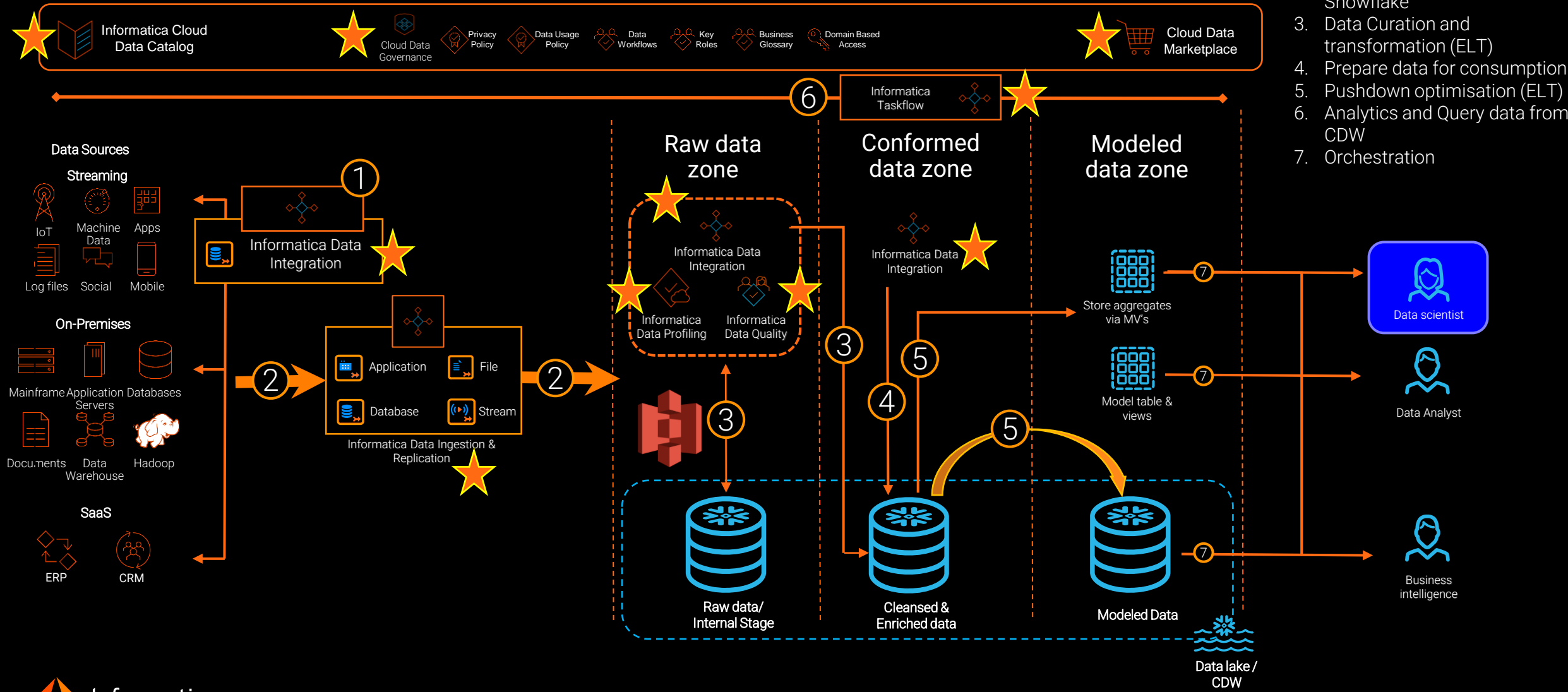
**Inexistent Data Profiling, Catalog and Governance**

**Code-based data profiling & curation**

**No Solution for Data Sharing**

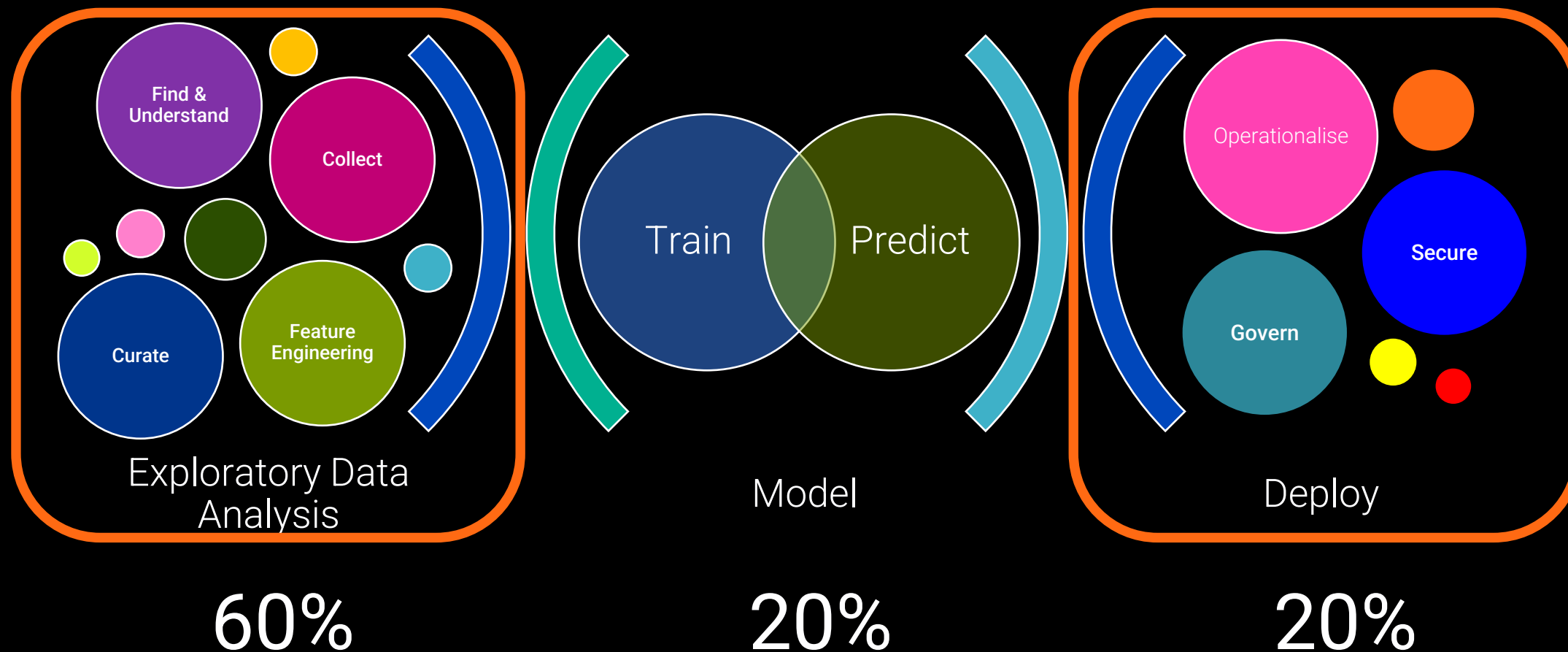


Modern Data Architecture



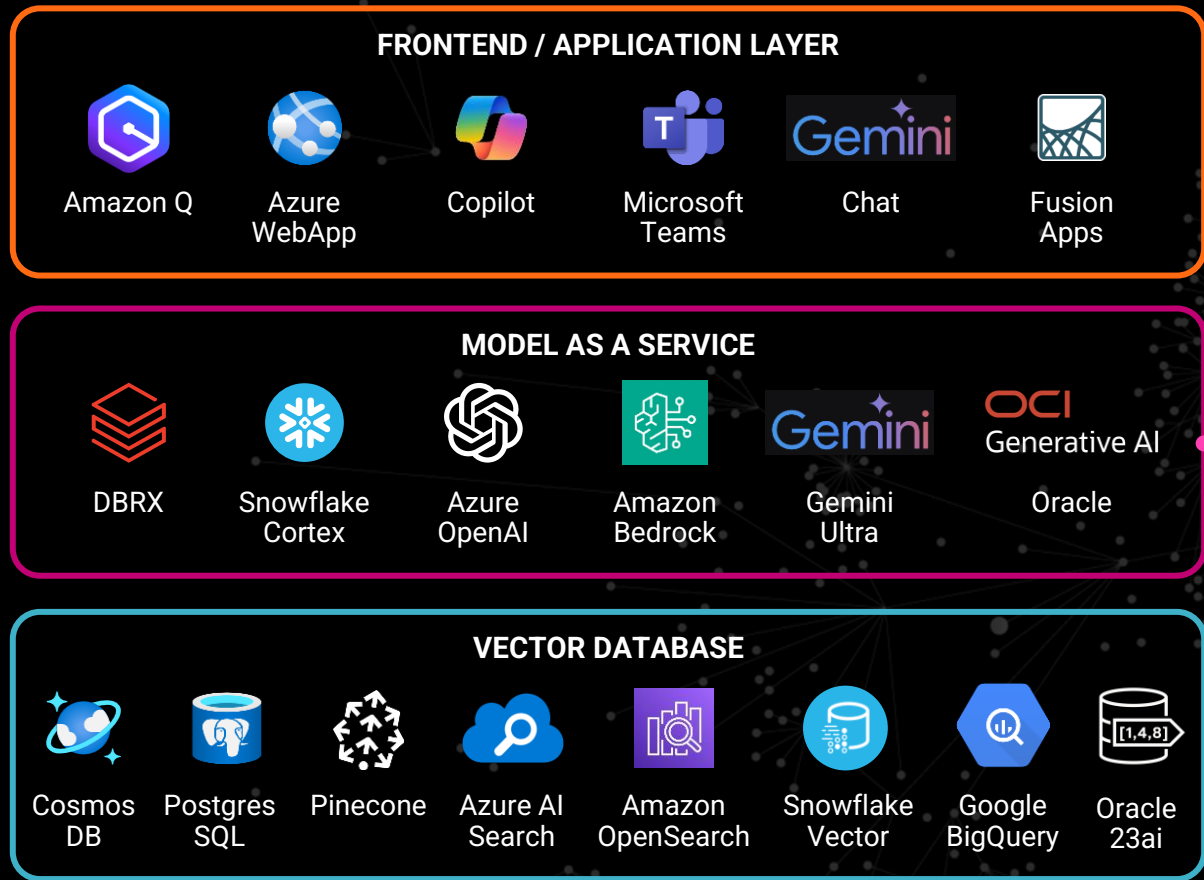
The Data Science Workflow

Streamlined Data Management is the Data Science Supercharger



Making Sense of the Generative AI Landscape

Key building blocks and options cloud ecosystems provide

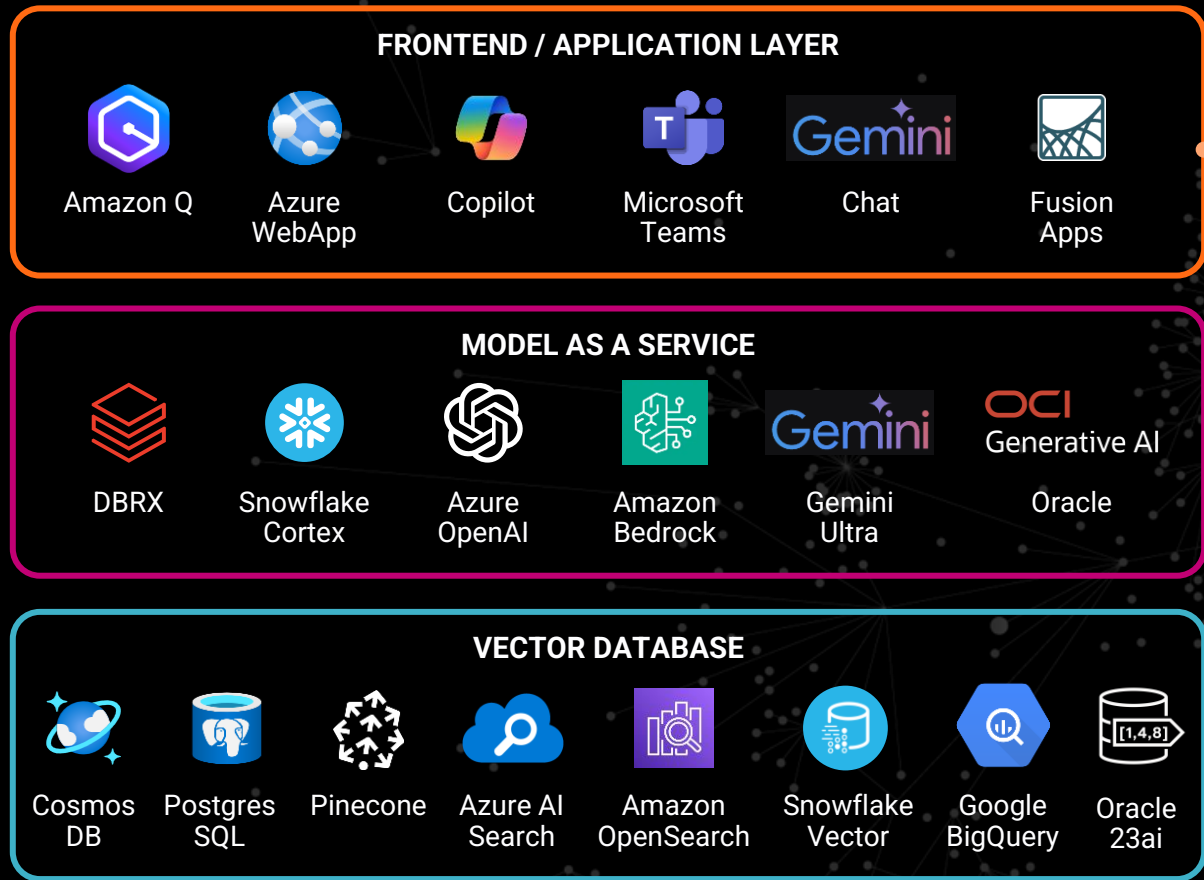


Model-as-a-Service

- There are hundreds and hundreds of open source and commercial, LLM and SLM models available today
- Model-as-a-service provide hosting of models and model management and execution

