

Modern Data Management in the era of AI

Jonathon Bowring

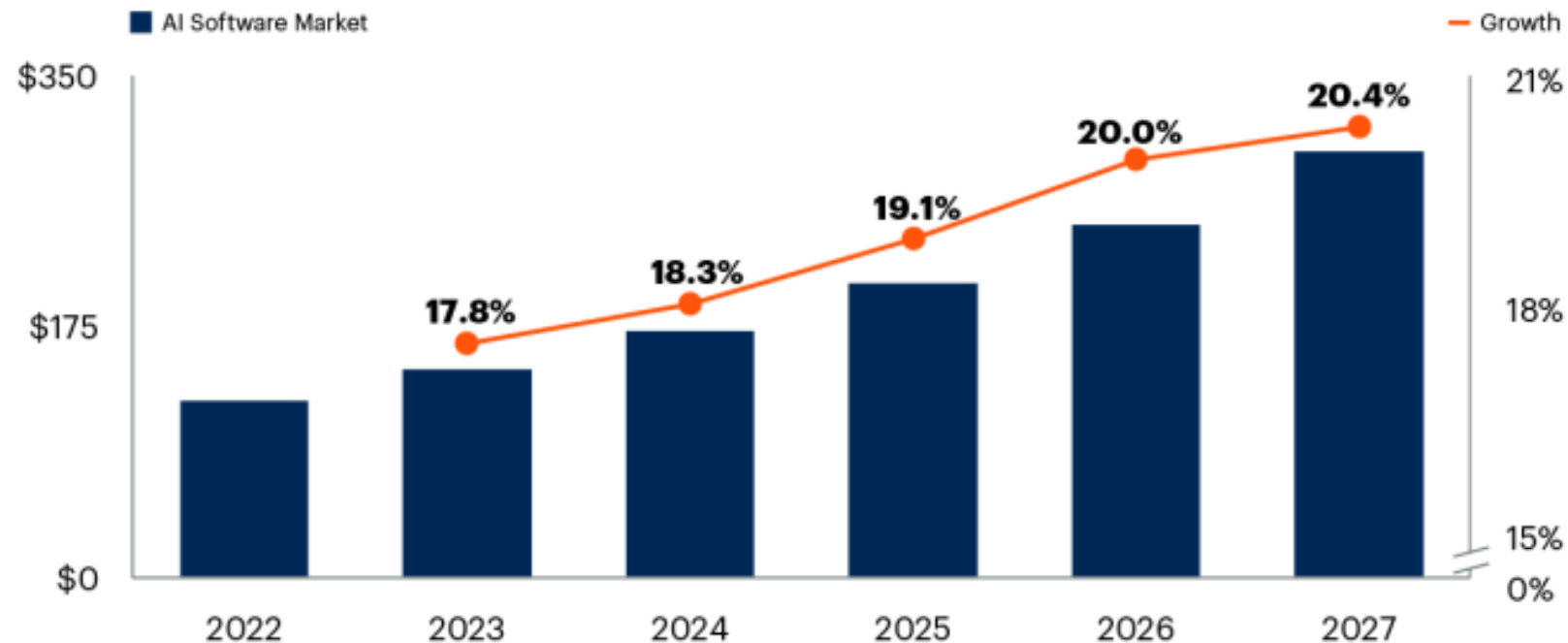




AI Market

According to Gartner

AI Software Forecast and Growth
In Millions of U.S. Dollars

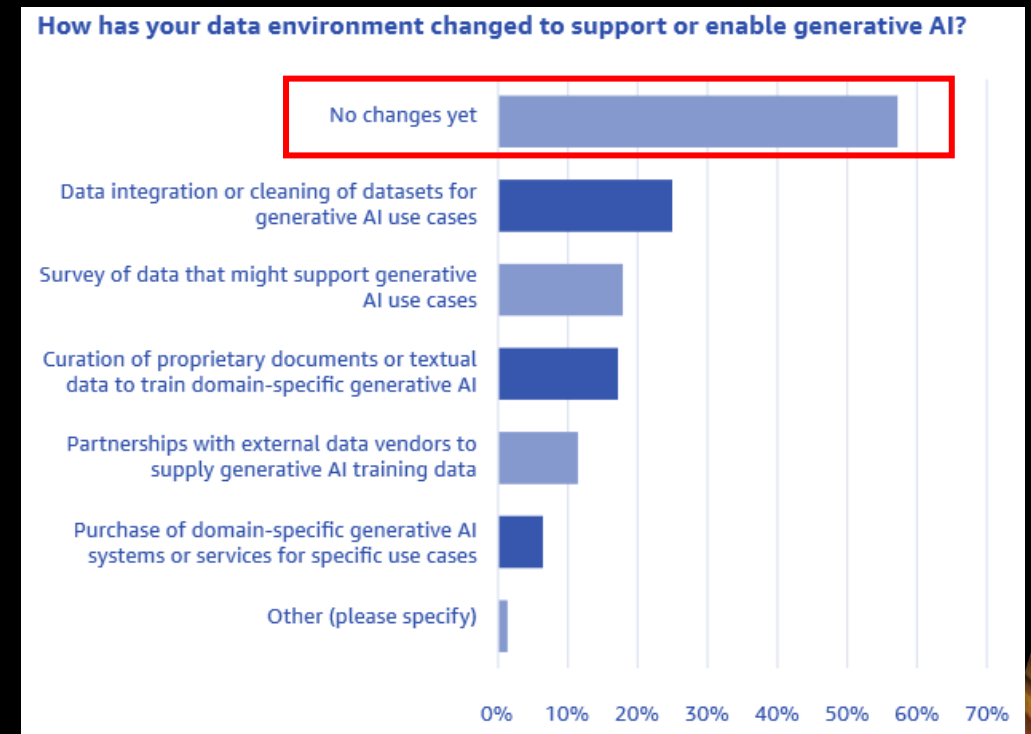


- By 2027, spending on AI software will grow to **\$297.9 billion** with a CAGR of **19.1%**.
- Over the next five years, the market growth will accelerate from 17.8% to reach **20.4%** in 2027.
- Generative AI software spend will rise from 8% of AI software in 2023 to **35%** by 2027.

NOTE: [1] "Forecast Analysis: Artificial Intelligence Software, 2023-2027, Worldwide." (2023), [Gartner.com](https://www.gartner.com)

Generative AI Opportunities

93% of CDOs Feel Data Strategy is Crucial to Generative AI Value



The global generative AI market size is projected to grow from \$43.87 billion in 2023 to **\$667.96** billion by 2030, at a **CAGR of 47.5%** during the forecast period¹

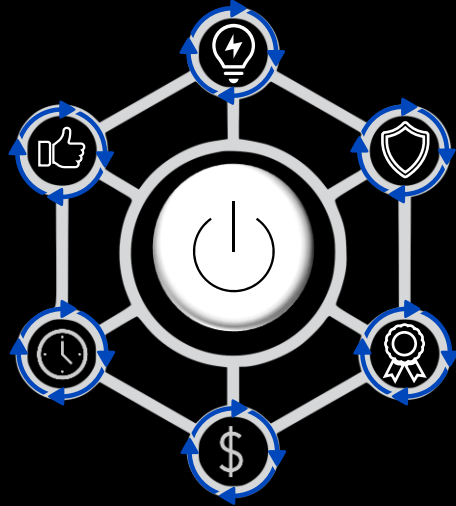
SOURCE: "CDO Agenda 2024: Navigating Data and Generative AI Frontiers." (2023), [AWSstatic.com](https://awsstatic.com)

NOTE: [1] "Generative AI Market Size." (2023), [FortuneBusinessInsights.com](https://fortunebusinessinsights.com)

© Informatica. Proprietary and Confidential.

Enterprise Use Cases

Generative AI Examples by Industry



Retail

- + Personalised Product Recommendations
- + Inventory Management
- + Customer Experience Enhancement
- + Price Optimisation
- + Supply Chain Optimisation

Financial Services

- + Fraud Detection and Prevention
- + Personalised Financial Advice
- + Credit Scoring Models
- + Risk Management
- + Automated Document Processing

Consumer Goods

- + Product Design and Development
- + Market Trend Analysis
- + Quality Control
- + Demand Forecasting
- + Targeted Marketing

Healthcare

- + Drug Discovery and Development
- + Personalised Medicine
- + Medical Imaging Analysis
- + Predictive Analytics for Patient Care
- + Synthetic Data Generation for Research

Oil and Gas

- + Exploration and Reservoir Simulation
- + Predictive Maintenance
- + Supply and Demand Forecasting
- + Safety and Risk Assessment
- + Energy Efficiency Optimisation

Common AI Challenges

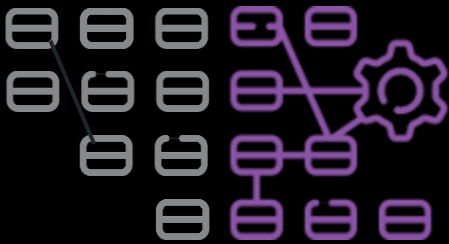
Key points to consider



Regulatory and Legal Challenges

74% surveyed listed Data Privacy as their top concern.¹

- The evolving nature of generative AI raises regulatory and legal challenges.
- Businesses must stay abreast of relevant laws and regulations, such as intellectual property rights and data protection, to ensure compliance and mitigate legal risks.



Data Availability and Complexity

68% surveyed listed Misinformation as their next highest concern.¹

- Implementing and maintaining generative AI systems requires vast amounts of high-quality, holistic, and governed data to effectively train the data models
- Businesses cannot afford to invest in training their staff, collaborating with external experts, or hiring a village to ensure effective AI outcomes



Data Quality & Reputation

- AI relies on vast amounts of data, raising concerns about data quality.
- Businesses must implement robust measures to ensure data is accurately observed information to maintain customer trust and safeguard their reputation.

NOTE: [1] "Who's Afraid of AI?" (2023), [WPROMOTEcom](#)

There is NO AI without Trusted Data

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

WHY AI FAILS?

POOR QUALITY DATA

- Data is critical when it comes to securing successful AI outcomes.
- Trusted data helps AI learn to make a recommendation or perform a task.
- The more fit for use data sourced, the better the outcomes

DO YOU WANT TO PREVENT AI FAILURE?

Cost of Poor Data Management¹

- *"Average cost of poor data quality at \$12.8 million per year"*
- *"AI models depends entirely on the accessibility, quality, granularity, and latency of your organisation's data"*

Data and AI bias

- 86% of Australian business and IT decision makers believe data bias will become a bigger concern as AI/ML use increases
- Yet over a third (34%) have not looked at technology and tools to help in tackling it.

