

Modern Data Management in the era of AI

Jonathon Bowring

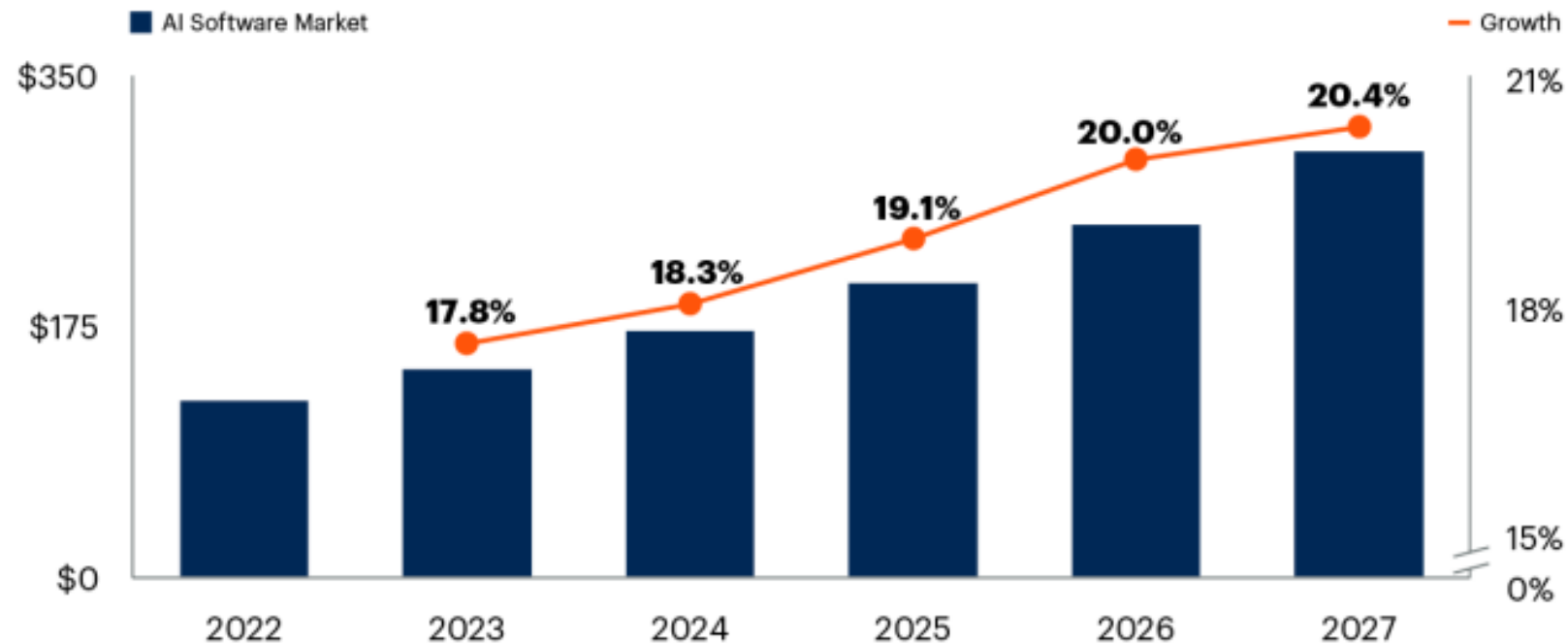




AI Market

According to Gartner

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



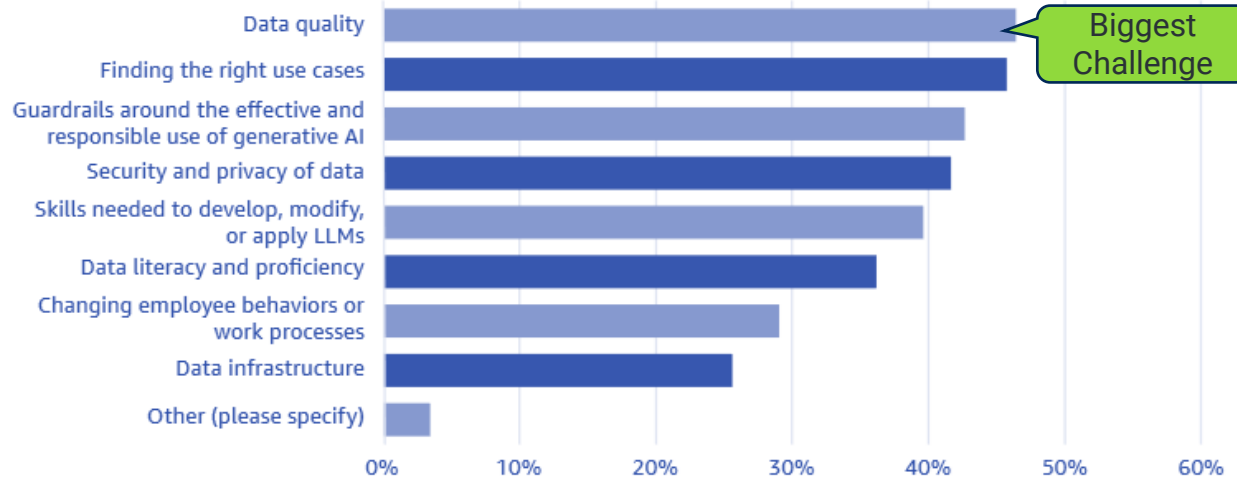
- By 2027, spending on AI software will grow to **\$297.9 billion** with a CAGR of **19.1%**.
- Over the next five years, the market growth will accelerate from 17.8% to reach **20.4%** in 2027.
- Generative AI software spend will rise from 8% of AI software in 2023 to **35%** by 2027.