Making Sense of the Generative AI Landscape

Key ecosystem components

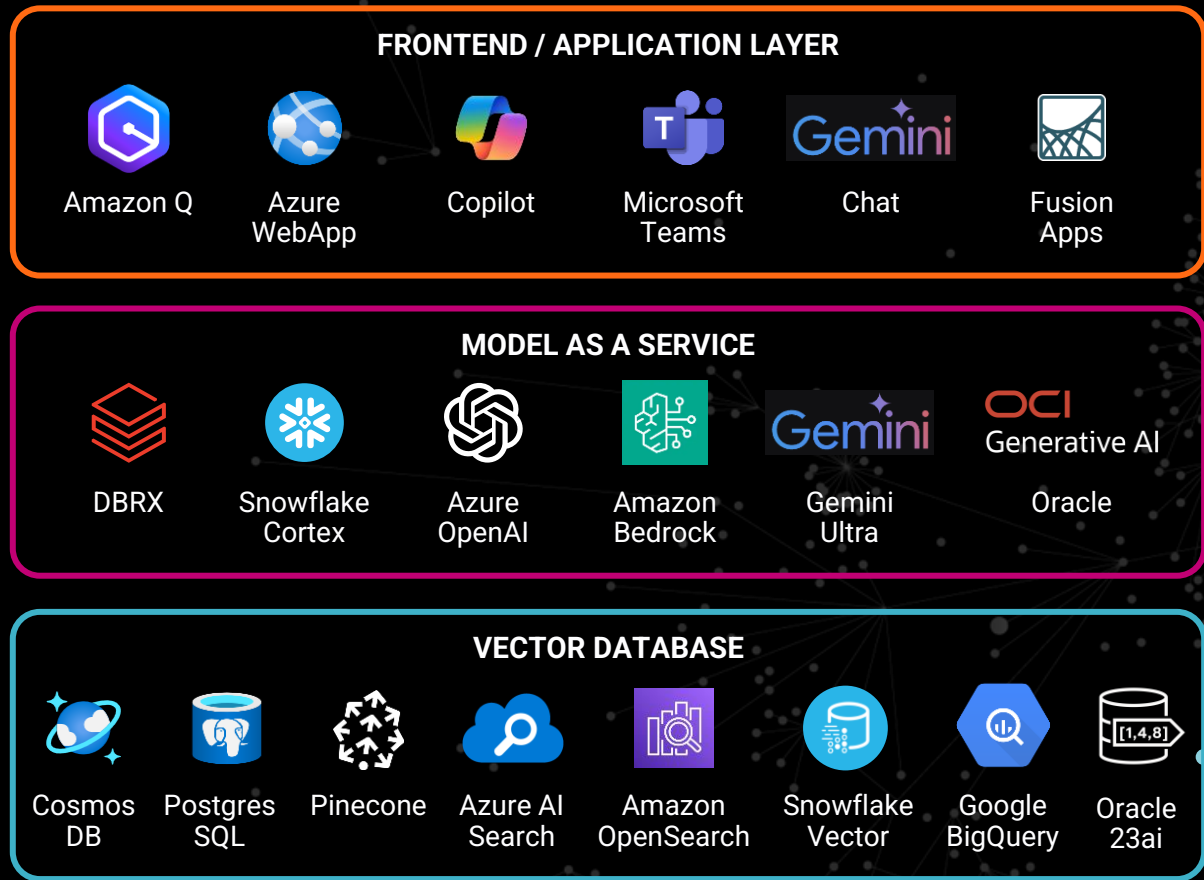


Front End/Application Layer

- Frameworks for building Generative AI applications
- End user collaboration end points that expose Generative AI capabilities to their users (Microsoft Teams, Microsoft Co-pilots, Amazon Q, etc...)

Making Sense of the Generative AI Landscape

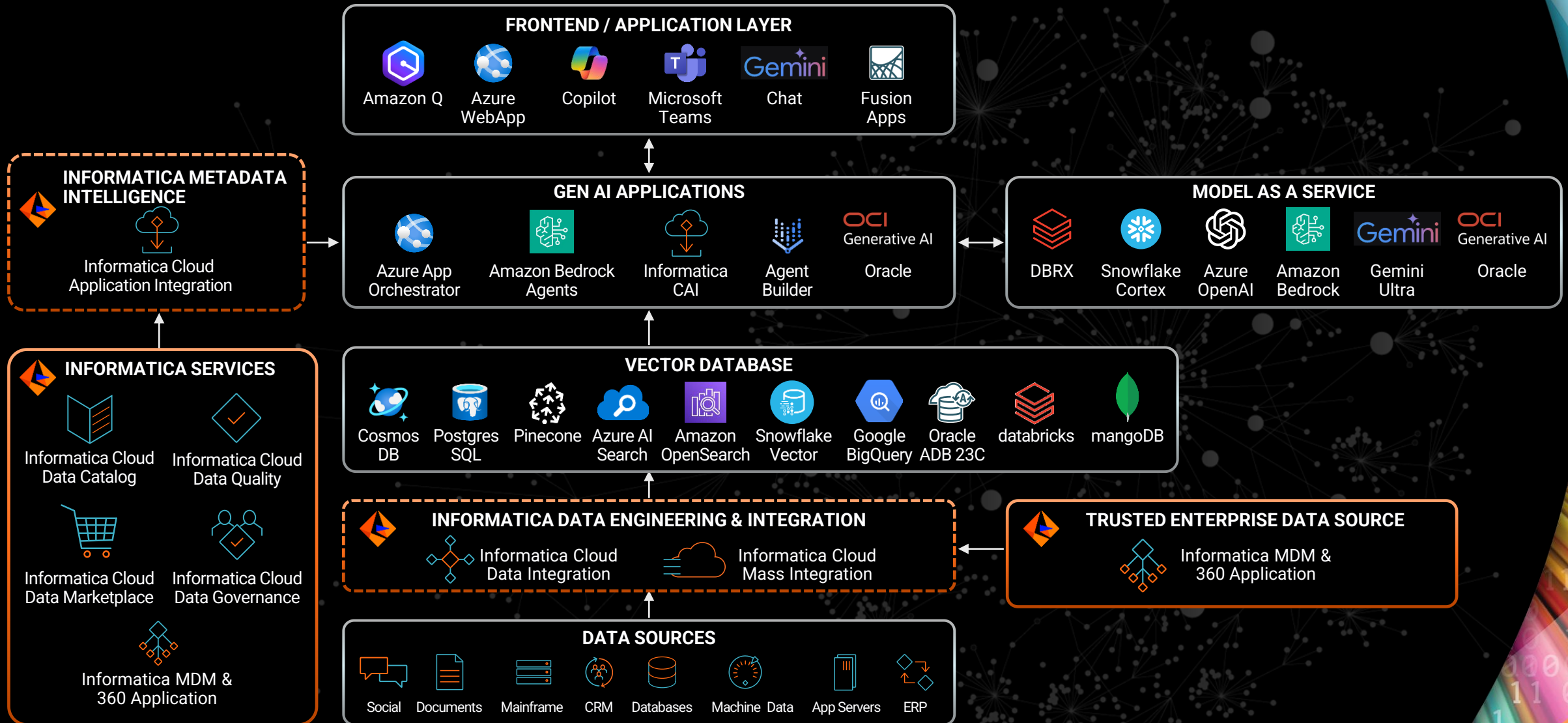
Key ecosystem components



Vector Databases/Vectorization

- Key for RAG use cases to embed other data and services into Generative AI conversational applications
- Stores vectorized form of data to enable vector search to retrieve relevant data as context to inference queries
- Some ecosystems have one or more Vector Databases, others are offering new Vector column types in their database/data lake services

Enabling Enterprise Generative AI Applications



Bringing Enterprise Data to Generative AI

Emerging Best Practices for Architects

Modern Data Architecture & AI Trends in 2024



- Multi-cloud & Hybrid Data Management Platforms
- Data Products
- RAG & AI Agents
- AI Governance



- Modern Data Stack
- Data Mesh

State of Enterprise Data for Gen AI

Grounding Gen AI in Enterprise Data is important

Internet Data

- Majority of data that LLMs are trained on is publicly available, internet data
- Unstructured and Semi-Structured
- Optimised for Search and Retrieval (SEO)
- Accessible to anybody with an internet connection

Enterprise Data

- Estimated 90% of the world's data is private data, owned by organisations
- Enterprises are more experienced with working on structured data
- Not tagged for retrieval
- Subject to role-based access control

Simplified RAG Pipelines

Unstructured Corpus

- Formats: DOCX, PDF, PPT, Email, HTML
- Sources: Sharepoint, File Servers, Data Lakes, Websites, Email Servers

Structured Sources

- APIs, ERP, CRM, MDM, DWH, Lakehouse, Graph Databases

Preprocess

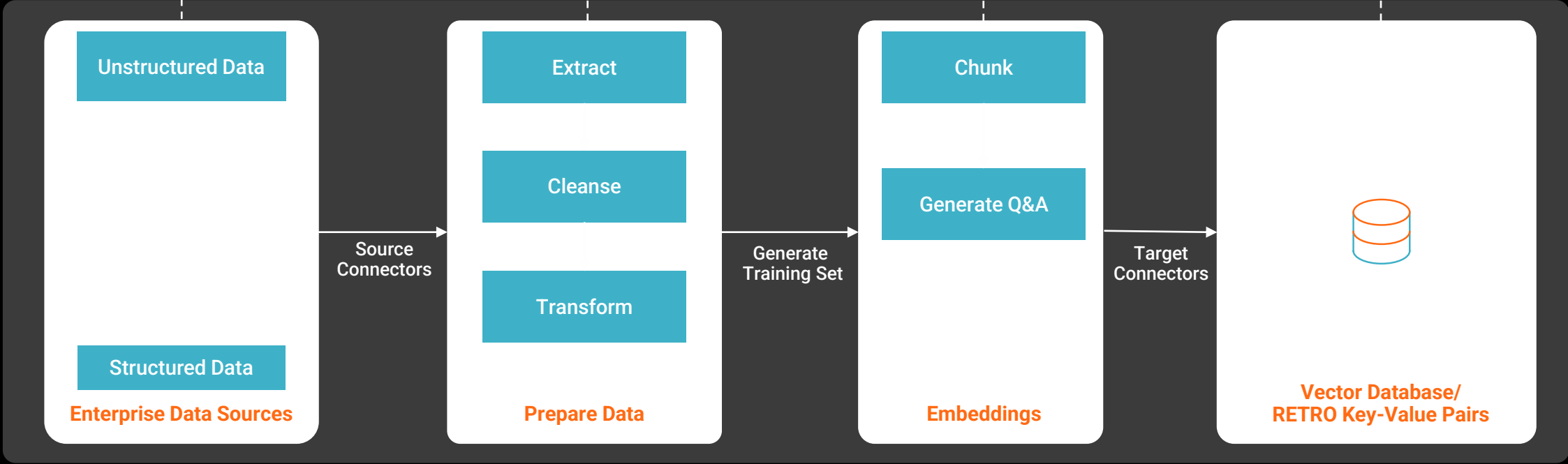
- Extract: Includes metadata, layouts, content
- Cleanse: Remove confidential data, filter spam/ads/irrelevant text, duplicate content
- Transform: Classify, Original Text to JSON

Vectorise

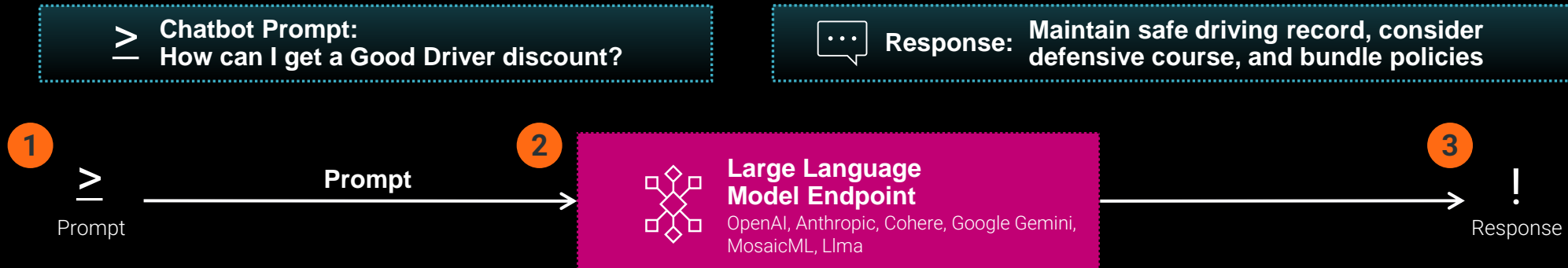
- Chunk: Optional step to convert long documents to manageable chunks
- Embed: Convert text to vectors, using embedding models