MS Chatbot Spew Racist Tweets²

- Used uncleansed Twitter interactions for training ML models
- Within 16 hours, more than 95,000 tweets posted, and rapidly turned overtly racist, misogynist, and anti-Semitic

Bad AI Leads to Attorney Sanctions²

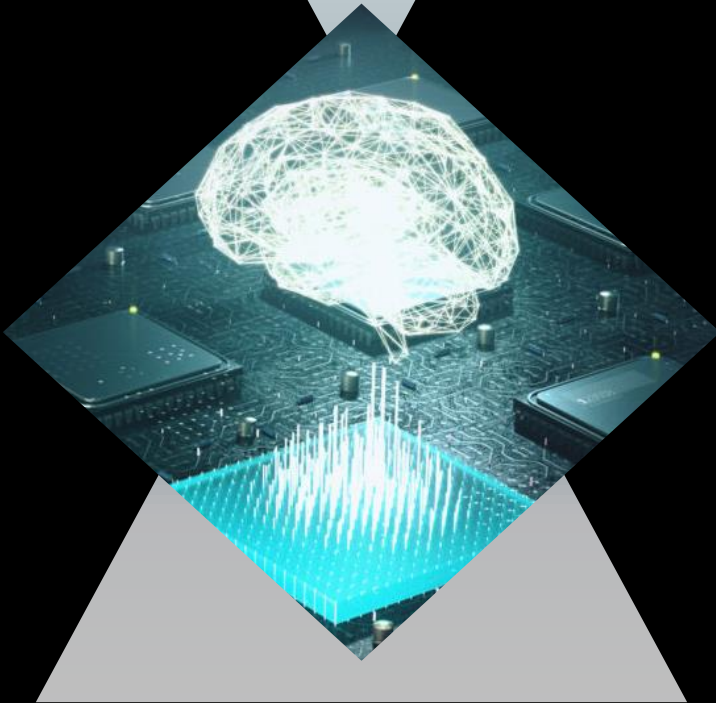
- Used generative AI to find prior cases to support a case
- At least 6 of the AI findings did NOT exist
- As of June 2023, Schwartz was facing possible sanctions by the court.

Misdiagnosis due to Errant AI²

- Hospitals used AI to diagnose & triage COVID-19 patients faster
- AI model was flawed because it was trained on the wrong data set
- Misabeled data or data from unknown sources was a common culprit.

NOTE: [1] "The steep cost of a poor data management strategy." (2023), [CIO.com](#) NOTE: [2] "Famous Analytics and AI Disasters." (2023), [CIO.com](#)

AI WORLD



The Great Data Divide

- Unequal data access
- Low use of data to inform decision-making; 77% cite data intelligence is a challenge
- CDOs rank poor data literacy as one of the top internal roadblocks

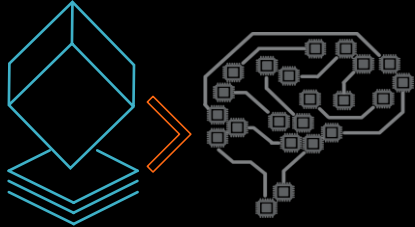
Decentralised Data Products Dilemma

- Decentralised teams expected to produce datasets without relevant skills or tools
- Distributed data creation leading to declining data quality
- Dark data proliferation; 50% of useful data is not analysed

Overwhelmed Data Teams

- Slow data supply chains due to overwhelming demand for data engineering resources
- Data Engineers spend 40% of their time in documentation, requirements gathering and maintaining pipelines

3 Key AI Themes



Data Management for AI

Your own Artificial Intelligence (AI) solutions

- Need holistic, high-quality, governed data management (DM) to deliver trusted AI insights to fuel your competitive advantage.



AI-Powered Data Management

AI Powered Data Management Solutions

- AI-powered insights to data management
- Dramatically accelerates data discovery, integration, quality, governance, and master data management.



GenAI Data Management

Data to All

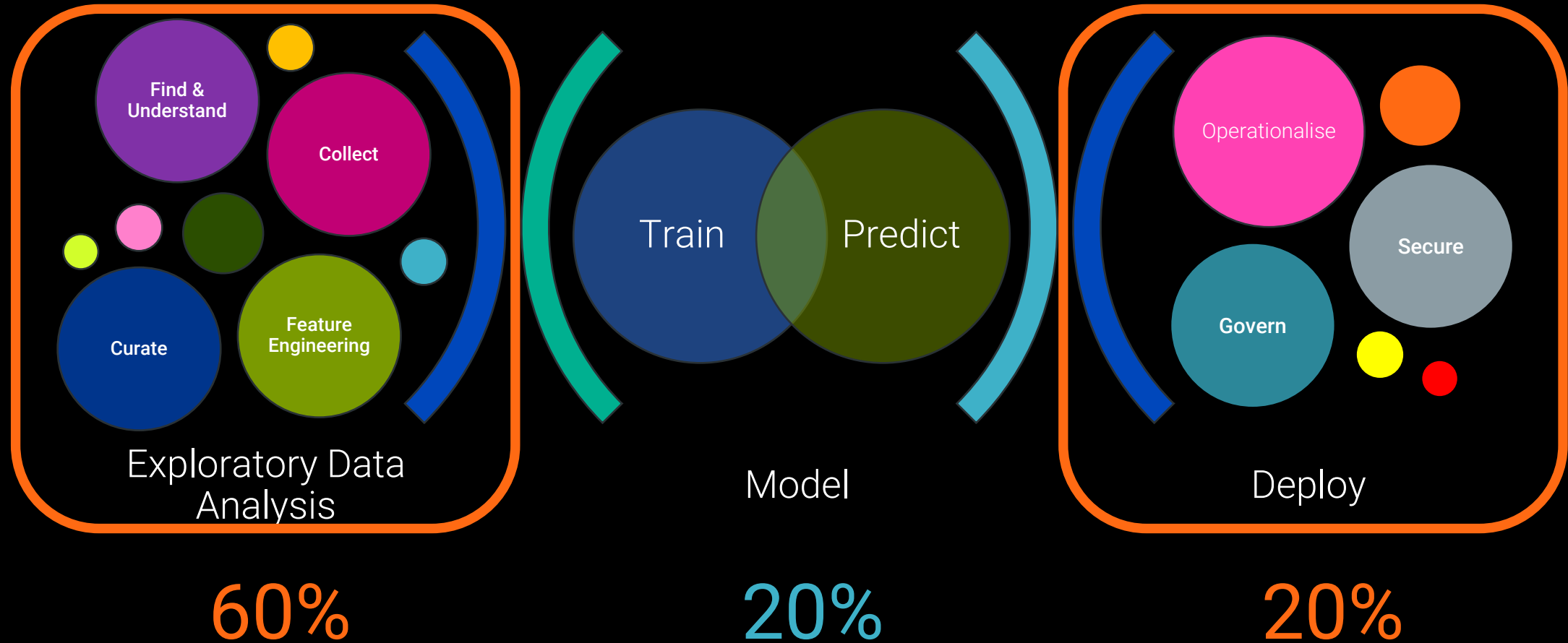
- Dramatically simplifies how organisations access and manage data
- Turbocharges data democratisation
- Exponentially increases the productivity of data teams

First AI Theme

Data Management for AI

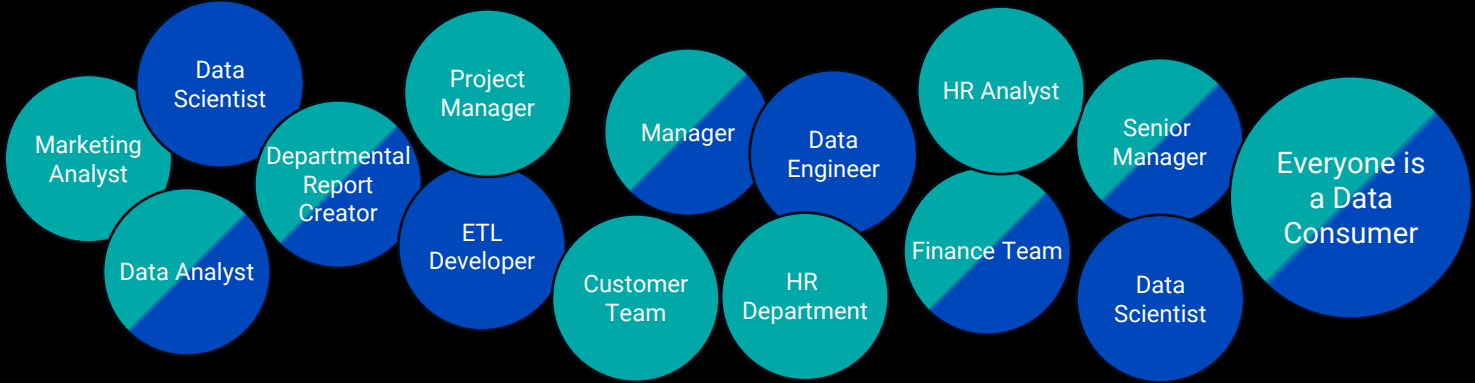
The Data Science Workflow

Streamlined Data Management is the Data Science Supercharger



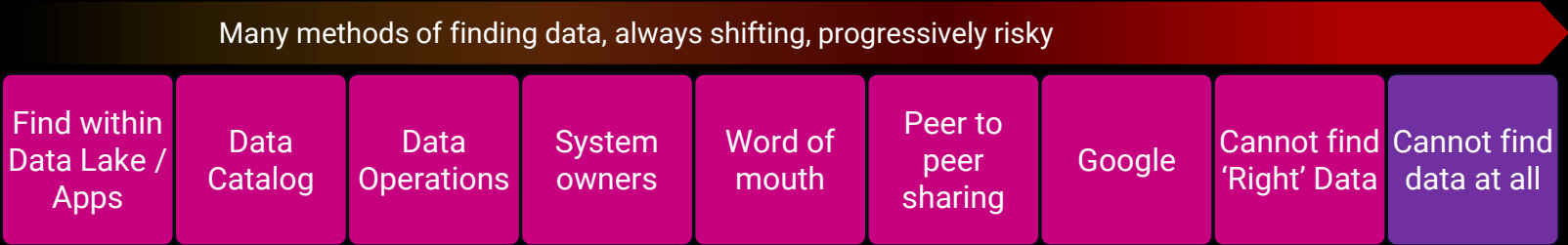
Discovering and Understanding Data

Data Consumers



- Technical skillset
- Non-technical skillset
- Hybrid skillset

Finding Data



Demand and usage **tracking is patchy**, as access requests are broadly distributed across the data estate

Can't find the right data?