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

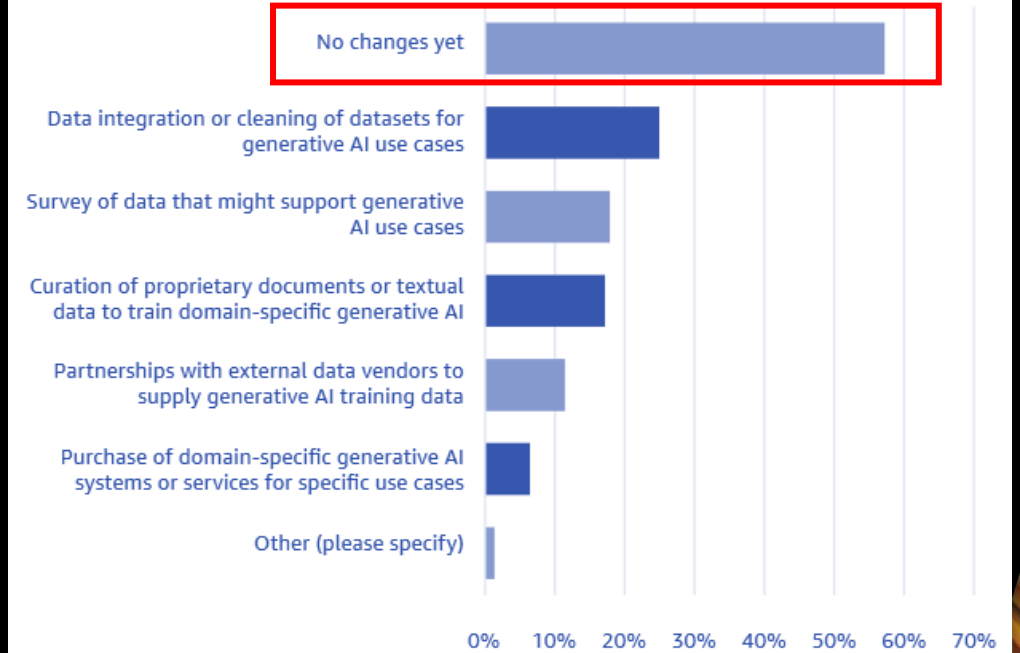
Generative AI Opportunities

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

What is the biggest challenge for your organization in realizing the potential of generative AI?



How has your data environment changed to support or enable generative AI?



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

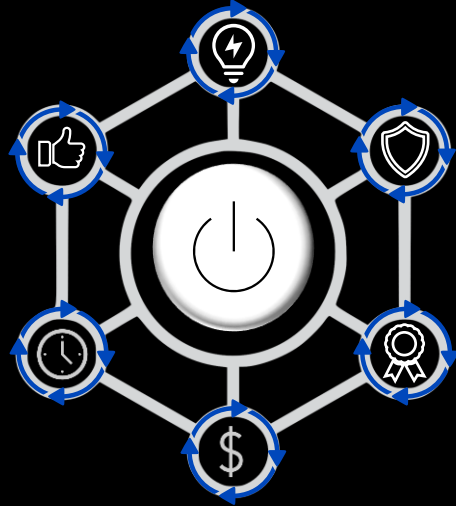
SOURCE: "CDO Agenda 2024: Navigating Data and Generative AI Frontiers." (2023), [AWSstatic.com](https://awsstatic.com)