Semantic Store

- Vector DB: Pinecone, Qdrant, Elastic
- RETRO: Key Value Stores like Mongo

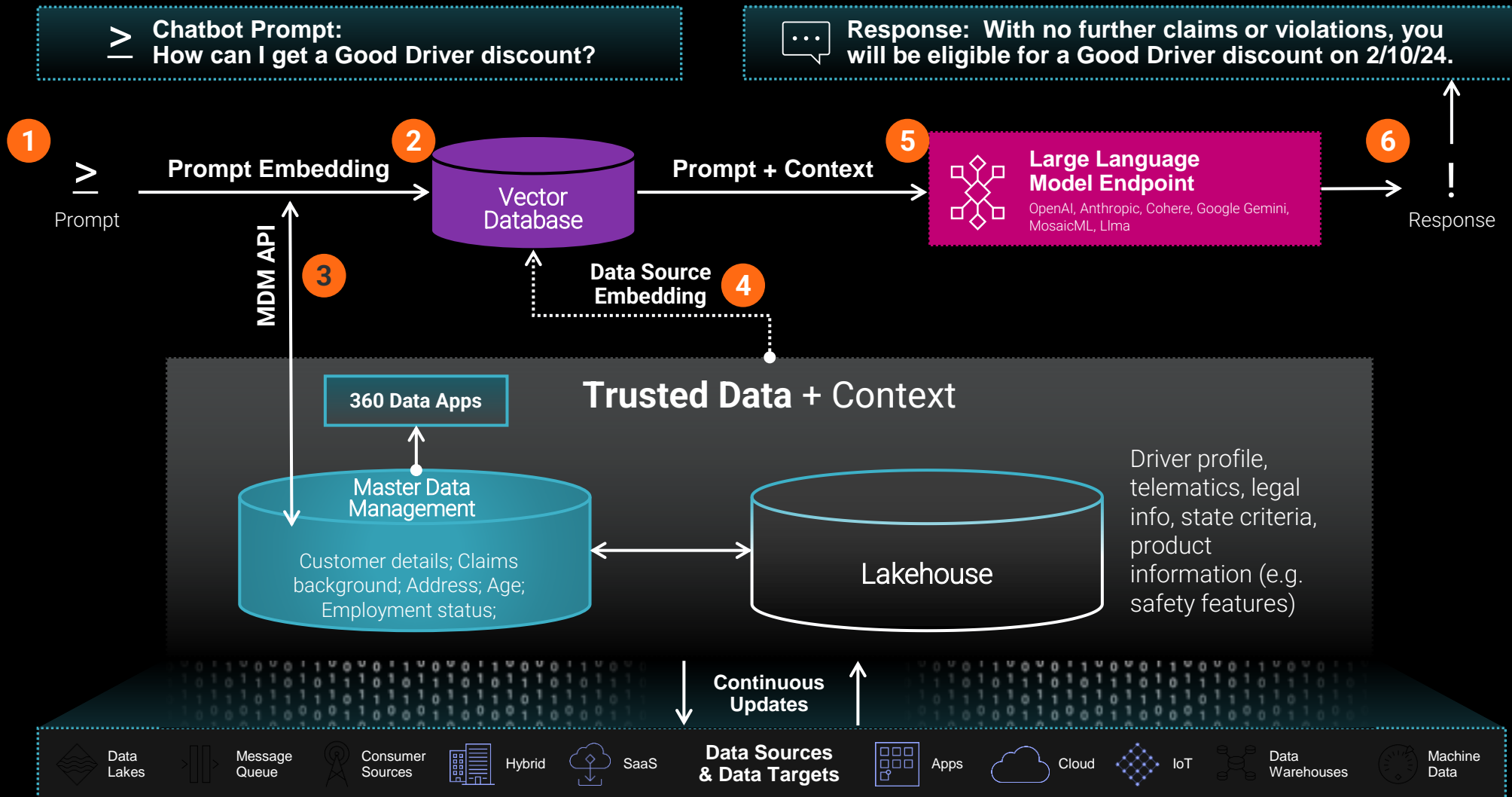


Example: User Query with No Context

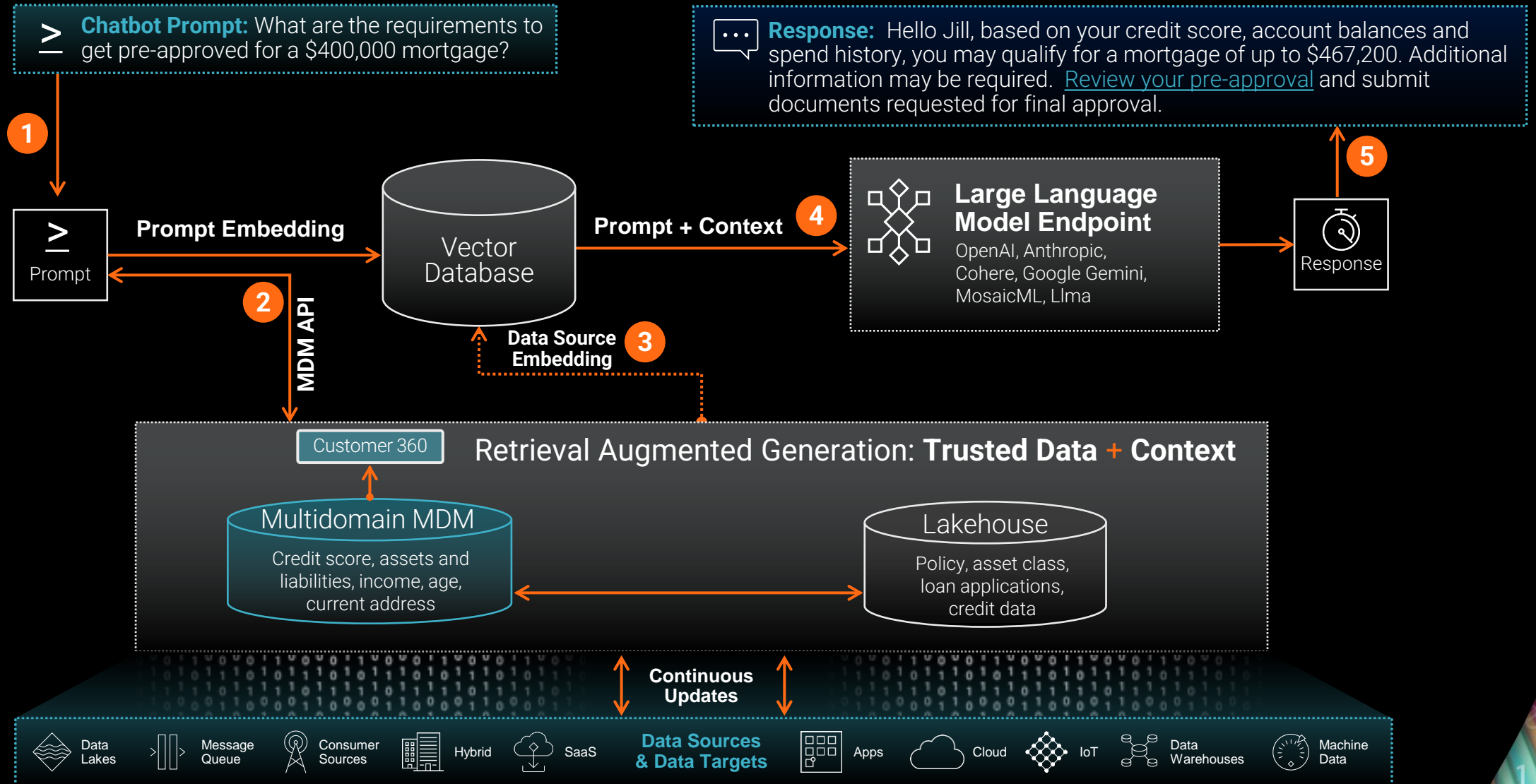


- Generic response
- No context based on user
- Response is helpful, but not tailored or contextual

Modern Data Architecture and RAG for Trusted Data & Context



Grounding AI with Trusted Data



Simplified AI Agent Calling

APIs

- API Source Connectors

REACT

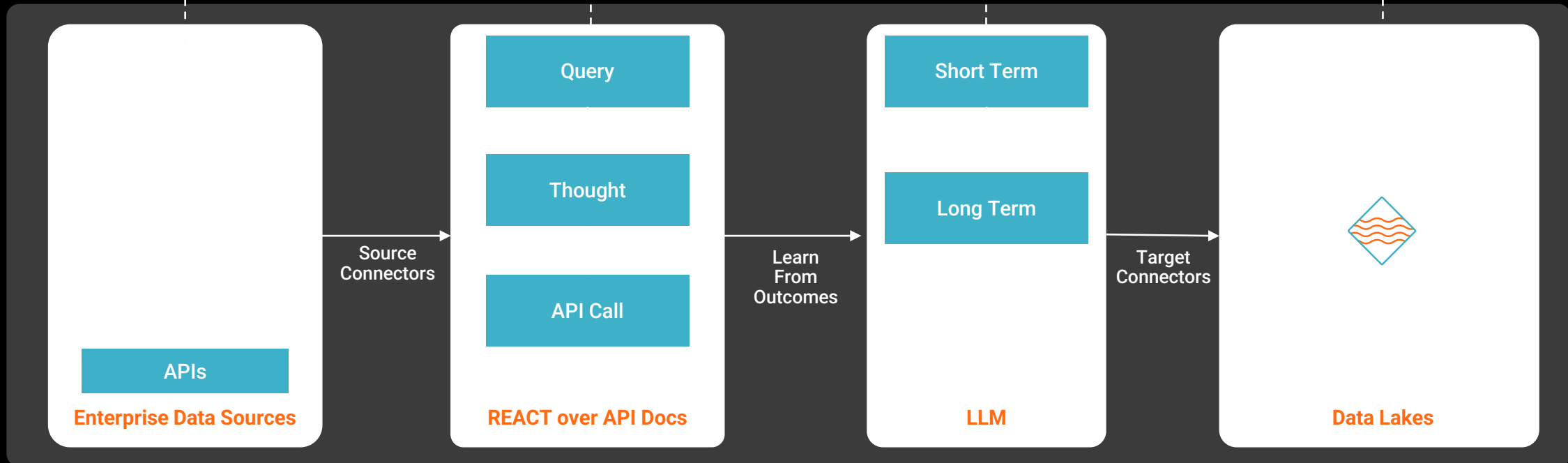
- Learn functional calling and tool use from API Documentation
- Use OODA Loops while inferencing
- New learning approaches also use OODA loops to fine tune with self play

Memory

- Short Term - Run Trials to check outcomes
- Long Term – Identify successful patterns to learn and store information in data lakes for fine tuning

Fine-Tuning Training and Validation Data

- Data Lake Targets: S3, Other File Systems



- AI system queries or "calls" external services or APIs, such as web search engines, knowledge bases, or specialised APIs, to retrieve additional information or perform specific tasks. This information are then incorporated into the AI system's response or decision-making process.
- Global Bank in APJ used Informatica's iPaaS to orchestrate AI agent calling and Google BERT models to build FinServ business processes
- A virtual assistant like Siri or Alexa might use an AI agent calling to retrieve weather information from a weather API, look up movie showtimes from a cinema database, or perform calculations using a math API.

Simplified Fine-Tuning Pipelines

Unstructured Corpus

- Formats: DOCX, PDF, PPT, Email, HTML
- Sources: Sharepoint, File Servers, Data Lakes, Websites, Email Servers

Structured Sources

- APIs, ERP, CRM, MDM, DWH, Lakehouse, Graph Databases

Preprocess

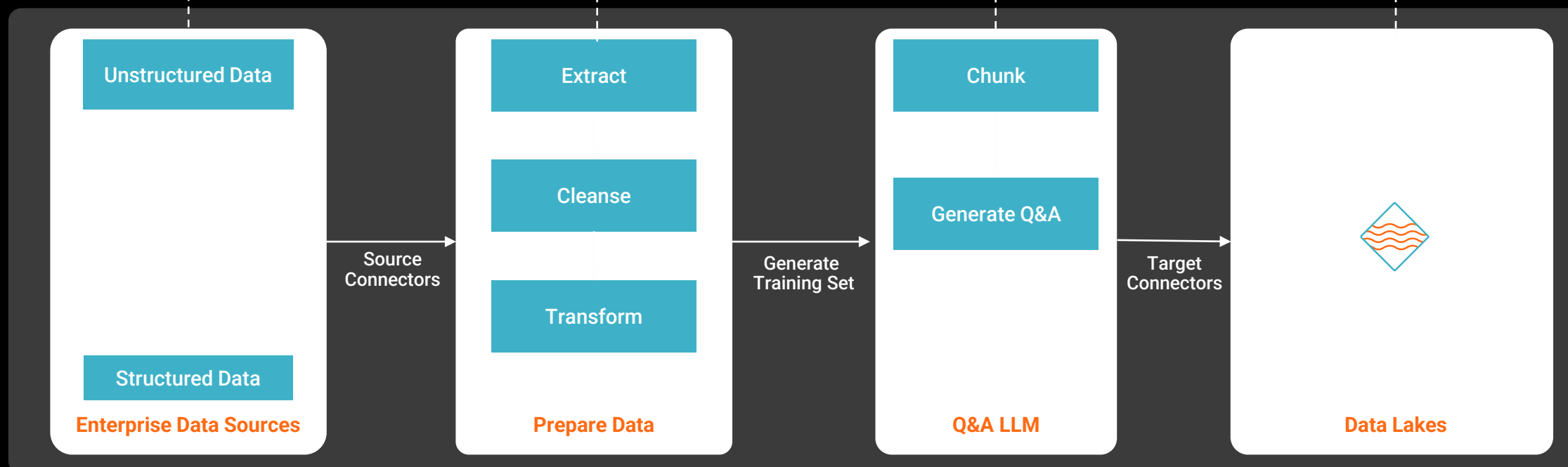
- Extract: Includes metadata, layouts, content
- Cleanse: Remove confidential data, filter spam/ads/irrelevant text, duplicate content
- Transform: Classify, Original Text to JSON

Generate Q&A

- Chunk: Optional step to convert long documents to manageable chunks
- Gen Q&A: Convert text to question and answers using LLMs