Ensuring that Data is Fit for Business Use

Profile



Data Quality Checks:
Turn by turn
directions for data
engineering instead
of "heading west".

Standardise

Example 1: KitKat
Kit-Cat
Kit Kat } → **Kit-Kat**

Example 2: North Central
Midwest
Great Lakes } → **Central**

Parse

Text Data: Call Center Comments, Social
Media, Product Descriptions

I love my Pink 64GB iPad Pro!!!

Color	Size	Product	Model
Pink	64 GB	iPad	Pro

Fuzzy Lookup/Match

Source Data

System of Record

Land O Lakes	Land O'Lakes
Biluxi	Biloxi
KitKat	Kit-Kat

Validate

- ✓ Currency code must be consistent with country code
- ✓ Employee ID must be unique
- ✓ If customer tier is bronze then max credit is 1000
- ✓ All ICD10 codes must have a verified description

Enrich

AddressL1: 1008 Avenue of the Americas
AddressL2: Suite 7
City: New York
State: NY
Zip Code: 10018-5402
Longitude: 40.7325525
Latitude: -74.004970

Ability to Source Data

Fragmented Landscape of Applications, Data, People...

Applications/SaaS



ERP



SaaS



IOT



CRM

~500+

cloud apps in use in most organisations

Data



Structured Data



Real-time



Unstructured Data



Batch Data

10 petabytes

of data stored by average enterprise.
180 zettabytes total by 2025

People



Data Architect



Business User



Developer



CIO/IT Leader



CDO



Data Scientist

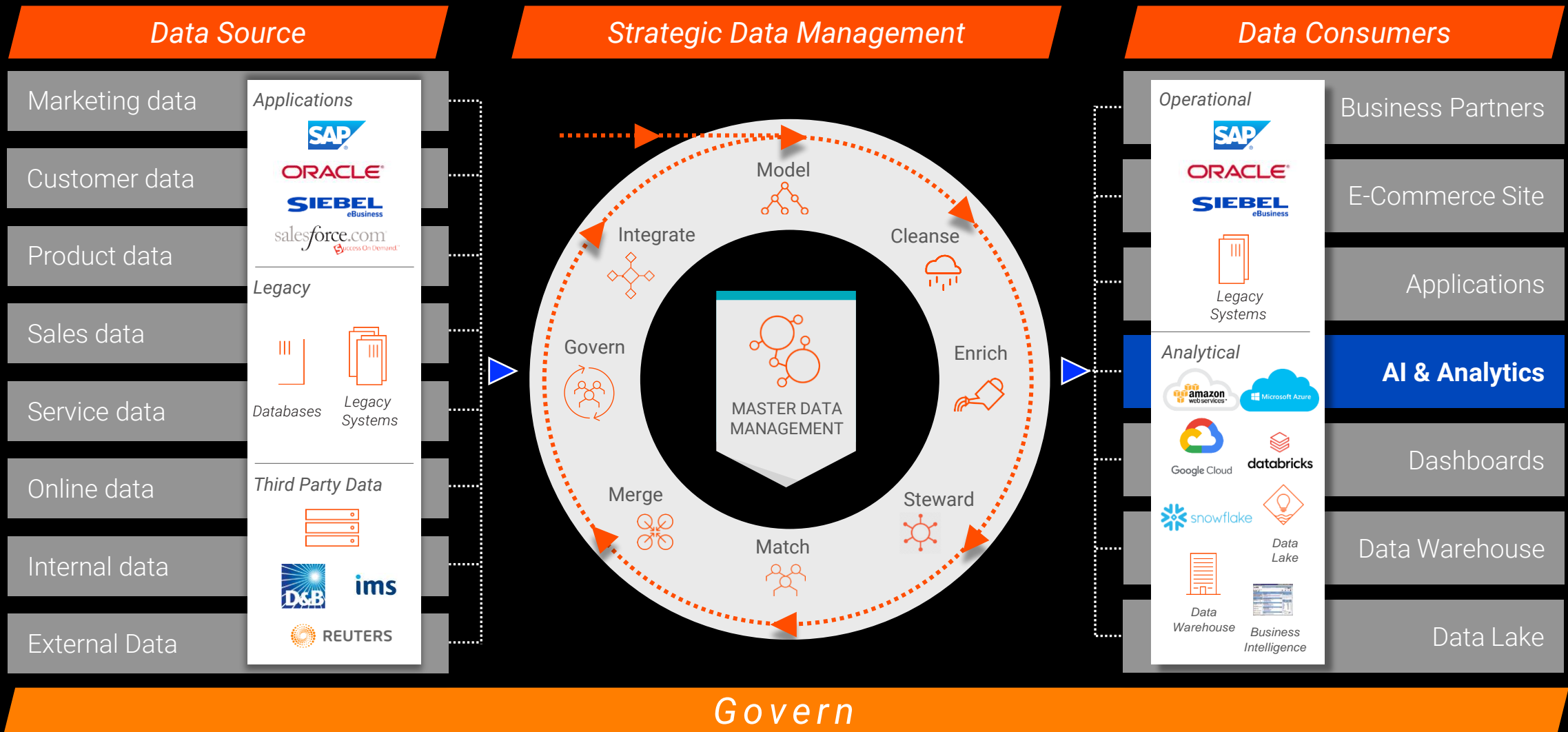


Data Engineer

88% of execs

believe technology democratisation is critical to ignite innovation

Consolidate Data into a Single Source of Truth



Single View of Your Critical Data Domains Enabling Better AI Outcomes

Trusted and Timely Data with Data Governance

Governance



Simplify trusted data delivery and end to end data governance, including AI Models governance

Democratisation



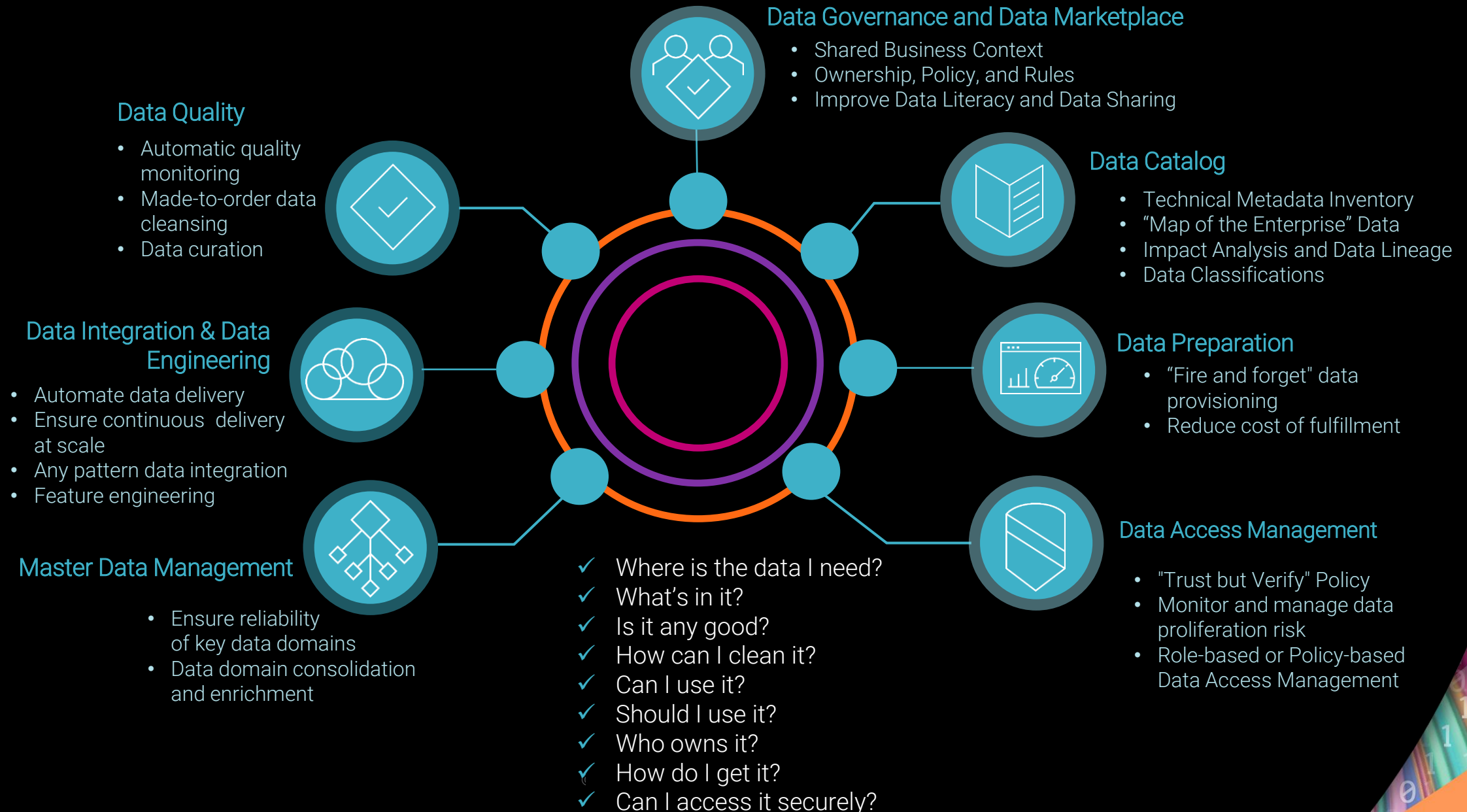
Accelerate time to value for data-driven outcomes, connecting data producers and consumers

Observability



Achieve reliable analytics insights by monitoring pipelines, data science models, and improving data accuracy