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

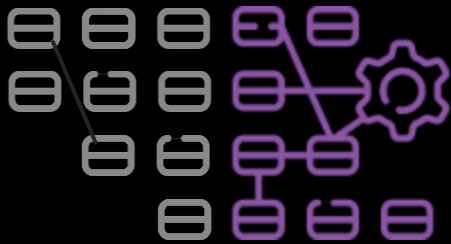
Key points to consider



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), [WPROMOTE.com](#)

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 uncleaned 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

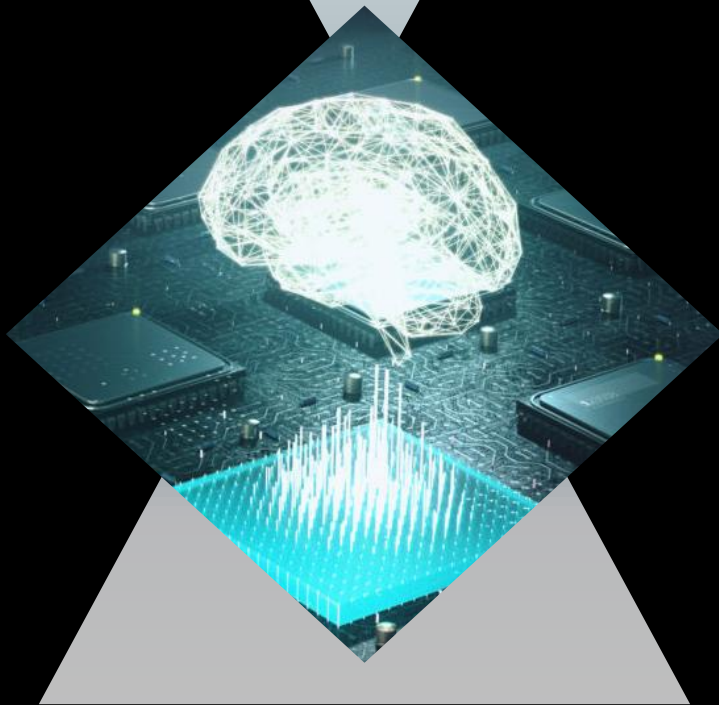
- Used generative AI to find prior cases to support a case
- At least 6 of the AI findings did NOT exist
- As of June 2023, Schwartz was facing possible sanctions by the court.