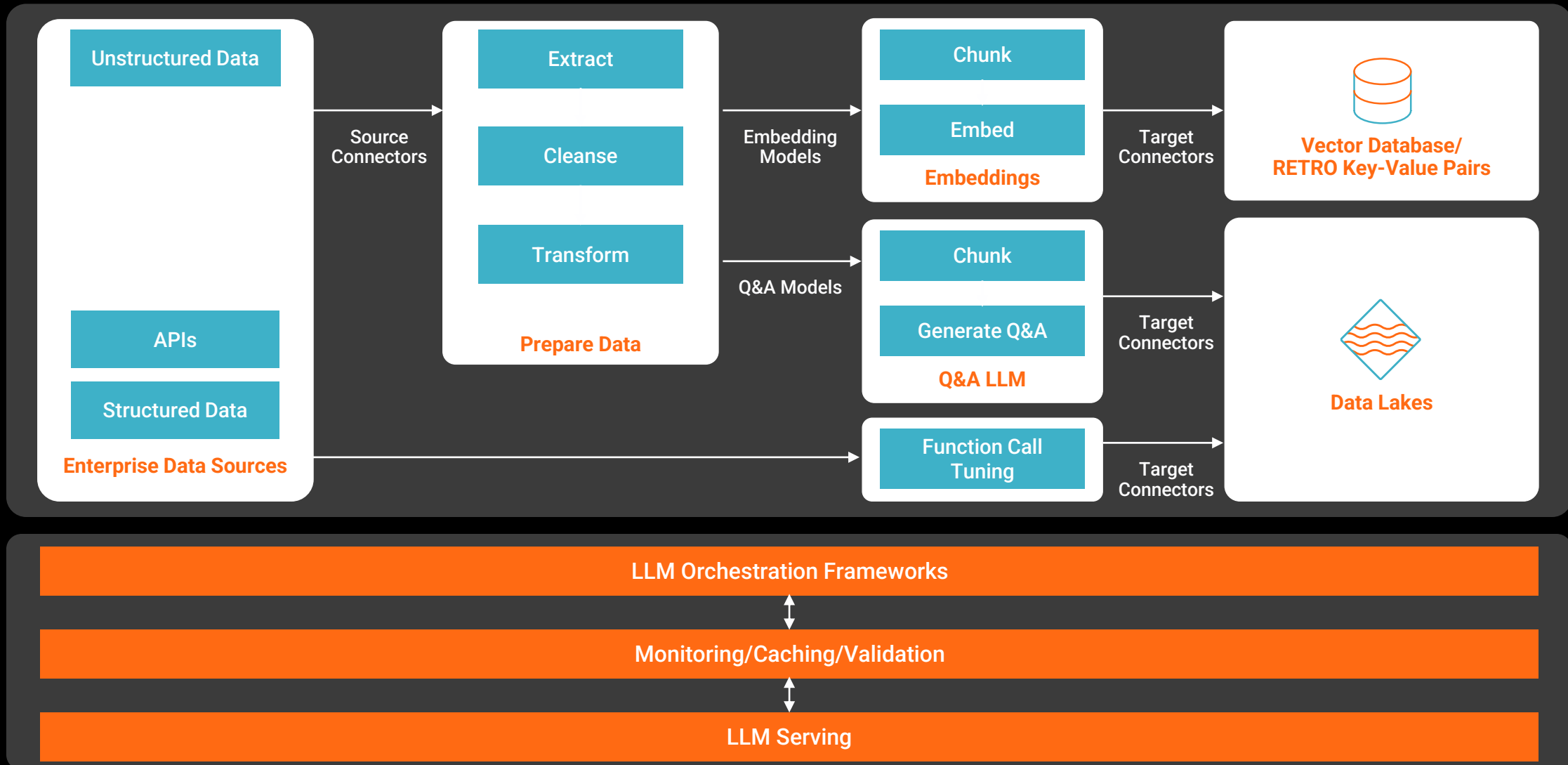
Fine-Tuning Training and Validation Data

- Data Lake Targets: S3, Other File Systems

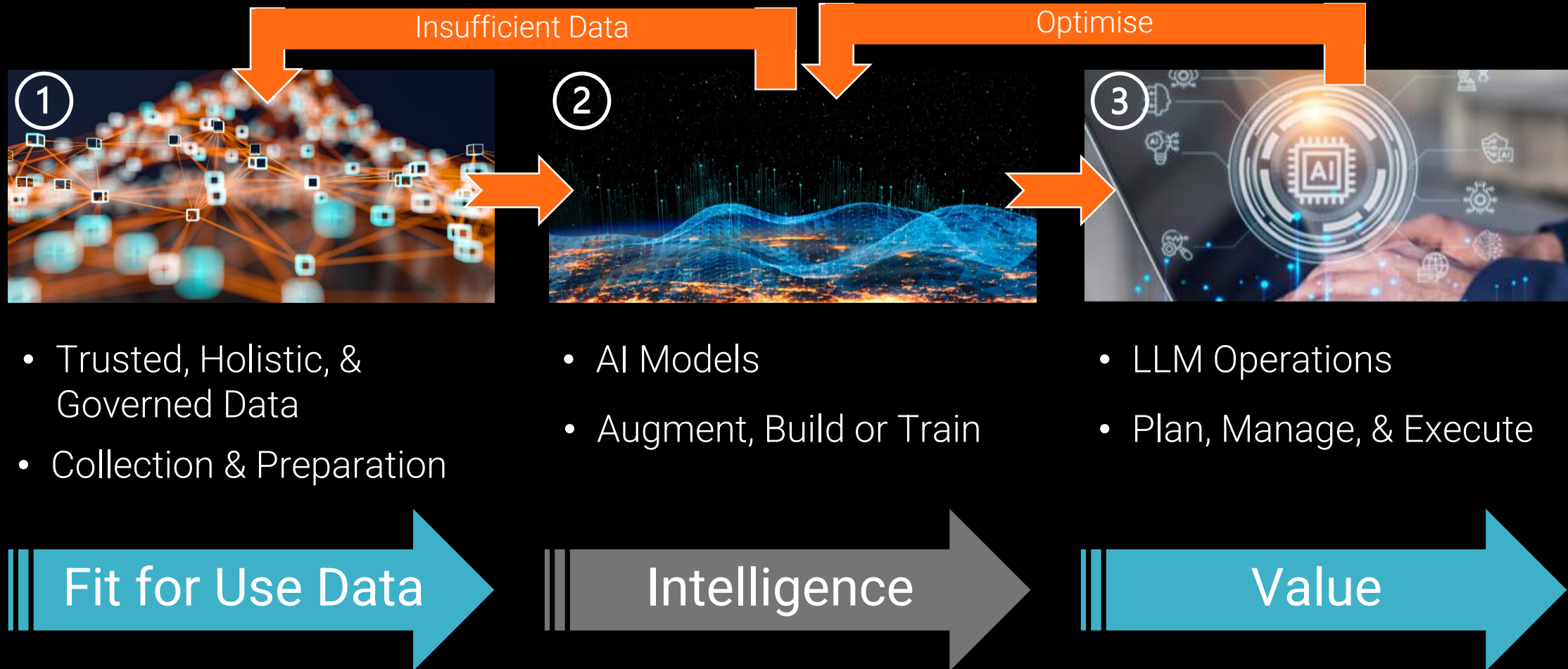


1. Start with a pre-trained model that has already learned a lot of general knowledge from a large dataset during its initial training.
2. Prepare a smaller, task-specific dataset related to the particular problem you want the model to solve (e.g., customer service dialogues if you want a chatbot, or medical images if you want a disease diagnosis model).
3. Load the pre-trained model and continue training it on the task-specific dataset, making adjustments to its knowledge and weights to better fit the new data.

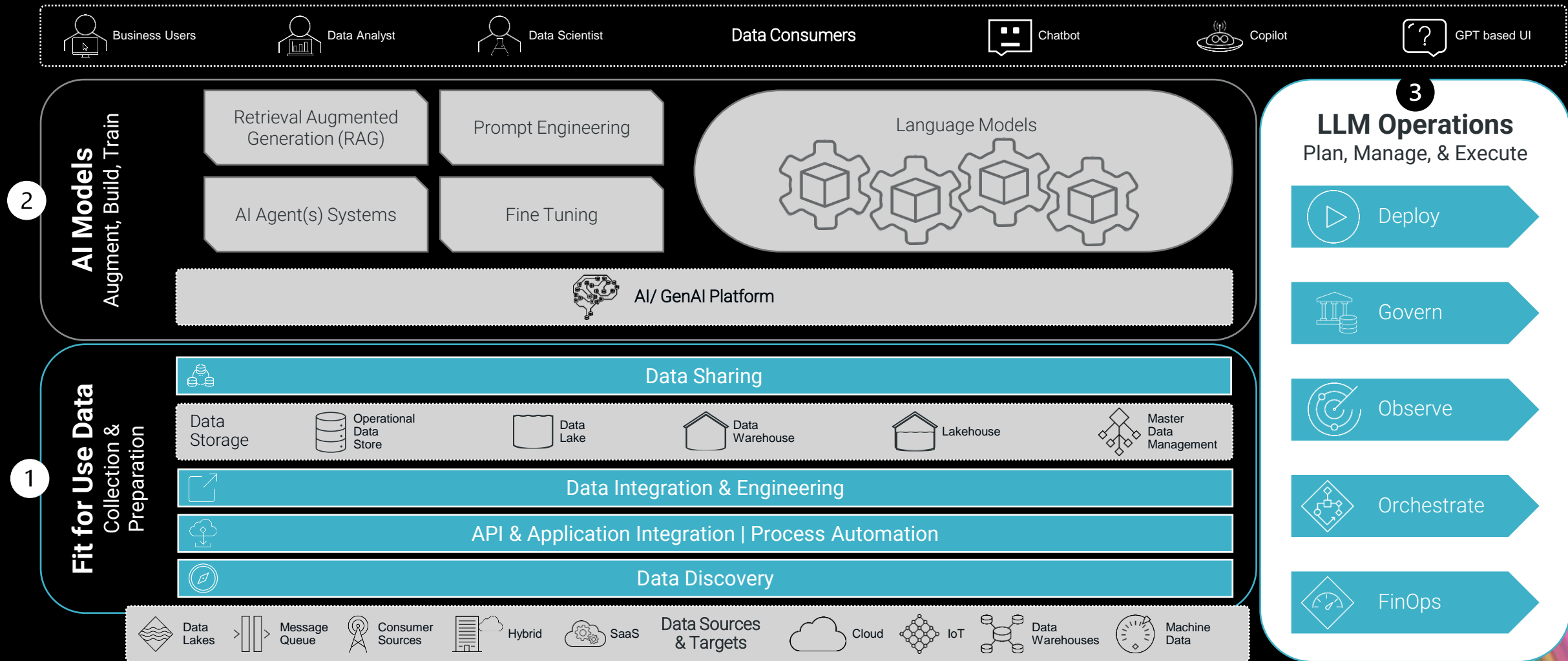
Simplified Enterprise Architecture for Gen AI



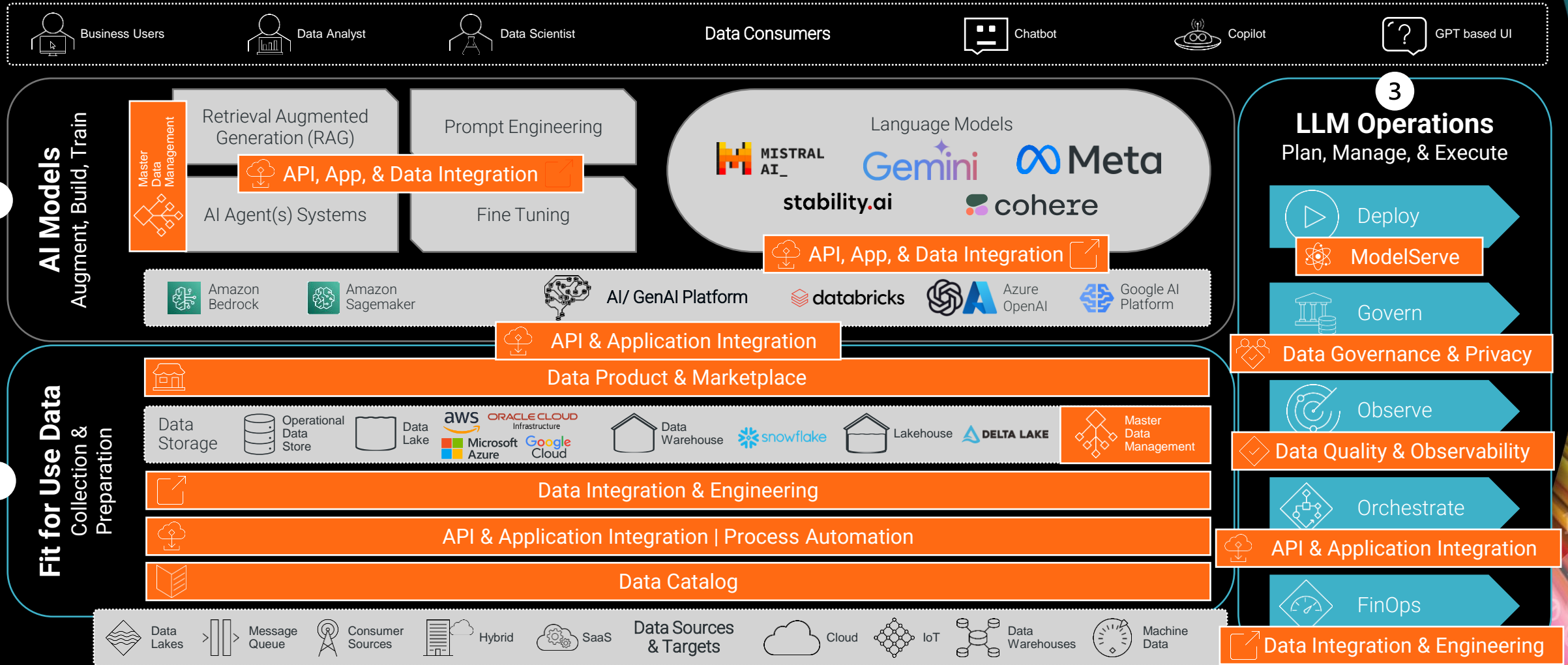
How AI/GenAI Works



AI Reference Architecture



Data Management for AI





GOAL

Our goal is to make the most of our data to drive the journey for our members.