Data Management Capabilities to drive AI Initiatives





Garbage IN



AI



Garbage OUT

Second AI Theme

AI for Data Management

Where data
& AI come to **LIFE**

Challenges that AI paired with Data Management will address

Scale & Agility



Policy & Trust



Skills & Literacy



AI Generated Classifications



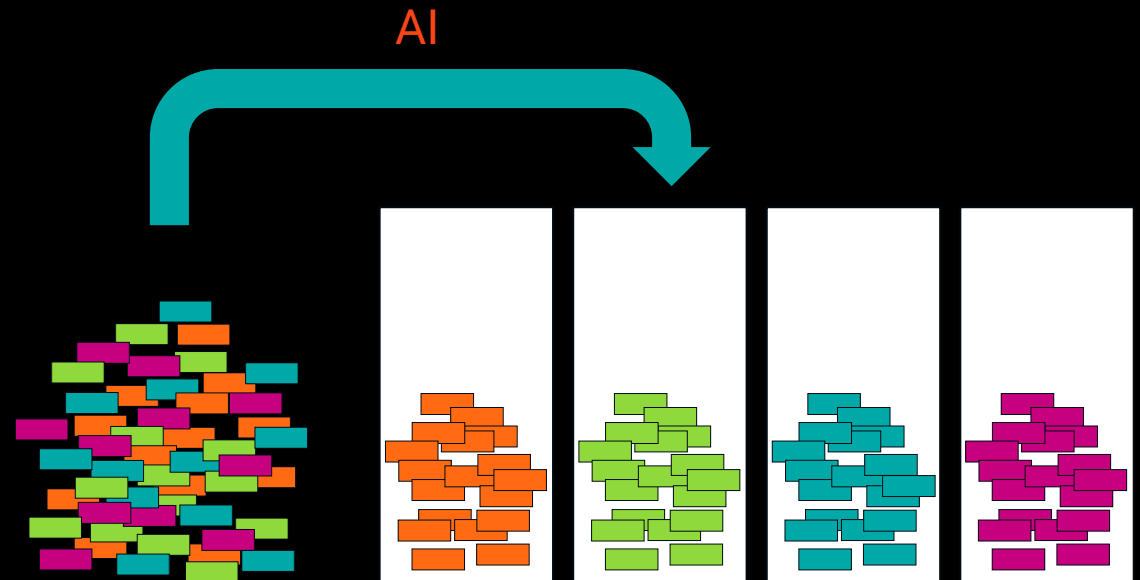
Key Highlights

- Conventional classification methods based on Regex, algorithmic pattern recognition and reference data requires a lot of manual development work initially to create organization specific classifications
- AI Generated Classifications can work without any access to data, using only metadata to generate classifications



Benefits

- AI Generated Classifications will automate the process of classification creation by using organization's metadata and data patterns
 - This is expected to reduce time to curate and create classifications by 70%, saving each organization by upwards of \$100K per year



AI Inferred Data Lineage



Key Highlights

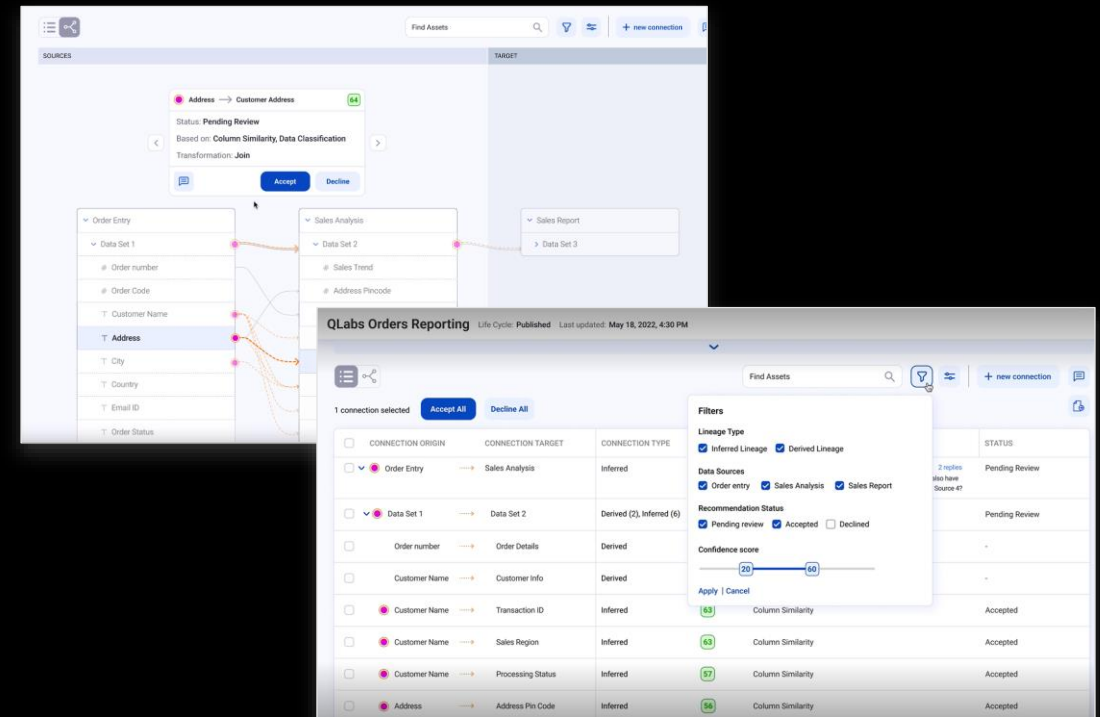
- Complement existing derived lineage from auto-connection assignment and code logic parsing
- Fill the gaps and extend derived data lineage with inferred lineage
- Curation Inferred Lineage by subject matter experts



Benefits

- Enhance productivity of experienced data management professionals with fully automated workflows
- Democratize data and data management by enabling non-technical users create specifications and perform basic data management tasks on their own

- Visual curation on lineage diagram
- Bulk curation on tabular format
- Filtering
- Collaboration and Comments



AI Powered Data Access Management



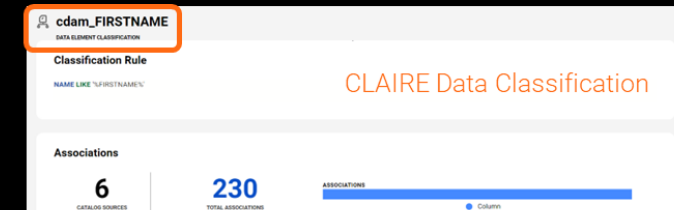
Key Highlights

- AI Generated Classifications allow to identify which fields are sensitive
- Access Control Policies are defined against Data Classifications
- CDAM applies policies dynamically on read or at runtime on write (based on context)



Benefits

- Minimize the potential for human error in manual classification processes
- Robust risk mitigation measures against unauthorized access to data
- Enhance efficiency with faster time-to-value and lower TCO, avoiding multi-party solution



CDAM Access Control Policy

Rule Names	Conditions	Transformations
Ofuscate_PersonName	Trigger this rule if: → User Group is any of [Data_Consumers]	Assign all of the following transformations to data classes: → Replace [cdam_BUSINESS_ID] with regex → Replace [cdam_FIRSTNAME] with consistent regex → Replace [cdam_LAST_NAME] with ****

User with Privileges

FIRST_NAME	SECOND_NAME	EMPLOYEE_NUMBER	COUNTRY	JOB_TITLE
Salvador	Meyer	84047	Ireland	Legal Administrator
Sara	Townsend	55686	France	Network Administrator
Alexis	Deleon	55801	Ireland	Compliance Officer
Gabrielle	Cuevas	13962	Germany	Media Buyer
Brecken	Berg	95088	Germany	Electrical Engineer
Emmalyn	Benjamin	88642	United Kingdom	IT Project Manager
Kyro	Mayo	91949	Germany	Media Buyer
Aaliya	Booth	78848	United Kingdom	Mechanical Engineer
Chaim	Ford	61428	Ireland	Compliance Officer
Alexandra	Sloan	49395	Ireland	Compliance Officer
Grace	Pamphala	75326	United Kingdom	Compliance Officer

User without Privileges

FIRST_NAME	SECOND_NAME	EMPLOYEE_NUMBER	COUNTRY	JOB_TITLE
Jakob	Adhithy	936042276038	Germany	Court Reporter
Xvlas	Corfcap	444607627894	United Kingdom	Marketing Research Analyst
Pkxq	xbrtcjaj	5484484031640	Germany	Project Manager
Mduo	qilcext	6872587489830	United States	Copywriter
Xqmz	uogzjem	2201799735622	Germany	Cloud Solutions Architect
Lmon	phcbeepz	7088887775853	Germany	Quality Control Analyst
Zrxq	myvaadga	3309725977403	Germany	Marketing Research Analyst
Ypmx	mpukmbad	332215463410	France	Web Developer
Bpwj	qccyvi	569455264405	France	Cybersecurity Analyst
Nuz	rodiezalf	3854298161402	Germany	Media Buyer
Xpic	zswgxmli	6649959732424	United Kingdom	Civil Engineer

AI Generated Data Observability



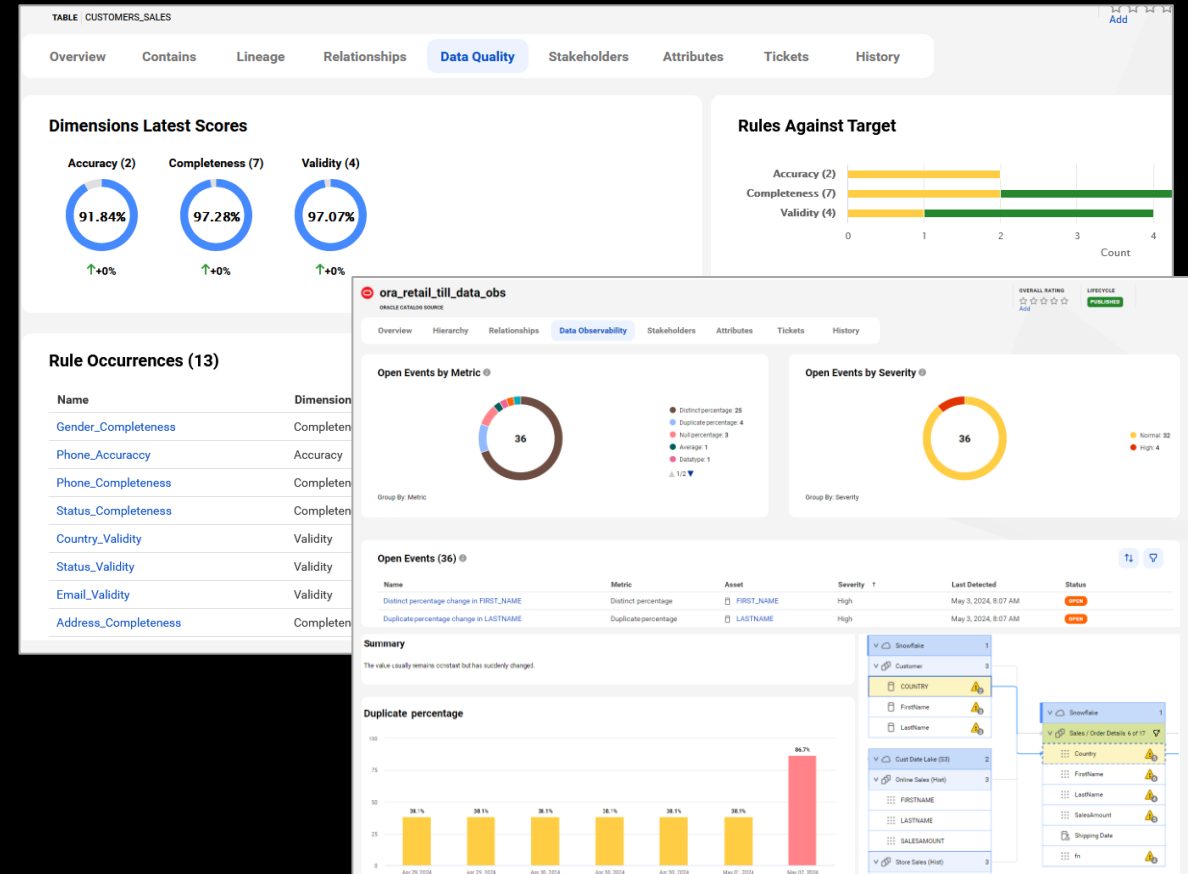
Key Highlights

- Automation of data quality rules application per governance policies
- Data quality execution and monitoring across all instances of the business entity