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

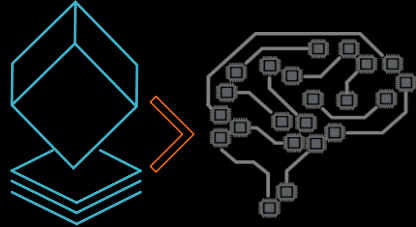
Decentralised Data Products Dilemma

- 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

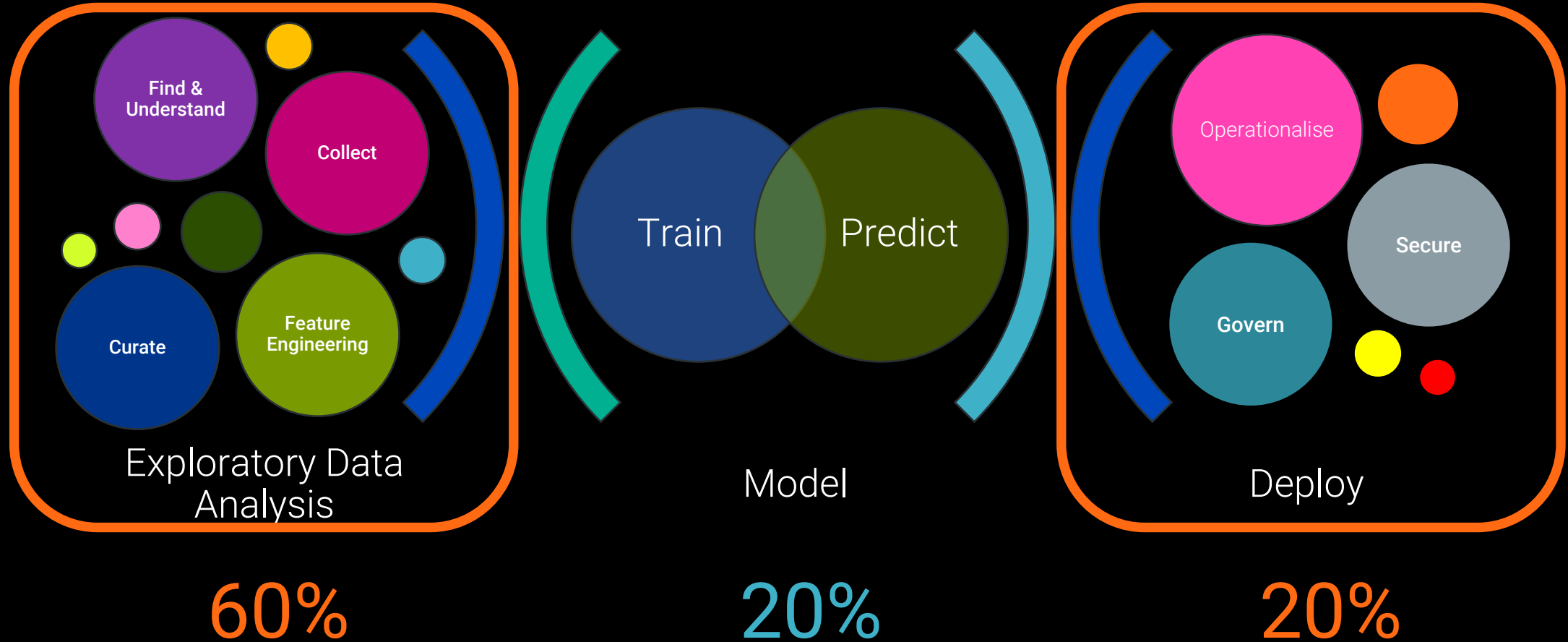
First AI Theme

Data Management for AI

Where data
& AI come to 

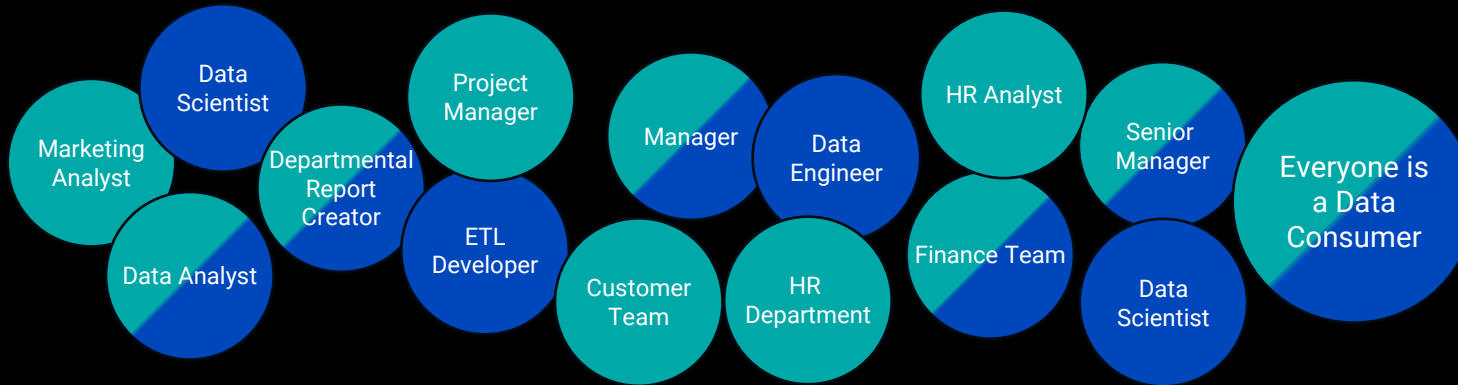
The Data Science Workflow

Streamlined Data Management is the Data Science Supercharger



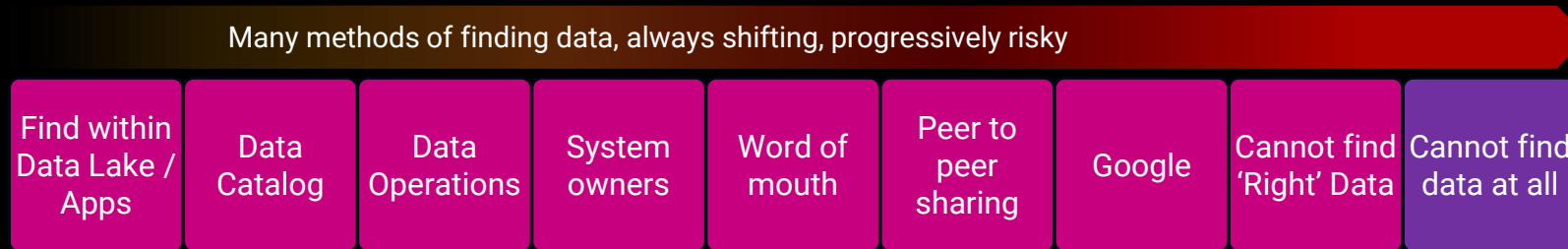