DATA MANAGEMENT
& AI STRATEGY

“With Informatica, we can now promote data-driven journeys across NRMA”

Single data platform strategy to consolidate all data on Google Cloud



Driving productivity with low-code modern data management



Deliver timely access to data, automatically

VALUE

Tenfold boost in productivity with automation; 60% Reduced Data Ingestion Time; 98% Less Time For Campaign Delivery; Self-service functions for democratised data on a modern, single data platform for connected, trusted insights

AI for Data Management & Data Democratisation

Organisations Require “Fit for Business Use” Data

Key Business Imperatives



IMPROVE
CUSTOMER
EXPERIENCE



GROW THE
BUSINESS



COMPLY WITH
REGULATIONS



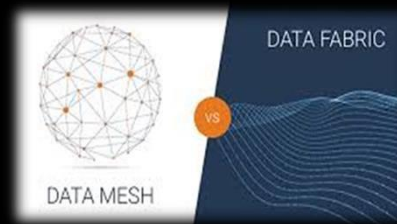
STRENGTHEN
RISK
MANAGEMENT



INCREASE
BUSINESS
AGILITY



Your Investments



Success Requires “Fit for Business Use” Data, and Intelligence & Automation are key enablers

TRANSPARENT

End to end visibility into the lineage of your data from creation to consumption

ACCESSIBLE

Access to data and applications that run your business

CLEAN

Clean and trustworthy data for business use

VALID

Valid and authoritative sources of the truth across systems

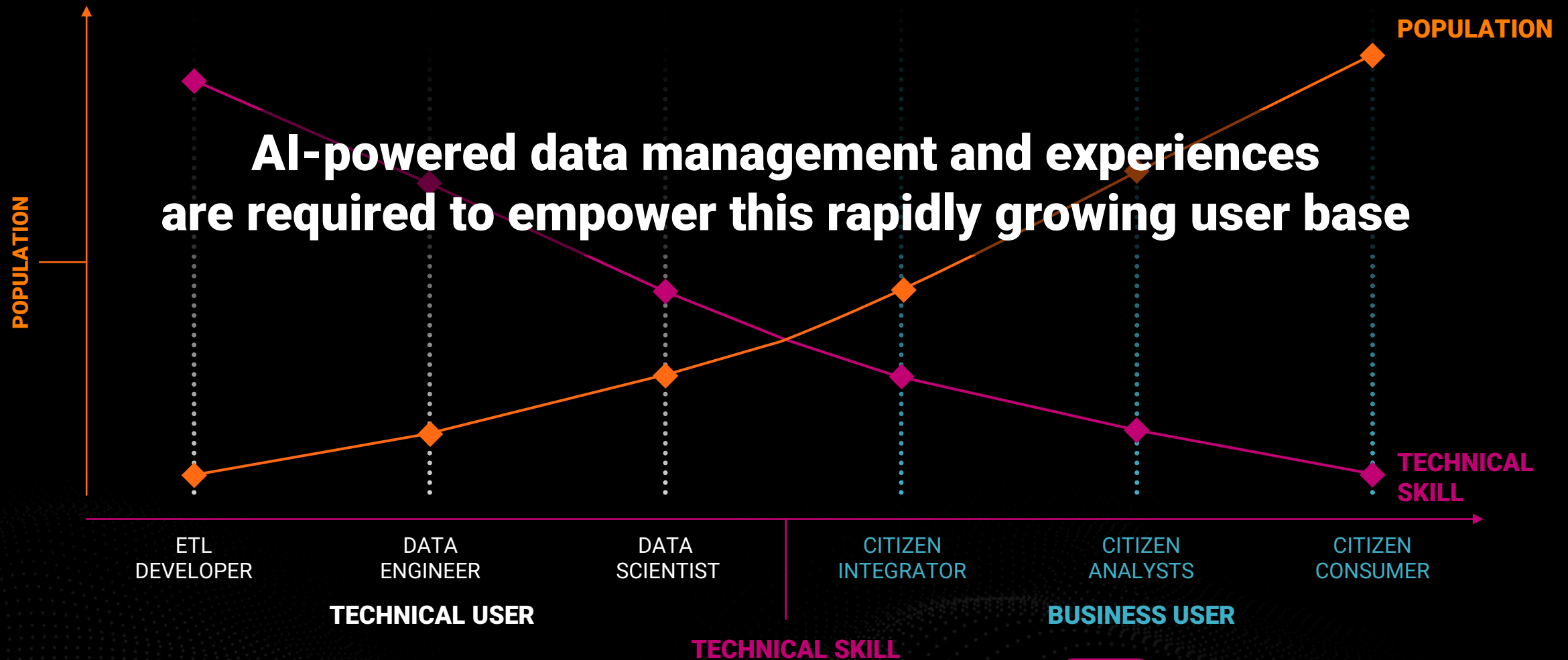
GOVERNED

Defined policies, standards, and established controls by data stewards across your business

UNDERSTOOD

Enabling every business user to get answers about the data they use every day

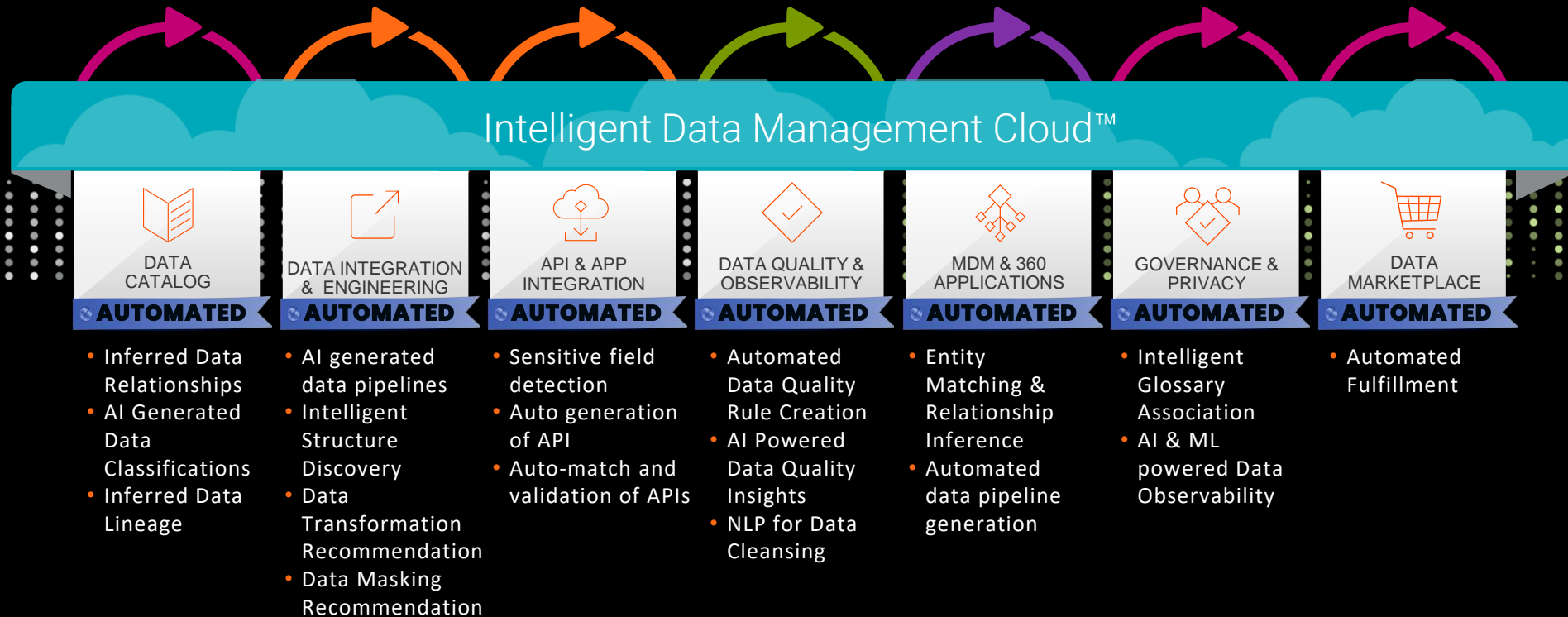
Inability to Address the Expanding Breadth of Data Users



80% of Data Analysts spend around 50% of their time cleaning & organising data¹

80%

CLAIRE[®] copilot - Data Management Intelligence & Automation



Latrobe University

GOAL

Develop a unified, trusted source of certified data to better service their customers, including students, partners and the community.

DATA MANAGEMENT & AI STRATEGY

“Informatica’s Customer 360 enables us to have a single source of truth, a holistic 360-degree view of data we can trust from across our organization for data analytics and insights”

Automated Data Classification to improve data governance and privacy controls



Uses AI to automate onboarding & mastering of students



Use trusted data to provide close to real-time insights

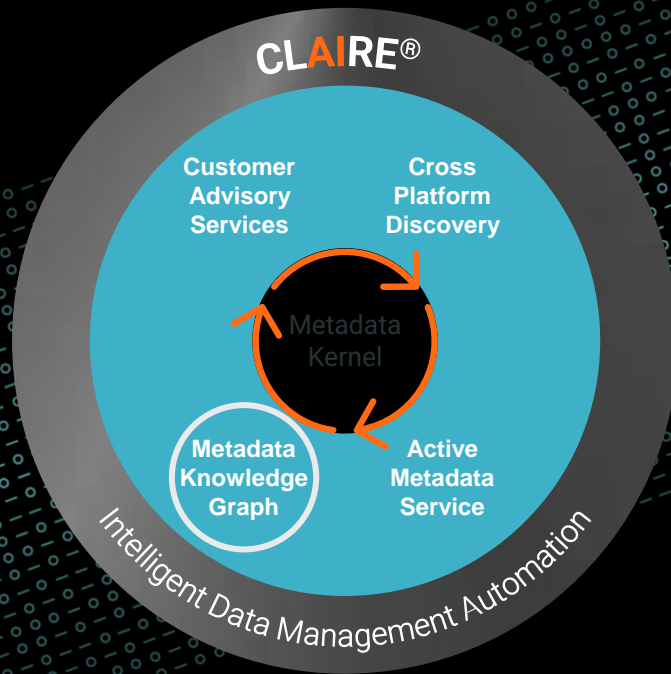
VALUE

Their AI model can predict within two weeks of a student's failure; AI classifies data for privacy regulation; near real-time insights

CLAIRE[®]: AI-Powered Data Management

CLAIRE[®] copilot

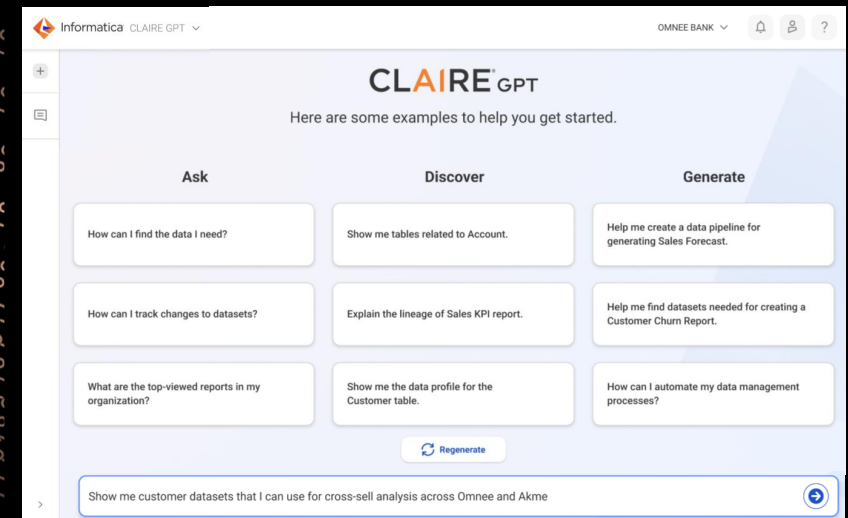
AI-POWERED METADATA INTELLIGENCE AND AUTOMATION



Data Management Automation

CLAIRE[®] GPT

NATURAL LANGUAGE INTERFACE TO DATA



Generative AI-powered Data Management

Gen AI to Simplify Data Management for All Users

CLAIRE® GPT Possibilities



Informatica CLAIRE GPT

OMNEE BANK

CLAIRE[®] GPT

Here are some examples to help you get started.

Ask

Discover

Generate

How can I find the data I need?

Show me tables related to Account.

Help me create a data pipeline for generating Sales Forecast.

How can I track changes to datasets?

Explain the lineage of Sales KPI report.

Help me find datasets needed for creating a Customer Churn Report.

What are the top-viewed reports in my organization?

How can I automate my data management processes?

Analysts

What are the original sources of data for the customer retention rate report?.

Explain the lineage for the customer retention rate data.

Are there any anomalies or outliers in the customer retention rate data that need to be investigated?

Show me customer datasets that I c

Business Users

Who are our most valuable customers in terms of spending over the last year?

What products or services do customers frequently buy together?

What are the trends in customer retention rate over the last 4 quarters?

Data Engineers

Where can I find the datasets I need to calculate Customer Acquisition Cost (CAC)?