Benefits



- Data Quality Automation takes away mundane work and saves 100s of hours of work thus reducing the manual work by as much as 80%
- Provides a more complete Data Quality view by automatically increasing coverage to over 60%



AI Data Parsing and Structure Discovery



Key Highlights

- File pattern recognition and structure derivation
- Usage of custom ML-powered NER (named entity recognition) and NLU (natural language understanding) mechanisms to identify fields and field types
- Data and schema drift



Benefits

- Automatic file ingestion and onboarding processes to extract and use information out of complex files (machine data, application data, logs, non-relational formats...)
- Reduce errors to detect PII information, mitigating risk and improving security

The screenshot displays the 'Intelligent Structure 1' interface. The top section, 'Intelligent Structure Details', includes fields for Name (Acme_Log), Location (Default), and Description (Claire on BDM demo). Below these is a 'Sample File' field with 'Acme_log.txt' and a 'Discover Structure' button. A sample log entry is shown below. The bottom section, 'Visual Model', displays a hierarchical tree structure of the data. The root is 'table', which branches into 'element'. 'element' further branches into 'IP', 'value', 'token', 'element', 'statusCode', and 'number'. The 'element' node under 'value' branches into 'datetime' and 'HTTPRequ...'. 'datetime' branches into 'HTTPMethod', 'URI', and 'protocol'. The 'HTTPRequ...' node branches into 'HTTPMethod', 'URI', and 'protocol'. To the right of the visual model is a table with columns A through F, showing data rows with values like IP, value, token, datetime, HTTPMethod, and URI.

A	B	C	D	E	F
IP	value	token	datetime	HTTPMeth...	URI
0.0.0.0.0.0.1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d\l\omcat	[18/Sep/2014:14:54:24 +0300]	GET	/dx-console/ HTTP/1.1' 302 -
0.0.0.0.0.0.1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d\l\omcat	[18/Sep/2014:14:54:25 +0300]	GET	/dx-console/com.informatica.b2b.ds.Main/main.jsp HTTP/1.1' 302 -
0.0.0.0.0.0.1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d\l\omcat	[18/Sep/2014:14:54:25 +0300]	GET	/dx-console/login.jsp HTTP/1.1' 200 4472
0.0.0.0.0.0.1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d\l\omcat	[18/Sep/2014:14:55:47 +0300]	POST	/dx-console/_spring_security_check HTTP/1.1' 302 -
0.0.0.0.0.0.1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d\l\omcat	[18/Sep/2014:14:55:47 +0300]	POST	/dx-console/_spring_security_check HTTP/1.1' 302 -

AI Self Integrating Systems



Key Highlights

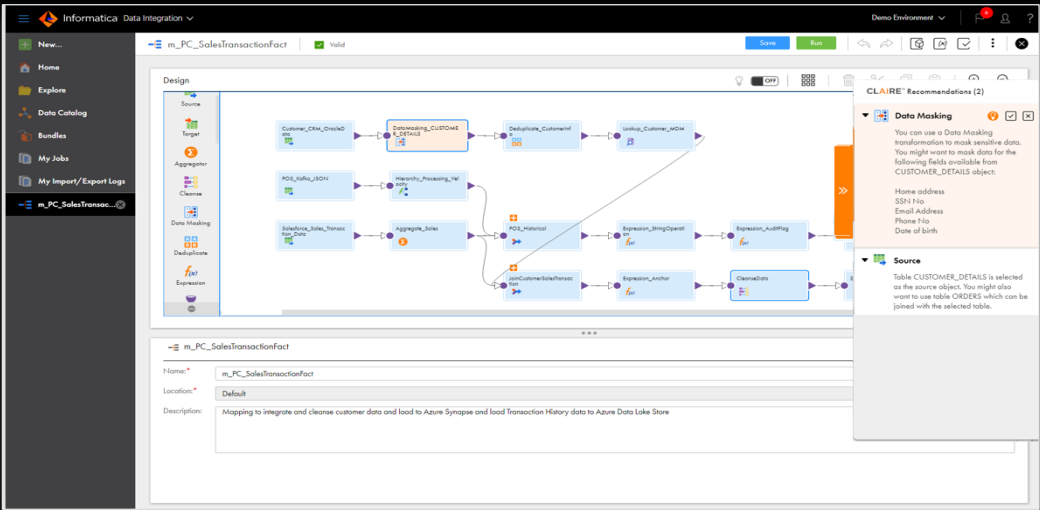
- **Fully automated data pipeline** generation
- Support for unstructured, semi-structured and hierarchical data sources
- Automated creation data masking, cleansing and standardization rules
- Support for join, union, normalization and denormalization of data sources



Benefits

- **Automate, accelerate and simplify** all data management activities
- **Democratize** data and improve productivity for all Personas
- **Fully autonomous DI** experience

Fully Automated Data Pipeline Generation

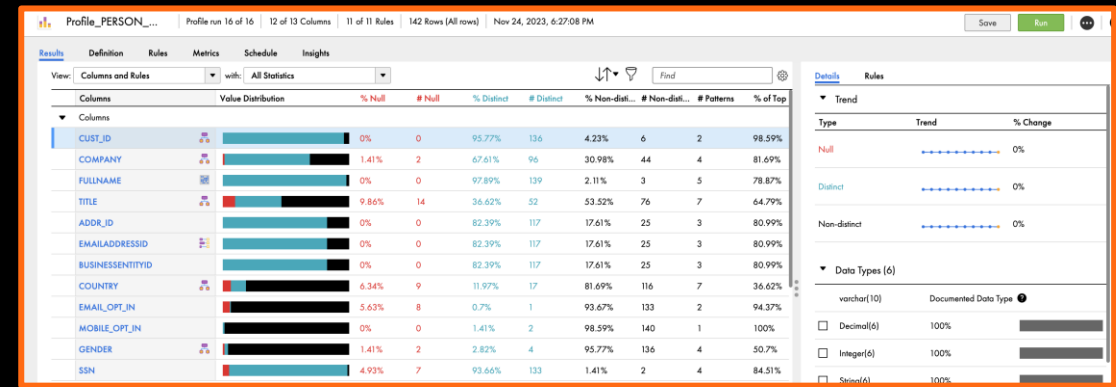
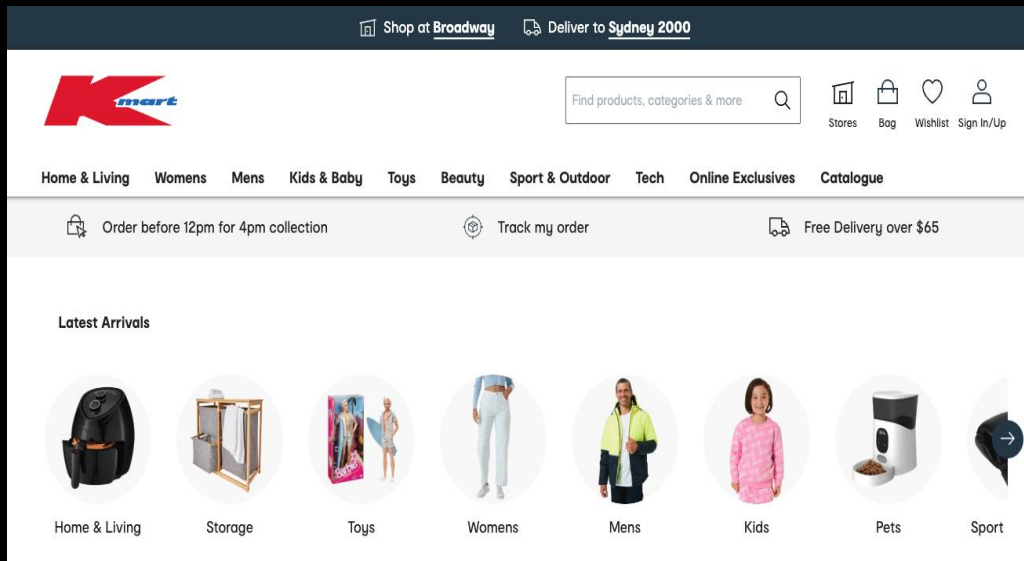


AI

	Design Time	Run Time	Operations
Type	Manual vs. Automated		Avg Time
Manual	Manual		Months/Weeks
Augmented	Manual	Automation	Days
Autonomous	Manual	Automation	Mins/Hours



- Kmart has more than 30,000 employees and operates 234 stores across Australia and New Zealand, with its head office located in Mulgrave, Melbourne.
- Improved data quality; **quadrupled** online sales by **introducing products 4X faster**

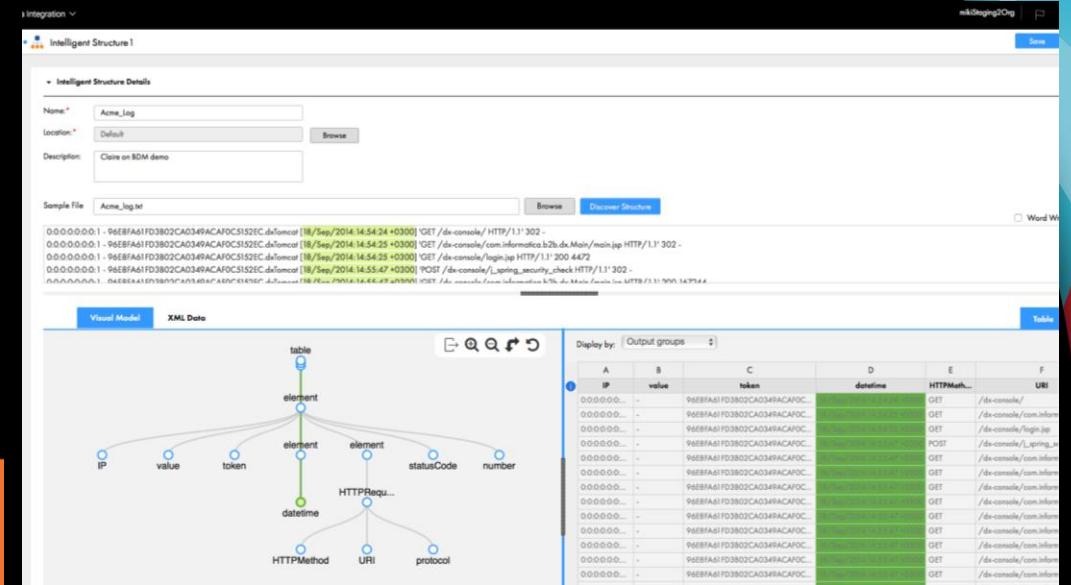
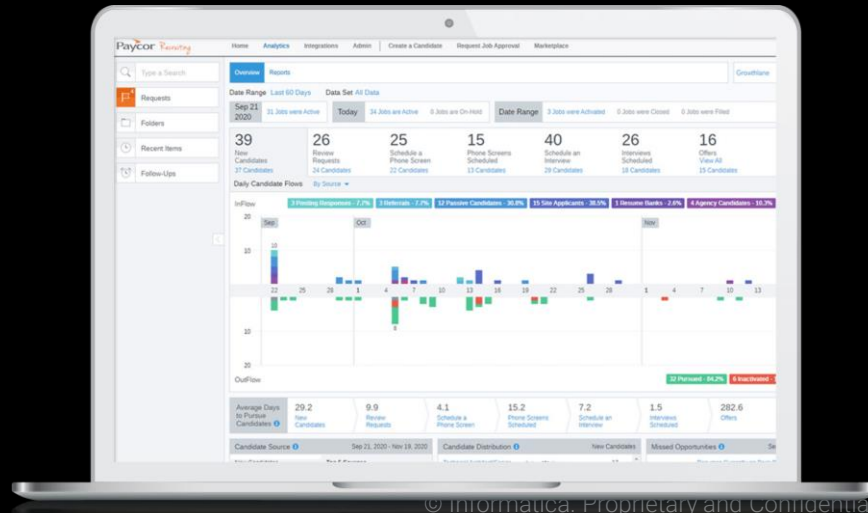


View:	All Insight Types	in:	All Columns	Hide Rejected Insights	Find
Insight Statement	Score	Insight Type	Columns	Status	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	TITLE	Approved	
<input type="checkbox"/> The number of tokens in the column has a high standard deviation.	Medium	Column Token Deviation	TITLE		
<input type="checkbox"/> The data may contain special characters	High	Special Characters	TITLE	Approved	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	FULLNAME	Approved	
<input type="checkbox"/> The number of tokens in the column has a high standard deviation.	Medium	Column Token Deviation	FULLNAME		
<input type="checkbox"/> The data may contain special characters	Medium	Special Characters	FULLNAME	Approved	
<input type="checkbox"/> Numeric values found outside the 95% standard deviation range.	Medium	Number Value Distribution	MOBILE_OPT_IN	Approved	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	ADDR_ID		
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	EMAILADDRESSID		
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or valu...	High	Completeness Check	EMAIL_OPT_IN	Approved	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	CUST_ID		
<input type="checkbox"/> Numeric values found outside the 95% standard deviation range.	Low	Number Value Distribution	CUST_ID		
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or valu...	Low	Completeness Check	GENDER	Approved	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	GENDER		
<input type="checkbox"/> Unusual variation on the number of blank values, null values, and empty values in the column b...	High	Completeness Variation	GENDER	Approved	
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	BUSINESSENTITYID		
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or valu...	High	Completeness Check	COUNTRY		
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	COUNTRY	Approved	
<input type="checkbox"/> The number of tokens in the column has a high standard deviation.	High	Column Token Deviation	COUNTRY		
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or valu...	Medium	Completeness Check	SSN		

Latrobe University

Latrobe University

- Leverages AI to automate data classification as part of its governance program.
- Uses AI to automate onboarding & mastering of students.
- Utilises AI to automate data quality rule creation



- Recognized a 512% ROI moving to an AI-assisted low code/no-code service from Informatica.
- Witnessed a 75% performance improvement in pipeline development with the help of Informatica's CLAIRE engine

Up to
100X
faster than
manual efforts

Up to
95%
reduction in time
to production
for MDM

Up to
96%
reduction in
DQ efforts

Up to
92%
accuracy in
business terms
automation

Up to
1.25M
cost savings
for data
management

Third AI Theme

Gen AI Data Management

Where data
& AI come to **LIFE**

CLAIRE®: AI-Powered Data Management

CLAIRE AI copilot

IN PRODUCT AI-POWERED METADATA INTELLIGENCE AND AUTOMATION



AUTOMATION



INTELLIGENCE



CLAIRE®
UNIFIED METADATA
INTELLIGENCE



Data Management Automation

Activated Metadata to provide automated data management capabilities

Reduced manual effort and improved operational efficiency across data management tasks

Proactive recommendations and alerts for potential issues

INTELLIGENT GLOSSARY

SMART DOMAINS

COLUMN SIMILARITY

ENTITY MATCHING

SCHEMA MATCHING

DATA QUALITY INSIGHTS

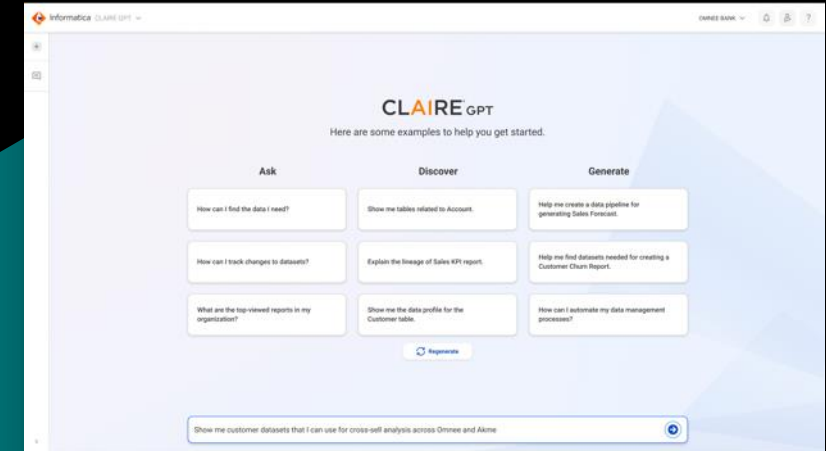
JOIN RECOMMENDATIONS

NEXT TX RECOMMENDATIONS

DATASET RECOMMENDATIONS

CLAIRE GPT

NATURAL LANGUAGE INTERFACE TO DATA



Generative AI-powered Data Management

Enhance productivity of experienced data management professionals with fully automated workflows

Democratise data and data management by enabling non-technical users to create specifications and perform basic data management tasks on their own

Reduce data management costs

NL TO ETL/ELT

NL TO DQ

NL TO PREP

NL TO GOVERNANCE

NL-BASED DISCOVERY

NL-BASED EXPLORATION

NL-BASED TESTING AND DOCUMENTATION

AUTOMATED FINOPS

CROSS PRODUCT EXP.

AI COPILOT

IN-CONTEXT DATA INTELLIGENCE

OPEN, POLYGLOT

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

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...





What Kind of Data Quality does @RETAIL_CUSTOMERS

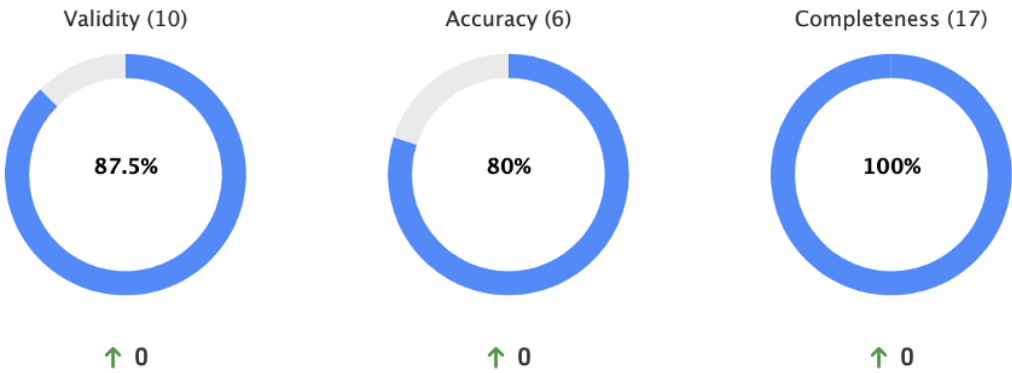


Finding out the kind of data quality does RETAIL_CUSTOMERS

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:





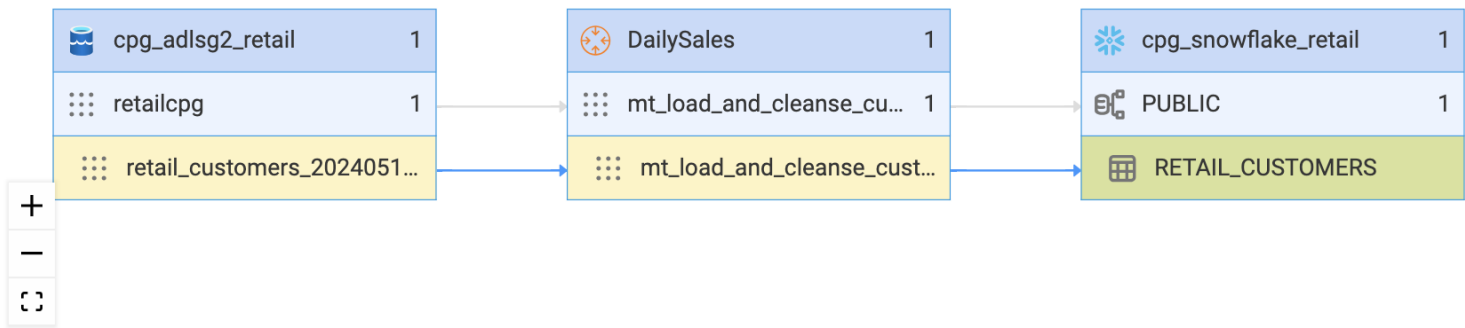
Show the lineage for @RETAIL_CUSTOMERS



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...