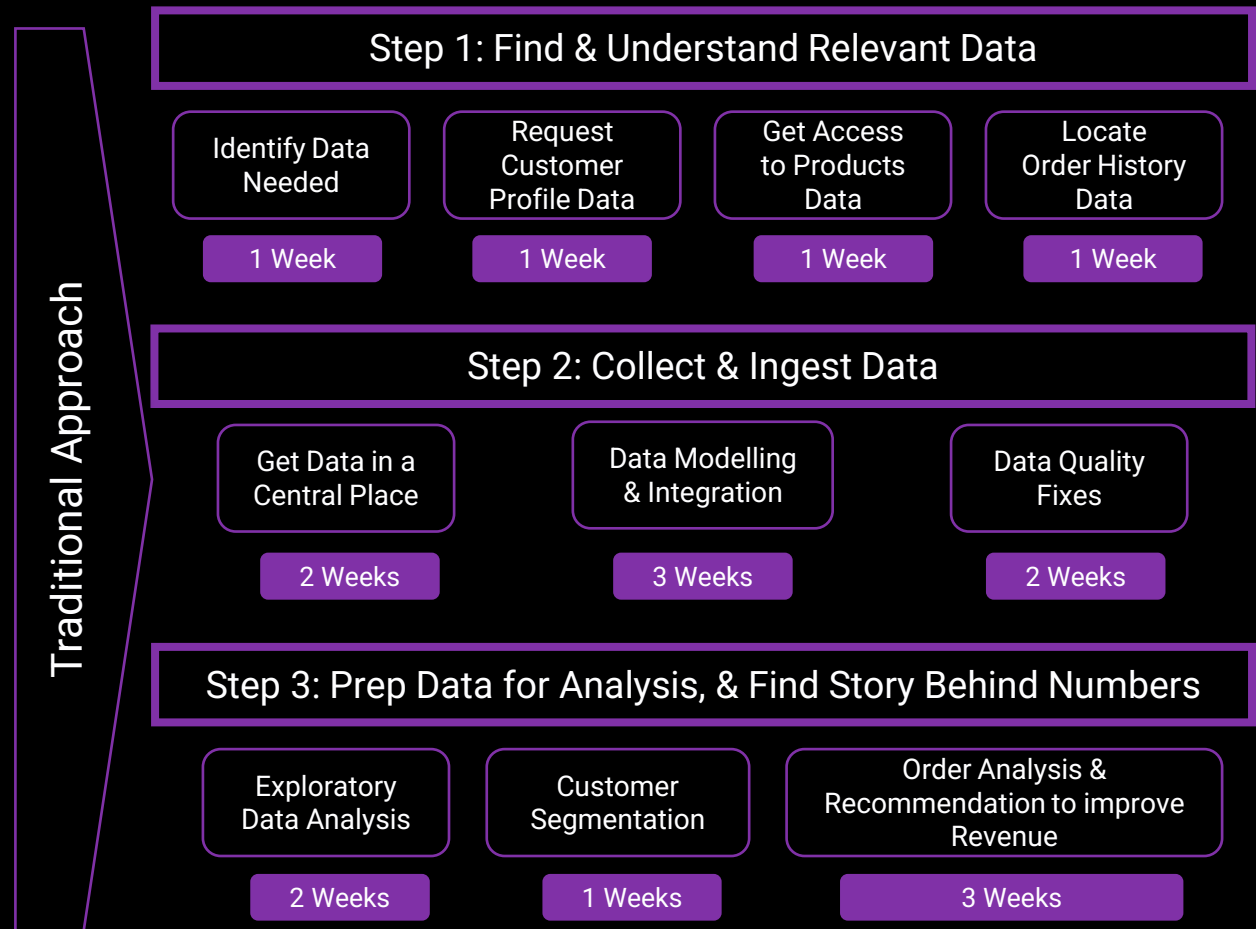
Are all the datasets I need to calculate CAC connected and accessible?.

Create a pipeline that calculates CAC for different marketing channels.

Business Analysis & Revenue

Example Scenario

- **Role:** Business Analyst at XYZ Retail Ltd.
- **Objective:** Understand Customer Behavior, Analyse Past Orders, and Recommend how to Improve Revenue
- **Strategy:** Use data to segment customers, identify top performing products, and underperforming categories



CLAIRE[®]GPT

Need help getting started? Try one of these sample prompts.

Discover

Show me the assets controlled by @claire
GPT_DEMO

Find catalog sources of type Snowflake,
Databricks, and Oracle

Show me the assets which are profiled in
last 30 days

Explore

Show data preview of
@CMI_SPECIALTY_CODES

Show data preview of @CMI_PHYSICAN

Create a mapping for
@CMI_SPECIALTY_CODES

Ask

How do you create a custom workflow in
CDGC?

Can mass ingestion cause any restrictions
when using SQL Server?

What API can be used to get the details of
an asset in CDGC?

What is CLAIRE GPT?



CG

What is Claire GPT?



Finding out what CLAIRE GPT is

CLAIRE GPT is a business-oriented tool powered by Informatica's AI machine learning engine, CLAIRE. It allows users to process, manage, and analyze data using natural language prompts. Users can discover data and explore metadata available in Data Governance and Catalog, including metadata from various sources like databases, data integration sources, data lakes, and data warehouses. Additionally, they can discover and explore business assets such as business glossaries, data sets, policies, and processes. By performing data-driven insights based on responses from CLAIRE GPT, users can determine the data sets that best meet their business requirements and create pipelines to transform their data.

Citations:

1. [CLAIRE GPT/Getting Started With Claire Gpt-Welcome to CLAIRE GPT](#)
2. [CLAIRE GPT/Using Claire Gpt-User persona based use cases](#)
3. [CLAIRE GPT/Administration-Administer CLAIRE GPT](#)



CG

Show me retail customer datasets



Searching for retail customer datasets

A small, circular loading spinner icon.
Generating response...

Ask CLAIRE...



CG

Show me retail customer datasets



Searching for retail customer datasets

Here are the retail customer datasets available across different resources.

1. **RETAIL_CUSTOMER**: This table is part of the **sandbox_CDW_Snowflake** resource in **Snowflake**. It resides under the **CDW_DEMO_DB** database in the **SANDBOX_PUBLIC** schema.
2. **RETAIL_CUSTOMERS**: This table does not have a description provided. It is part of the **cpg_snowflake_retail** resource in **Snowflake**. It resides under the **RETAIL_CPG** database in the **PUBLIC** schema.
3. **retail_customers_202405100903.csv**: This is a flat file present in the **cpg_adlsg2_retail** resource in **Adls Gen2**. It resides under the **cdpadlsgen2** storage account in the **retailcpg** container.

Found **6** of the available **10** records.

Here are the retail customer datasets found in different resources.

1. Datasets under the **sandbox_CDW_Snowflake** Snowflake resource in **CDW_DEMO_DB** database under the **SANDBOX_PUBLIC** schema:
 - **CUSTOMERS**: This is a table dataset.
 - **DM_CUSTOMERS_SALES**: This is another table dataset.
 - **CUSTOMERS_SALES**: This is also a table dataset.
2. Dataset under the **cpg_adlsg2_retail** Adls Gen2 resource in **cdpadlsgen2** storage account under the **retailcpg** container:
 - **retail_products_202405100903.csv**: This is a flat file dataset.
3. Datasets under the **cpg_snowflake_retail** Snowflake resource in **RETAIL_CPG** database under the **PUBLIC** schema:



Show me data preview for CUSTOMERS

Show me data preview for RETAIL_CUSTOMER

Get an overview of RETAIL_CUSTOMER

Ask CLAIRE...





Tell me more about @RETAIL_CUSTOMERS



Looking for more information about retail_customers

RETAIL_CUSTOMERS Overview

Description

RETAIL_CUSTOMERS is a table stored in a Snowflake environment. It has been profiled to assess its structure and quality comprehensively. However, no detailed description is provided.

Data Characteristics

- **Resource Type:** Snowflake
- **Number of Columns:** 1200
- **Data Quality:**
 - **Accuracy:** 95%, indicating a high level of reliability.
 - **Validity:** 92%, showing a good level of data correctness.
 - **Timeliness:** 88%, suggesting that the data is updated regularly.
 - **Uniqueness:** 90%, indicating a high level of distinct entries.
 - **Completeness:** 94%, indicating that most data entries are complete.
- **Profiled:** Yes
 - This indicates that RETAIL_CUSTOMERS has undergone profiling to assess its quality and structure comprehensively.



Get the data quality scores of RETAIL_CUSTO...

Find me an overview of RETAIL_CUSTO...

Find dataset with data lineage to RETAIL_CUSTO...

Ask CLAIRE...



Key Columns

The table includes several key columns that are significant for understanding retail customer information:

- CustomerID: A unique identifier for each customer.
- CustomerName: The name of the customer.
- Address: The address of the customer.
- Phone: The phone number of the customer.
- Email: The email address of the customer.

Glossary

It is associated with key business terms like "CustomerID" and "Email."

Data Entity Classification

It is classified under the "Customer Information" category, highlighting its importance in understanding and managing customer data.

Quality

- The quality has been assessed with metrics provided for accuracy, validity, timeliness, uniqueness, and completeness.

Stakeholdership

- RETAIL_CUSTOMERS has been profiled by the "Data Quality Assessment Team" who are responsible for maintaining its quality and structure.



Find an overview of RETAIL_CUSTO...

Find tables that can be joined with RETAIL_CUSTO...

Get dataset with data lineage to RETAIL_CUSTO...

Ask CLAIRE...



Quality

- The quality has been assessed with metrics provided for accuracy, validity, timeliness, uniqueness, and completeness.

Stakeholdership

- RETAIL_CUSTOMERS has been profiled by the "Data Quality Assessment Team" who are responsible for maintaining its quality and structure.



Show me what glossaries are associated with @RETAIL_CUSTOMERS



Showing the glossaries associated with retail_customers

Here are the glossaries associated with **RETAIL_CUSTOMERS**. They are part of the **Party Data** domain under the **Personal Data** subdomain in the **Data Subject Categories** subdomain.

1. **Customers (inc. historic)**: This term refers to individuals who have previously or currently received goods or services from the organisation.
2. **Customer**: This term refers to corporate entities or individuals that have an economic relationship of some sort with the company, excluding employment.



Find an overview of RETAIL_CUSTO...

Find tables that can be joined with RETAIL_CUSTO...

Get dataset with data lineage to RETAIL_CUSTO...

|





What Kind of Data Quality does @RETAIL_CUSTOMERS

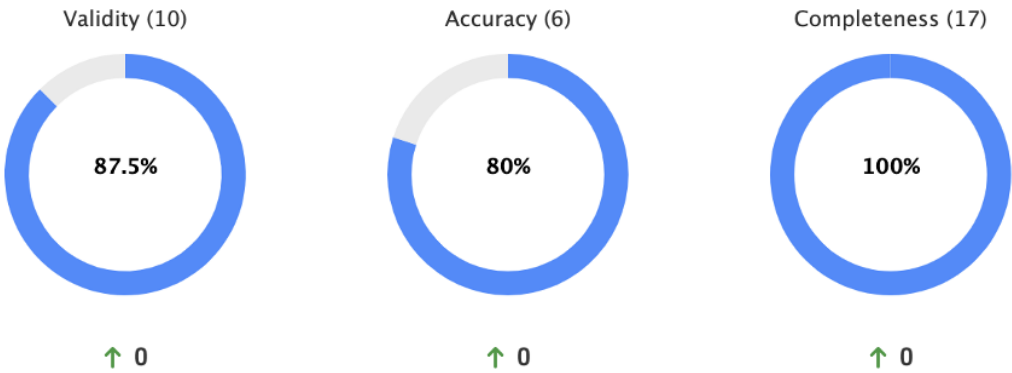


Finding out the kind of data quality does RETAIL_CUSTOMERS

The RETAIL_CUSTOMERS data quality analysis reveals the following:

- Validity (87.5):** The data shows high adherence to predefined formats or rules, indicating that the data meets the established criteria for validity, enhancing its reliability for analysis.
- Accuracy (80.0):** This score suggests that the data accurately reflects the real-world objects or events they are intended to represent, indicating that the data is reliable for representing customer information.
- Completeness (100.0):** The data is completely present and accounted for, with no missing entries, ensuring a comprehensive understanding of the customer data.
- Overall Data Quality (122.5):** This high score encapsulates the overall data quality, reflecting minor deficiencies across the measured dimensions. It underscores the high quality of the RETAIL_CUSTOMERS data, making it trustworthy and valuable for business operations and decision-making.

Data Quality Scores:



CG

Show me the profiling for @RETAIL_CUSTOMERS



Showing the profiling for retail_customers

The data profiling of the "RETAIL_CUSTOMERS" table provides insights into the key columns such as PHONE, ID, GENDER, LAST_NAME, RACE, FIRST_NAME, CITY, COUNTRY, and DOB. These insights are crucial for understanding the quality and robustness of the data, which significantly impacts business strategies and customer information management.

PHONE

- **Data Type:** String
- **Analysis of 980 Entries:**
 - **Distinctiveness:** High at 99.8%
 - **Duplication:** Low duplication at 0.2%
 - **Data Quality:** Very few nulls (0.2%), indicating high data quality.

ID

- **Data Type:** Integer
- **Analysis of 980 Entries:**
 - **Distinctiveness:** High at 99.8%
 - **Duplication:** Low duplication at 0.2%
 - **Data Quality:** Very few nulls (0.2%), indicating high data quality.



Find tables that can be joined with RETAIL_CUS...

Find me the data quality scores of RETAIL_CUS...

Show me data preview for RETAIL_CUST...

Ask CLAIRE...



GENDER

- **Data Type:** String
- **Analysis of 980 Entries:**
 - **Distinctiveness:** Very low at 0.2%
 - **Duplication:** Very high duplication at 99.8%
 - **Data Quality:** Very few nulls (0.2%), indicating high data quality.

LAST_NAME

- **Data Type:** String
- **Analysis of 980 Entries:**
 - **Distinctiveness:** High at 98.8%
 - **Duplication:** Low duplication at 1.2%
 - **Data Quality:** Very few nulls (0.2%), indicating high data quality.

RACE

- **Data Type:** String
- **Analysis of 980 Entries:**
 - **Distinctiveness:** Low at 9.5%
 - **Duplication:** High duplication at 90.5%
 - **Data Quality:** Very few nulls (0.2%), indicating high data quality.

FIRST_NAME



Find tables that can be joined with RETAIL_CUS...

Find me the data quality scores of RETAIL_CUS...

Show me data preview for RETAIL_CUST...

Ask CLAIRE...