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...





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...



Viewing sample data: 100 rows and 3 data types

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

CG

Create a mapping for the above

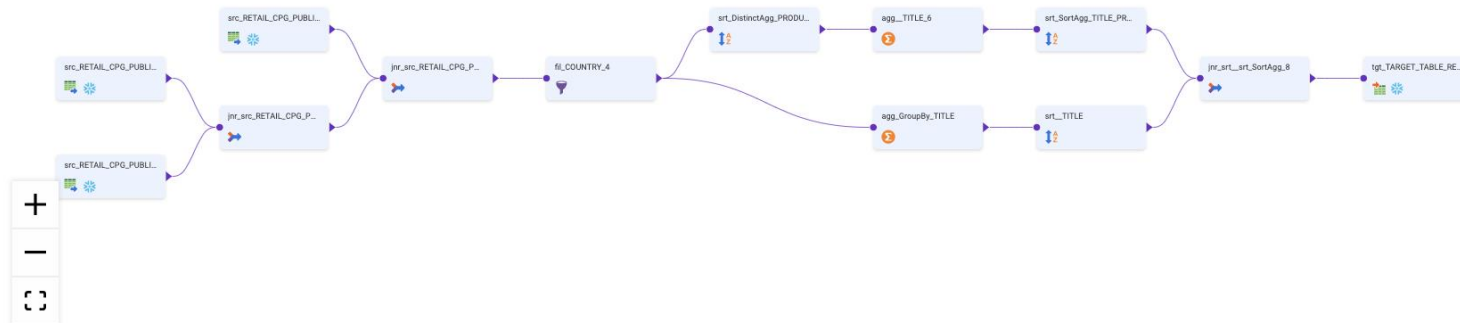


Creating a mapping

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



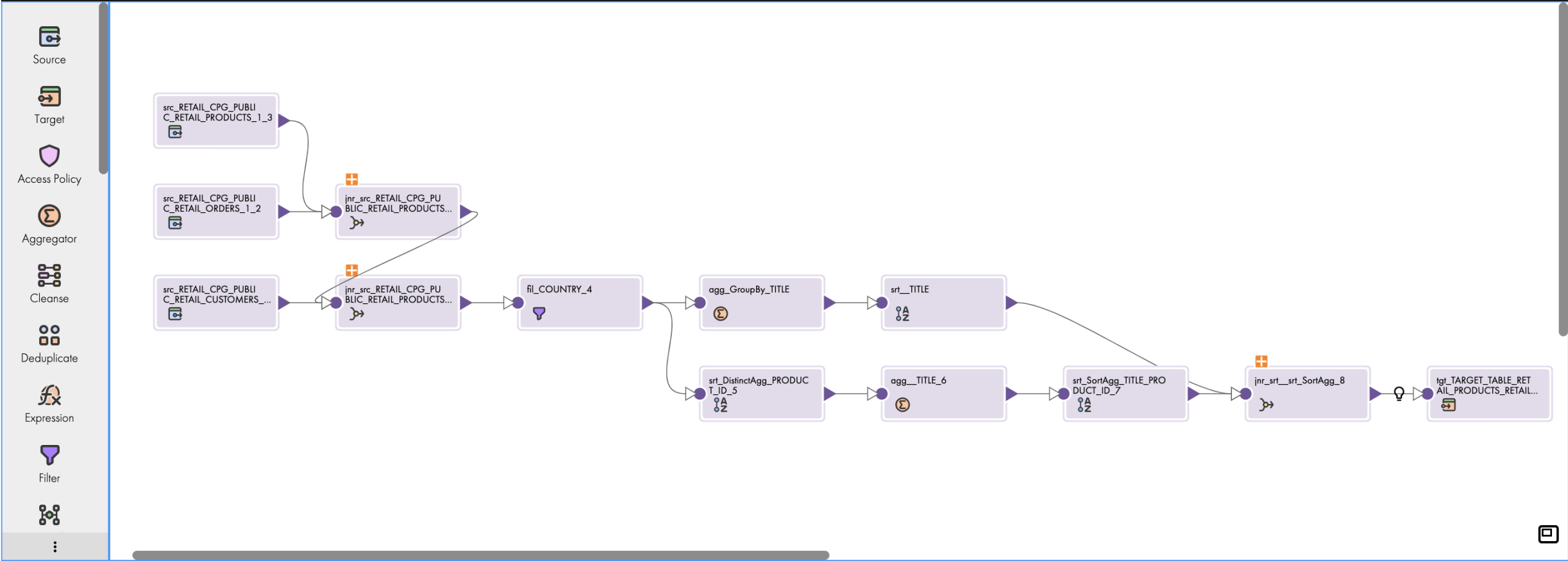
m_Snowflake_src_RETAIL_CPG_PUBLIC_RETAIL_CUSTOMERS_1_1_tgt_TARGET_TABLE_RETAIL_PRODUCTS_RE...

[Open in Data Integration](#)

Ask CLAIRE...



Design



Mapping3



Name:*

Mapping3

Location:*

Default

Browse

Description:

Increased Productivity With CLAIRe GPT

Data Engineer + CLAIRe GPT: Data pipeline creation



15 Weeks



8 Weeks

CLAIRe GPT

CLAIRe BENEFITS

CLAIRe GPT serves as an assistant SME helping data engineers understand the business context.

CLAIRe GPT helps find the right datasets needed for the task directly from the requirements spec.

CLAIRe GPT helps data engineers quickly explore the datasets and find data quality issues.

CLAIRe GPT generates the first draft of the pipeline from the finalized requirement spec. It also helps document and test the pipeline.

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



Great data for great AI

Informatica

Where data
& AI come to **LIFE**

DATA CONSUMERS



ETL Developer



Data Engineer



Citizen Integrator



Data Scientist



Data Analyst



Business Users

Intelligent Data Management Cloud™

DISCOVER &
UNDERSTAND



DATA
CATALOG

ACCESS &
INTEGRATE



DATA INTEGRATION &
ENGINEERING

CONNECT &
AUTOMATE



API & APP
INTEGRATION

CLEANSE &
TRUST



DATA QUALITY &
OBSERVABILITY

MASTER &
RELATE



MDM & 360
APPLICATIONS

GOVERN &
PROTECT



GOVERNANCE &
PRIVACY

SHARE &
DEMOCRATISE



DATA
MARKETPLACE

CLAIRE®

AI-Powered Metadata Intelligence & Automation

Connectivity

Metadata System of Record

DATA SOURCES



SaaS Apps
Sources

+



Mainframe



Applications



Databases

On-premises
Sources

+



IoT



Machine Data

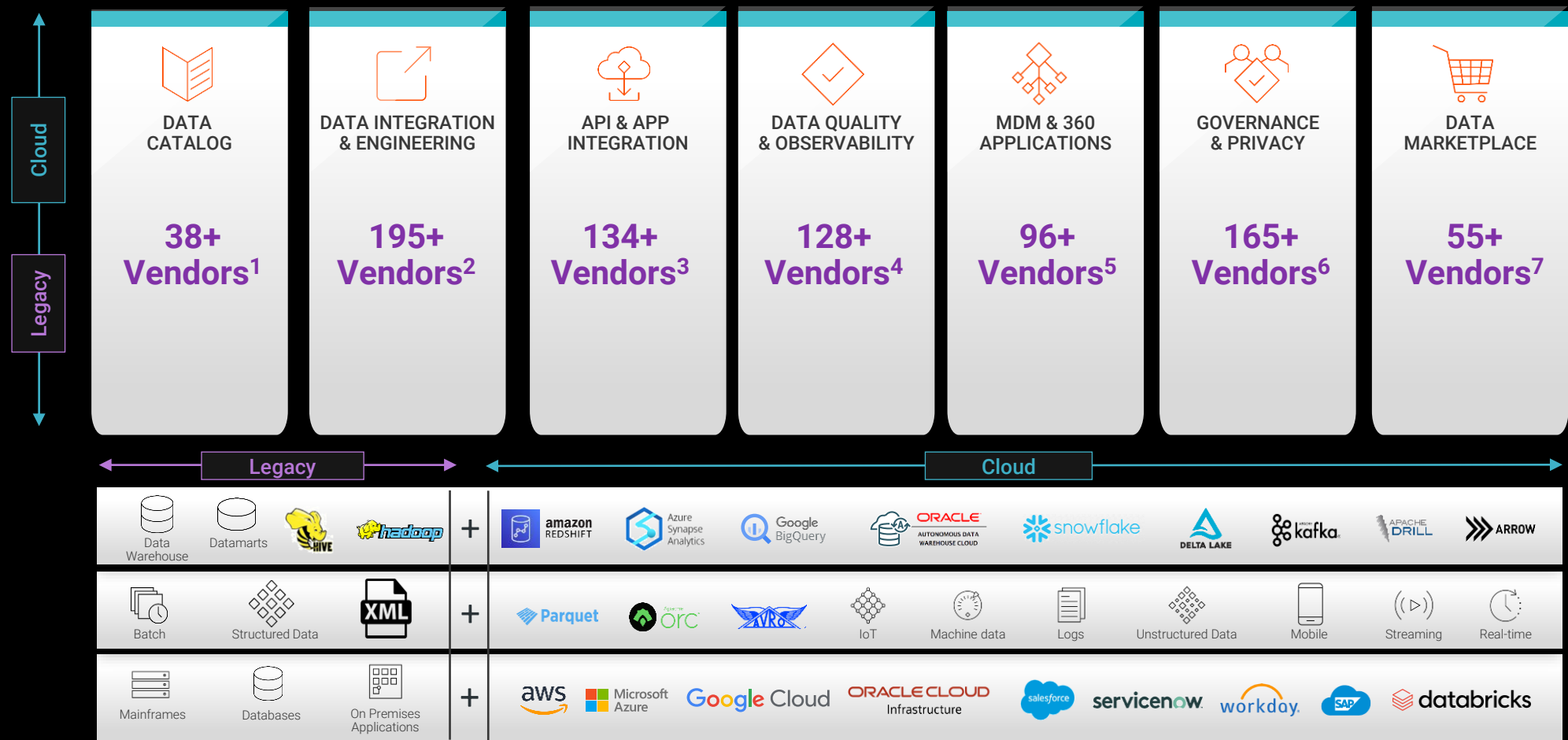


Logs

Real-time /
Streaming
Sources

Why Is Data Management Hard and Complex?