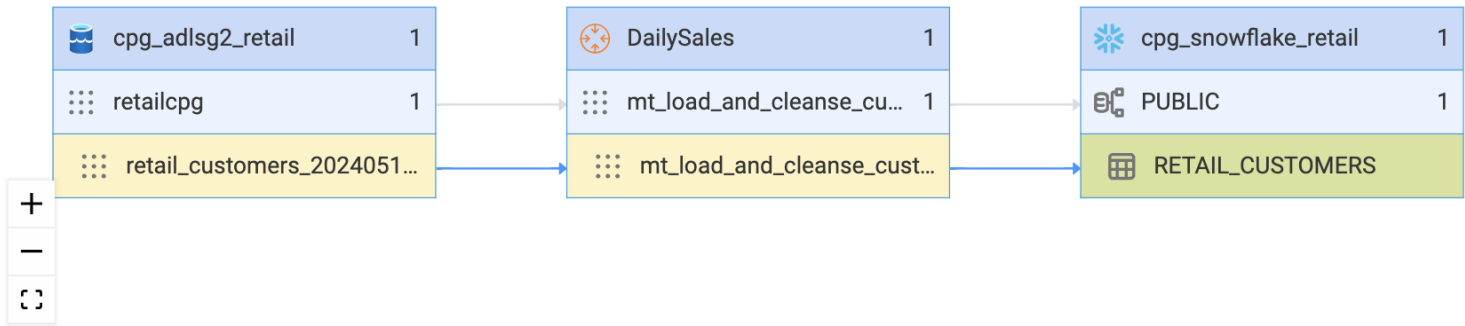
Show the lineage for @RETAIL_CUSTOMERS



Showing the lineage for retail_customers

- The lineage of **RETAIL_CUSTOMERS** involves a total of **2 catalog sources**.
- The data flow begins from the source **cpg_adlsg2_retail**, which includes the asset **retail_customers_202405100903.csv**.
- This data then flows into **RETAIL_CUSTOMERS** and proceeds towards the target resource **cpg_snowflake_retail**, which contains the asset **RETAIL_CUSTOMERS**.

Lineage of RETAIL_CUSTOMERS Table



Download



Show glossaries associated to RETAIL_CU...

Find me dataset with data lineage to RETAIL_C...

Show tables that can be joined with RETAIL_C...

Ask CLAIRE...



CG

show me the impact analysis of adding a column to @RETAIL_CUSTOMERS



Showing the impact analysis of adding a column to retail_customers

Impact Analysis of Changing RETAIL_CUSTOMERS:**• Upstream Impact:**

- Changes to RETAIL_CUSTOMERS could necessitate modifications in how data is ingested from the source catalog: **cpg_adlsg2_retail**.
- The single asset, **retail_customers_202405100903.csv**, might require adjustments in its format, structure, or data type to ensure compatibility with the new RETAIL_CUSTOMERS structure.

• Downstream Impact:

- Any changes in RETAIL_CUSTOMERS would directly influence the data being supplied to the downstream target: **cpg_snowflake_retail**.
- The single target asset, **RETAIL_CUSTOMERS**, might need updates to accommodate changes in the data's format, granularity, or semantics. This could impact the accuracy, timeliness, and relevance of insights derived from this data.

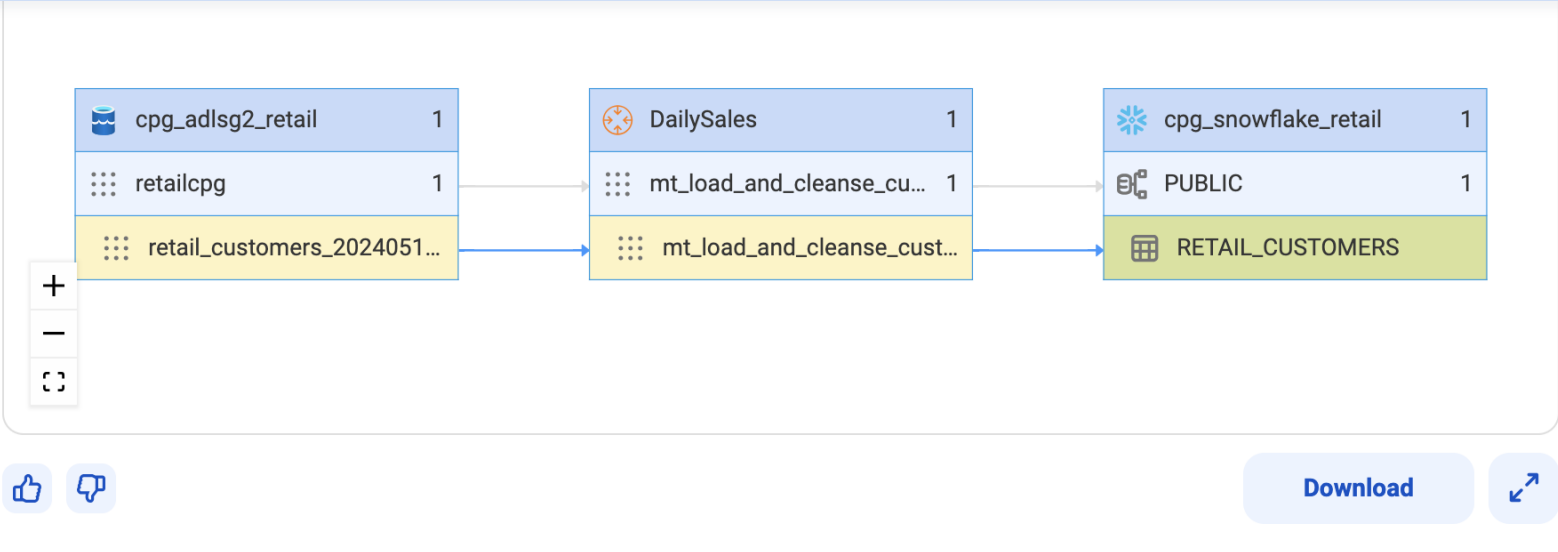
• Overall Ecosystem Impact:

- The coherence and efficiency of the entire data flow, from the initial data source through RETAIL_CUSTOMERS to the eventual data target, might be impacted. This necessitates a thorough review and potentially a redesign of the data processing, transformation, and analysis pipelines to align with the new RETAIL_CUSTOMERS configuration.
- Stakeholders relying on the insights generated from these data flows, such as retail analysts and customer relationship managers, may experience changes in the decision-making process due to the adjusted data inputs and outputs.

[Lineage of RETAIL_CUSTOMERS Table](#) 

Ask CLAIRE...





CG

Show me tables that can be joined with @RETAIL_CUSTOMERS



Searching for tables that can be joined with retail_customers

Here are the tables that can be joined with **RETAIL_CUSTOMERS**. They are part of the **cpg_snowflake_retail** resource in **Snowflake**, situated under the **RETAIL_CPG** database in the **PUBLIC** schema.

1. **RETAIL_ORDERS**: This table is related to the **RETAIL_CUSTOMERS** table.



Find me dataset with data lineage to RETAIL_ORDERS

Show an overview of RETAIL_ORDERS

Show me data preview for RETAIL_ORDERS

Ask CLAIRE...





Show me information about @RETAIL_ORDERS



Showing information about retail_orders

The RETAIL_ORDERS is a table stored in the cpg_snowflake_retail resource. Here's a summary of its key characteristics:

Name: RETAIL_ORDERS **Resource Name:** cpg_snowflake_retail **Type:** Table **Profiled:** Yes

Data Profiling

- **Profiled:** Yes
- **Last Profiled State:** Complete
- **Class Type:** com.infa.odin.models.relational.Table
- **Name:** RETAIL_ORDERS

Data Quality

- **Data Quality:** Not Available

Stakeholdership

- **Created By:** Scott Ligocki
- **Modified By:** Errick Coughlin
- **Certified:** No
- **Asset Lifecycle:** Published



Get tables that can be joined with RETAIL_ORDE...

Show me data preview for RETAIL_ORDE...

Get dataset with data lineage to RETAIL_ORDE...

Ask CLAIRE...



CG

Show me information about @RETAIL_ORDERS



Showing information about retail_orders

The RETAIL_ORDERS is a table stored in the cpq_snowflake_retail resource. Here's a summary of its key characteristics:

Name: RETAIL_ORDERS **Resource Name:** cpq_snowflake_retail **Type:** Table **Profiled:** Yes

Data Profiling

- **Profiled:** Yes
- **Last Profiled State:** Complete
- **Class Type:** com.infa.odin.models.relational.Table
- **Name:** RETAIL_ORDERS

Data Quality

- **Data Quality:** Not Available

Stakeholdership

- **Created By:** Scott Ligocki
- **Modified By:** Errick Coughlin
- **Certified:** No
- **Asset Lifecycle:** Published

[Get tables that can be joined with RETAIL_ORDE...](#)[Show me data preview for RETAIL_ORDE...](#)[Get dataset with data lineage to RETAIL_ORDE...](#)

Ask CLAIRE...





Show me a data preview of @RETAIL_CUSTOMERS



Showing a data preview of retail_customers

TABLE | RETAIL_CUSTOMERS

cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 16

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



A AREA	A City	A Country	Date_Of_Birth	A EMAIL	A First_Name
Main	Catania	Italy	21 Dec 1985	bgretham5y@g...	Maite
Main	Venlo	Netherlands	08 Jul 1949	gkarle52@csm...	Adelie
North	Paris	France	25 Feb 1948	cburganemn@b...	Maelyss
Main	Best	Netherlands	18 Oct 1954	lcowlesk4@ute...	Frederique
Main	Padova	Italy	09 Jan 1965	mlampskb@joo...	Clea
Main	Den Haag	Netherlands	21 Jul 1970	rthurlbornnx@fl...	Publicite
Main	Mestre	Italy	22 Jan 1985	goneilna@amaz...	Marta
North	Paris	France	24 Jan 2000	bmartyntsevlz...	Yu
North	Paris	France	23 Sept 1954	ccarletonei@ac...	Ma
North	Paris	France	04 Feb 2001	rskinner5d@im...	Personnalisee

Viewing sample data: 100 rows and 4 data types

Combine RETAIL_CUSTOMERS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...





Combine @RETAIL_CUSTOMERS with RETAIL_ORDERS



Combining retail_customers with retail_orders

TABLE | RETAIL_CUSTOMERS

cpq_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 25

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

A AREA	A City	A Country	Date_Of_Birth	A EMAIL	A First_Name
Main	Barcelona	Spain	22 Aug 1995	hkarsh0@fema....	Yao
Main	Barcelona	Spain	22 Aug 1995	hkarsh0@fema....	Yao
Main	Barcelona	Spain	22 Aug 1995	hkarsh0@fema....	Yao
Main	Washington	Sweden	10 Mar 1955	cyaakov2@reve...	Ye
Main	Washington	Sweden	10 Mar 1955	cyaakov2@reve...	Ye
Main	Springfield	Sweden	30 Mar 1950	eswaine3@peo...	Penelope
North	Sannois	France	28 Jun 1952	lmatiashvili5@d...	Yenora
Main	New York City	Sweden	31 Mar 1965	mhabens6@spr...	Marylene
Main	Sparks	Sweden	21 Dec 1997	gboddis9@tech...	Aurelie
Main	Sparks	Sweden	21 Dec 1997	gboddis9@tech...	Aurelie

Viewing sample data: 100 rows and 5 data types

- Combine RETAIL_CUSTOMERS with RETAIL_ORDERS
- Create a mapping for the above

Ask CLAIRE...





How many customers we have per country?



Finding out how many customers we have per country in RETAIL_CUSTOMERS

TABLE | **RETAIL_CUSTOMERS**

cpq_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 2

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



A Country	# COUNT(*)
Italy	21
Netherlands	51
France	457
Germany	44
Spain	42
United Kingdom	22
Sweden	332
Switzerland	8
Singapore	1
NULL	2

Viewing sample data: 10 rows and 2 data types

Combine RETAIL_CUSTOMERS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...



Italy	21
Netherlands	51
France	457
Germany	44
Spain	42
United Kingdom	22
Sweden	332
Switzerland	8
Singapore	1
NULL	2

Viewing sample data: 10 rows and 2 data types



Show Code

Download



How many customers live in the city Paris?



Finding out how many customers live in the city Paris in RETAIL_CUSTOMERS

33



Show Code

Combine RETAIL_CUSTOMERS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...





Show Code



Which top 5 countries have the most number of customers



Figuring out the top 5 countries with the most number of customers in RETAIL_CUSTOMERS

TABLE | RETAIL_CUSTOMERS

cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 2

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

×

A Country	# Number_of_Customers
France	457
Sweden	332
Netherlands	51
Germany	44
Spain	42

Viewing sample data: 5 rows and 2 data types



Show Code

Download



Combine RETAIL_CUSTOMERS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...





Figuring out which countries have the most orders in RETAIL_CUSTOMERS

TABLE | RETAIL_CUSTOMERS

cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 2

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

×

A Country	# total_orders
France	457
Sweden	332
Netherlands	51
Germany	44
Spain	42
United Kingdom	22
Italy	21
Switzerland	8
NULL	2
Singapore	1

Viewing sample data: 10 rows and 2 data types



Show Code

Download



Combine RETAIL_CUSTOMERS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...



[Show Code](#)[Download](#)

Which top 5 cities has the most orders?



Figuring out the top 5 cities that have the most orders in RETAIL_CUSTOMERS



TABLE | [RETAIL_CUSTOMERS](#)



cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 2



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



A City	# total_orders
Paris	33
Washington	11
Marseille	10
Nantes	10
Chicago	7

Viewing sample data: 5 rows and 2 data types

[Show Code](#)[Download](#)

Combine RETAIL_CUSTOMERS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...



[Show Code](#)[Download](#)

show me the top 5 product titles that have the highest total revenue



Finding the top 5 product titles that have the highest total revenue in retail_products



TABLE | RETAIL_PRODUCTS

cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 2

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



A TITLE	# total_revenue
Ergonomic Cotton Bag	6421.891431
Durable Cotton Bench	6039.671604
Sleek Steel Table	5808.653650
Incredible Silk Shoes	5803.588293
Mediocre Silk Bottle	5352.653974

Viewing sample data: 5 rows and 2 data types

[Show Code](#)[Download](#)

Combine RETAIL_PRODUCTS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...





show customers that live in the country France or Italy with the number of orders, average order amount, and number of products, and product title purchased with the revenue generated in for each product



Showing customers that live in the countryfrance or italy with the number of orders, average order amount, and number of products, and product title purchased with the revenue generated in for each product in retail_products



TABLE | RETAIL_PRODUCTS

cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 5

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



A TITLE	# Number_of_orde...	# Average_order_...	# Number_of_prod...	# Revenue
Incredible Silk Shoes	26	122.910031692308	26	3195.660824
Intelligent Granite ...	18	139.376185277778	18	2508.771335
Small Cotton Chair	29	63.459520793103	29	1840.326103
Synergistic Marble ...	25	104.860768200000	25	2621.519205
Ergonomic Cotton ...	25	131.736799160000	25	3293.419979
Lightweight Linen ...	24	40.885450416667	24	981.250810
Lightweight Paper ...	18	74.820339500000	18	1346.766111
Ergonomic Silk Table	18	66.658319888889	18	1199.849758
Synergistic Rubber...	15	47.899319933333	15	718.489799
Incredible Aluminu...	12	137.206993750000	12	1646.483925



Combine RETAIL_PRODUCTS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...





Showing customers that live in the countryfrance or italy with the number of orders, average order amount, and number of products, and product title purchased with the revenue generated in for each product in retail_products

TABLE | RETAIL_PRODUCTS

cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 5

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

×

A TITLE	# Number_of_orde...	# Average_order_...	# Number_of_prod...	# Revenue
Enormous Steel Wa...	18	107.105990944444	18	1927.907837
Mediocre Cotton T...	17	43.729770705882	17	743.406102
Small Granite Gloves	22	112.191645454545	22	2468.216200
Aerodynamic Grani...	25	42.465341640000	25	1061.633541
Fantastic Steel Knife	16	54.840186687500	16	877.442987
Aerodynamic Linen...	14	144.119922785714	14	2017.678919
Heavy-Duty Steel ...	11	54.255236818182	11	596.807605
Incredible Concret...	14	116.435405500000	14	1630.095677
Intelligent Paper Car	13	107.110623384615	13	1392.438104
Ergonomic Silk Key...	17	105.208409647059	17	1788.542964
Enormous Marble ...	19	55.667306947368	19	1057.678832

Viewing sample data: 100 rows and 3 data types



Show Code

Download

Combine RETAIL_PRODUCTS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...



SQL

```
SELECT
  R.TITLE,
  COUNT(RO.ID) AS Number_of_orders,
  AVG(RO.TOTAL) AS Average_order_amount,
  COUNT(DISTINCT RO.PRODUCT_ID) AS Number_of_products,
  SUM(RO.TOTAL) AS Revenue
FROM
  RETAIL_CPG.PUBLIC.RETAIL_PRODUCTS R
  JOIN RETAIL_CPG.PUBLIC.RETAIL_ORDERS RO ON R.ID = RO.PRODUCT_ID
  JOIN RETAIL_CPG.PUBLIC.RETAIL_CUSTOMERS C ON RO.CUST_ID = C.ID
WHERE
  (
    LOWER(C.COUNTRY) LIKE '%france%'
    OR LOWER(C.COUNTRY) LIKE '%italy%'
  )
GROUP BY
  R.TITLE
LIMIT
  100
```



Hide Code

Download



Combine RETAIL_PRODUCTS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...





Showing customers that live in the countryfrance or italy with the number of orders, average order amount, and number of products, and product title purchased with the revenue generated in for each product in retail_products

TABLE | RETAIL_PRODUCTS

cpg_snowflake_retail / RETAIL_CPG / PUBLIC Columns: 5

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

A TITLE	# Number_of_orde...	# Average_order_...	# Number_of_prod...	# Revenue
Enormous Steel Wa...	18	107.105990944444	18	1927.907837
Mediocre Cotton T...	17	43.729770705882	17	743.406102
Small Granite Gloves	22	112.191645454545	22	2468.216200
Aerodynamic Grani...	25	42.465341640000	25	1061.633541
Fantastic Steel Knife	16	54.840186687500	16	877.442987
Aerodynamic Linen...	14	144.119922785714	14	2017.678919
Heavy-Duty Steel ...	11	54.255236818182	11	596.807605
Incredible Concret...	14	116.435405500000	14	1630.095677
Intelligent Paper Car	13	107.110623384615	13	1392.438104
Ergonomic Silk Key...	17	105.208409647059	17	1788.542964
Enormous Marble ...	19	55.667306947368	19	1057.678832

Viewing sample data: 100 rows and 3 data types



Show Code

Download



Combine RETAIL_PRODUCTS with RETAIL_ORDERS

Create a mapping for the above

Ask CLAIRE...



Viewing sample data: 100 rows and 3 data types



Show Code

Download



CG

Create a mapping for the above

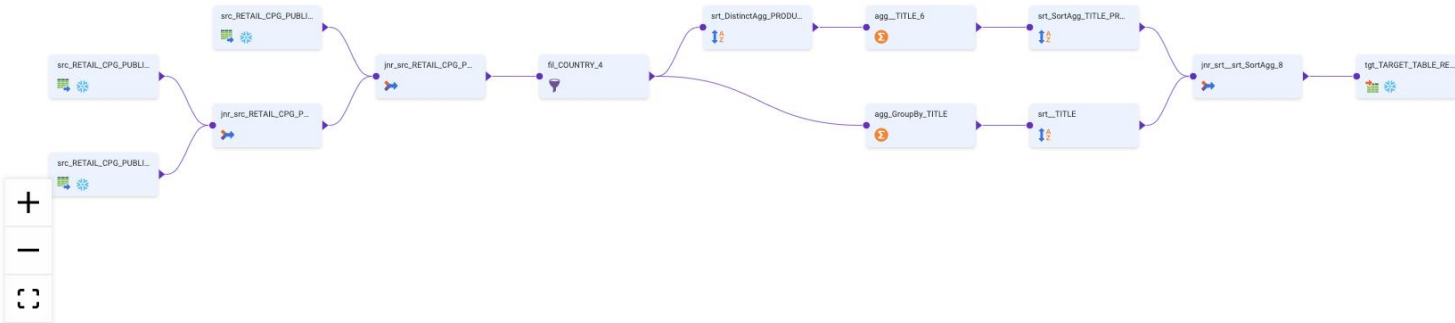


Creating a mapping

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



m_Snowflake_src_RETAIL_CPG_PUBLIC_RETAIL_CUSTOMERS_1_1_tgt_TARGET_TABLE_RETAIL_PRODUCTS_RE...



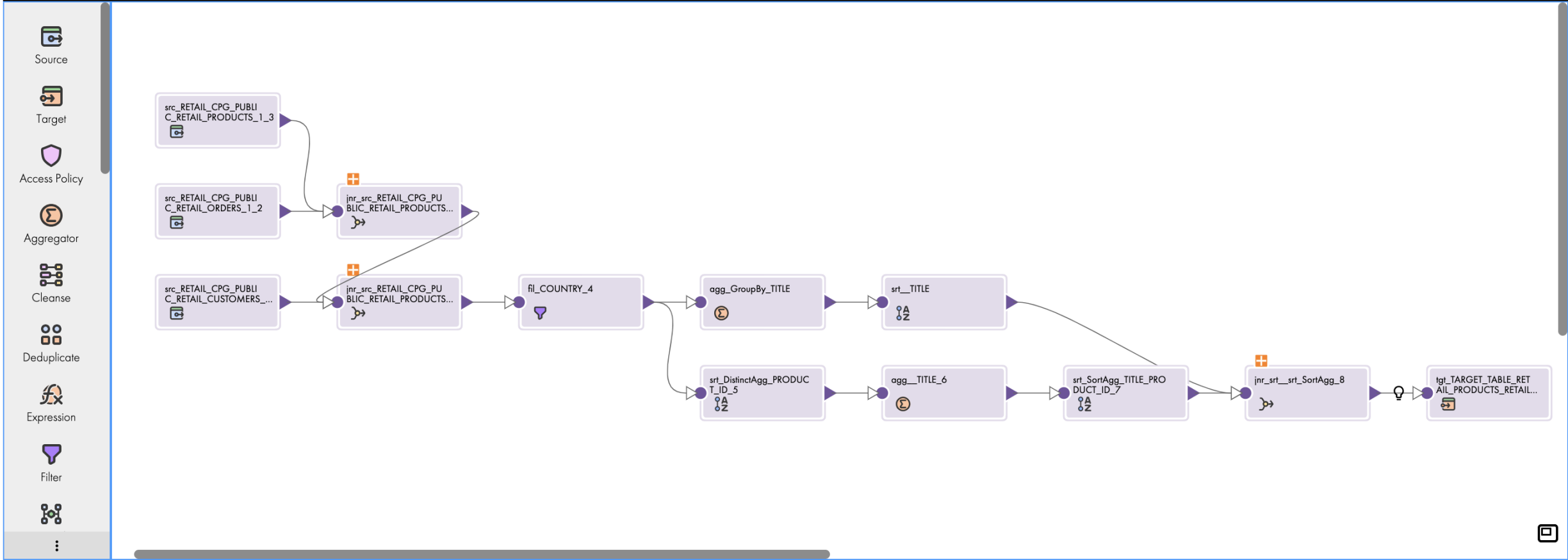
Open in Data Integration



Ask CLAIRE...



Design



Mapping3



Name:*

Mapping3

Location:*

Default

Browse

Description:

Deloitte Consulting

Innovating Customers with Generative AI

"With CLAI^{RE} 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."

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

Deloitte.

MAKING AN IMPACT THAT MATTERS

The Impact of GenAI for Data Management

Allowing the World to Speak the Language of Data



Data for All

Allow everyone in the organisation to be data driven, regardless of technical ability



Create Usable Data

Enable decentralised data producers to create, document, catalog, and test new data assets



Drive Exponential Productivity Growth for Your Data Team

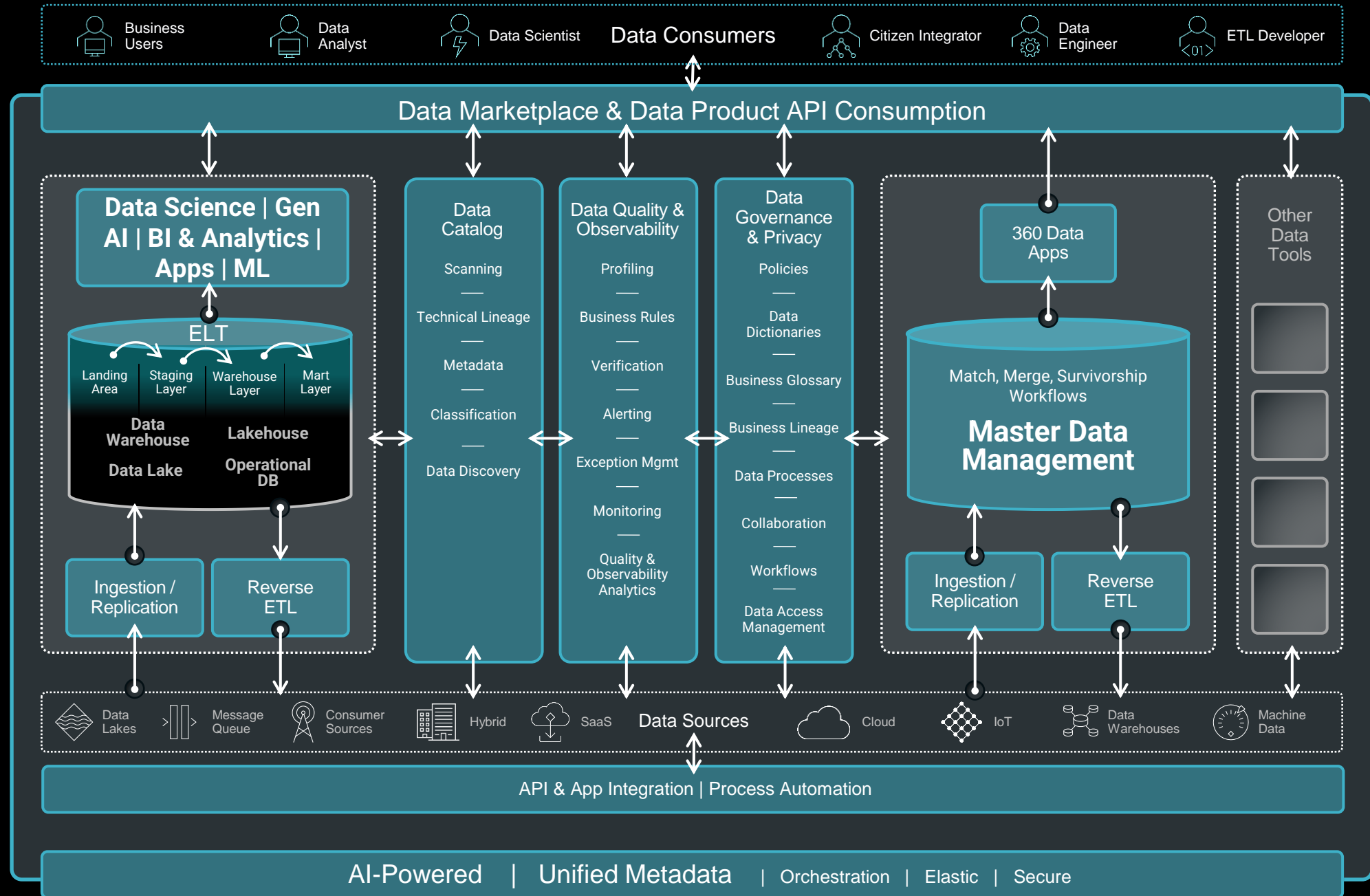
Drive exponential growth in productivity of your existing data professionals with AI

A decorative circular graphic on the left side of the slide. It features a dark blue outer ring with a light blue inner ring. Between these rings are various colored segments (orange, yellow, green, blue, purple) and binary code (0s and 1s) in white and light blue.

Wrap up:

AI Needs a Modern Data Architecture
& Data Management Needs AI

IDMC: AI-Powered Modern Data Architecture





Want to Continue the Conversation?

<https://www.linkedin.com/in/danielhein1>

dhein@informatica.com



Informatica
Experience
Lounge

Thank you

Where data
& AI come to **LIFE**