Numerous Vendors Across Multi-Cloud and Hybrid Ecosystems



¹Source: [Trust Radius, Data Catalog Software, May 2023](#)

²Source: [Trust Radius, Data Integration Tools Software, May 2023](#)

³Source: [Trust Radius, API Management Tools, May 2023](#)

⁴Source: [Trust Radius, Data Quality Software, May 2023](#)

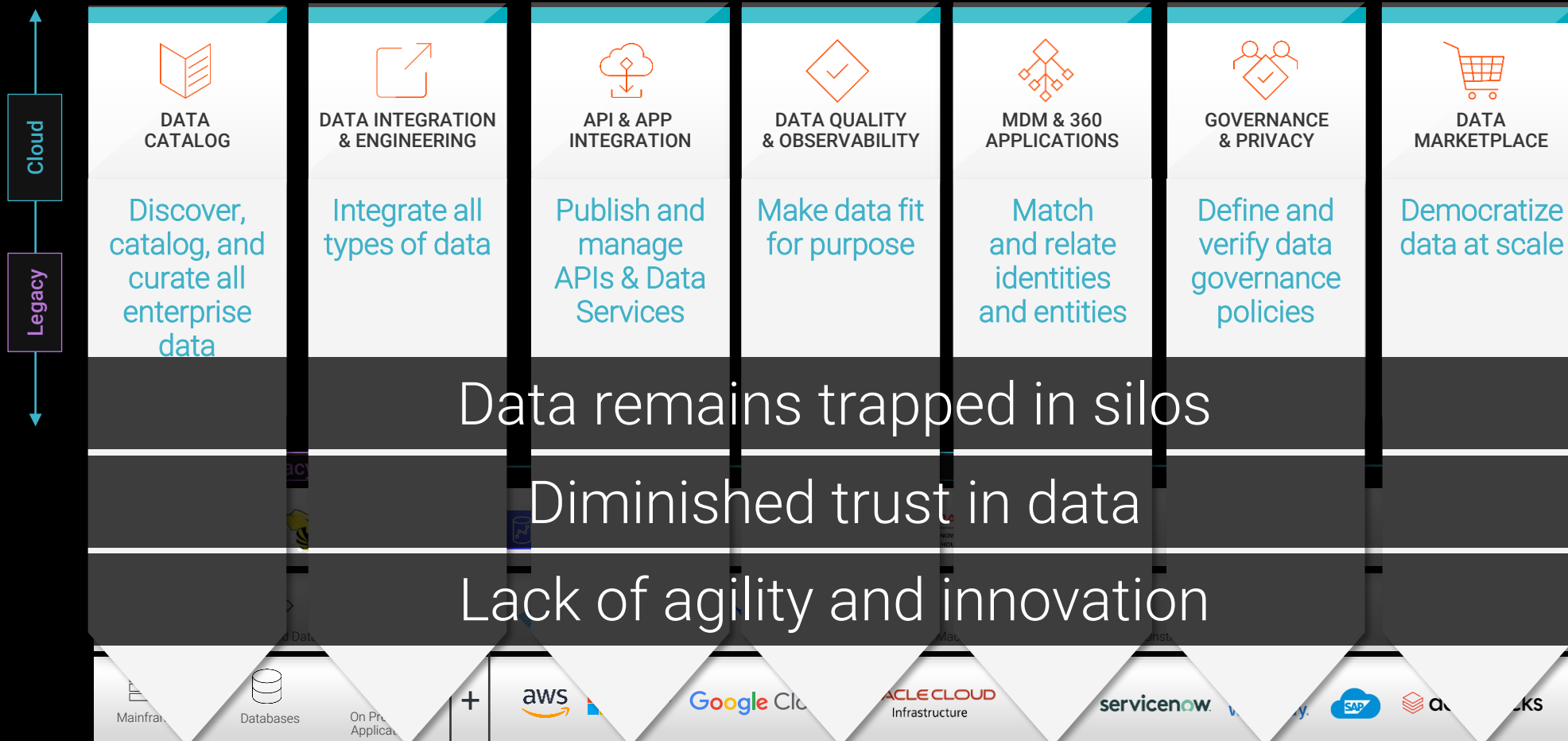
⁵Source: [Trust Radius, Master Data Management \(MDM\) Tools, May 2023](#)

⁶Source: [Trust Radius, Data Governance Software, May 2023](#)

⁷Source: [Trust Radius, Data Collaboration Tools, May 2023](#)

Piecemealed Solutions

Force Customers to “**stitch**” niche provider solutions together

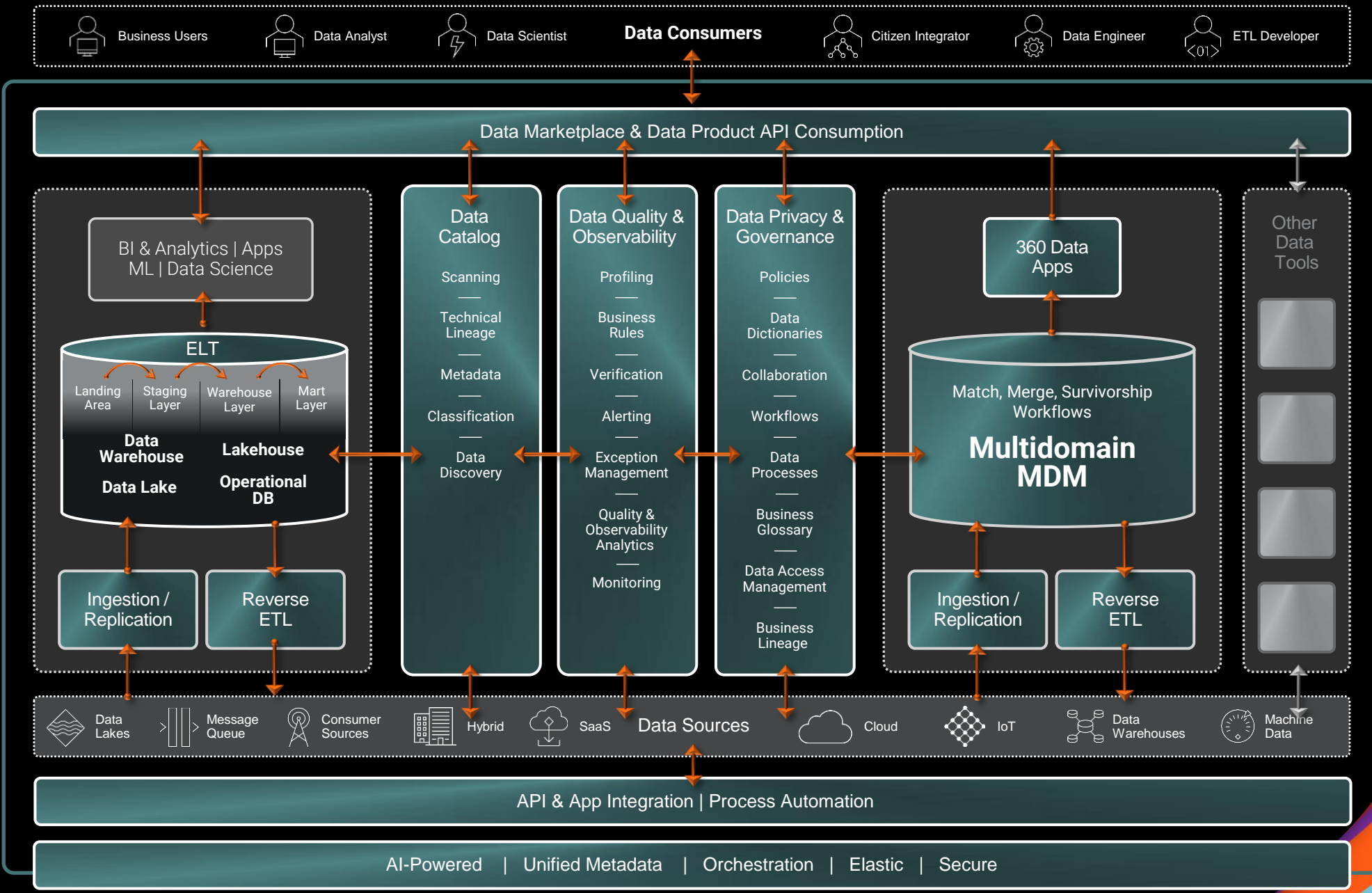


```

0 0 0 1 1 1 0
1 1 0 0 0 1 0 1
0 0 1 1 0 0 1
1 1 1 0 1 0 0 1
0 1 1 0 1 0 1 1
1 0 0
1 0 1 0 1 1
1 1 0 0 0 1 0
1 1 1 0
0 0

```

Modern Data Architecture for Analytics and AI



Customer Leadership—Global and Across All Key Verticals

AUTO / TRANSPORTATION



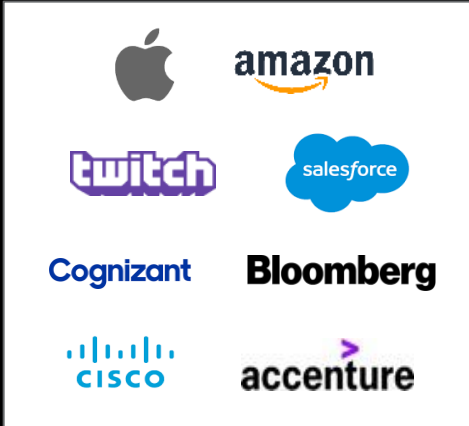
BANKING / INSURANCE



RETAIL



TECHNOLOGY / SERVICES



10,000+
Customers

9

of Fortune 10

GOVT. / PUBLIC SECTOR / EDU.



HEALTHCARE / PHARMA



MANUFACTURING



ENERGY / UTILITIES



85

of Fortune 100

Analysts Recognise Informatica Leadership

Best Data Management Products



Gartner® Magic Quadrant™

Data Integration Tools, 2022
Data Quality Solutions, 2022
Overall Rating—Positive, 2022
MDM Solutions, 2021

FORRESTER® The Forrester Wave™

Enterprise Data Catalogs For DataOps, Q2 2022
Enterprise Data Fabric, Q2 2022



MarketScape for Data Catalogs, 2022



iPaaS Technology Value Matrix, 2022



GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally, and MAGIC QUADRANT is a registered trademark of Gartner, Inc. and/or its affiliates and are used herein with permission. All rights reserved. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Want to Continue the Conversation?

Let's connect and discuss a broad range of topics, including:

- Creating value from your data investments
- How to grow data maturity in your organisation
- Aligning your data strategy with business priorities
- Overall data strategy and business adoption



CDO Insights 2024:
Charting a Course
to AI Readiness



Informatica
Experience
Lounge

Thank You!



Jonathon Bowring
Field Engineer at Informatica