Discovering and Understanding Data

Data Consumers



- Technical skillset
- Non-technical skillset
- Hybrid skillset

Finding Data

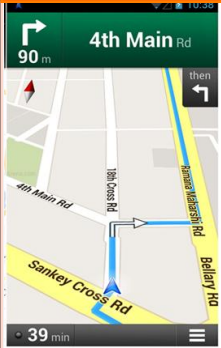


Demand and usage **tracking is patchy**, as access requests are broadly distributed across the data estate

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

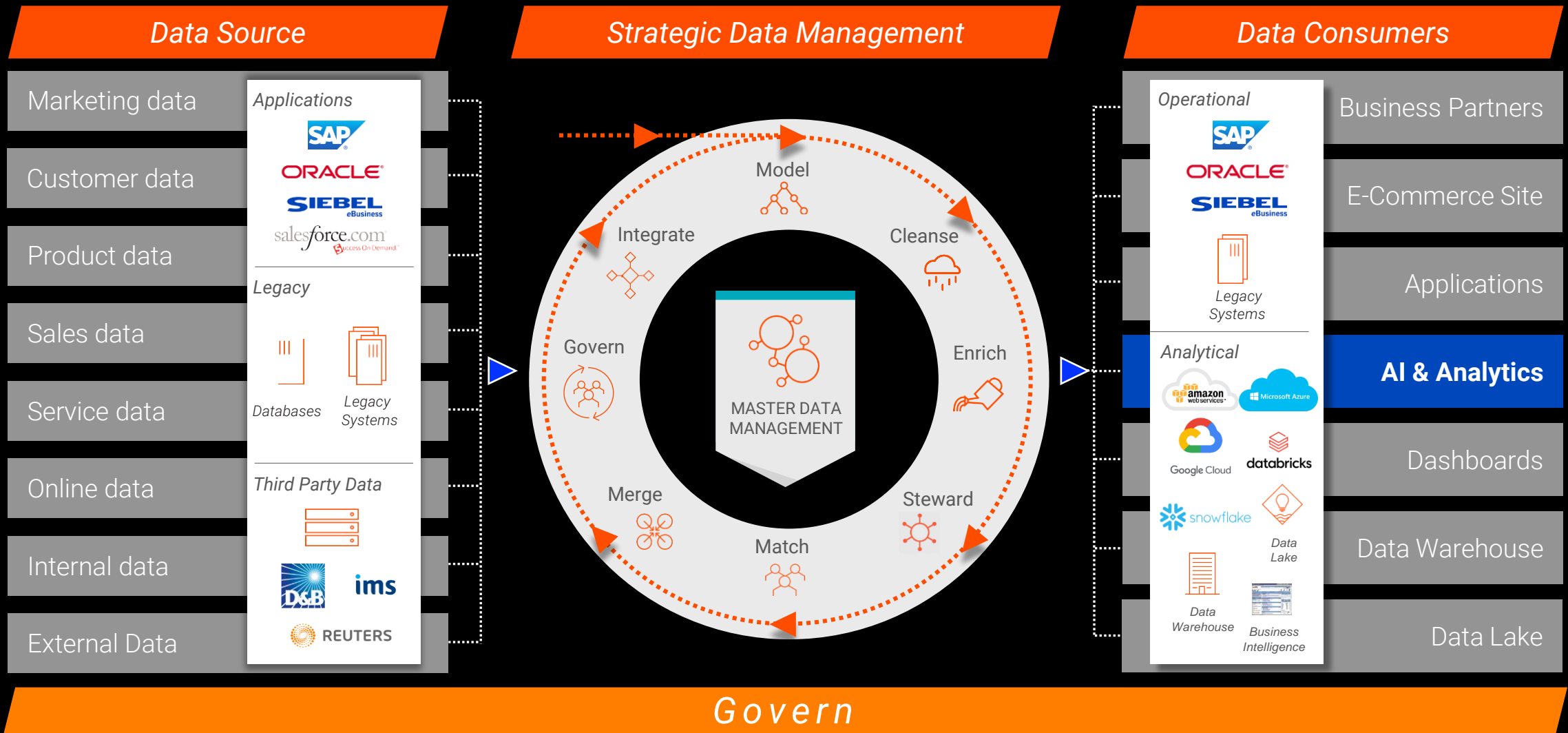


Data Architect Business User Developer

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



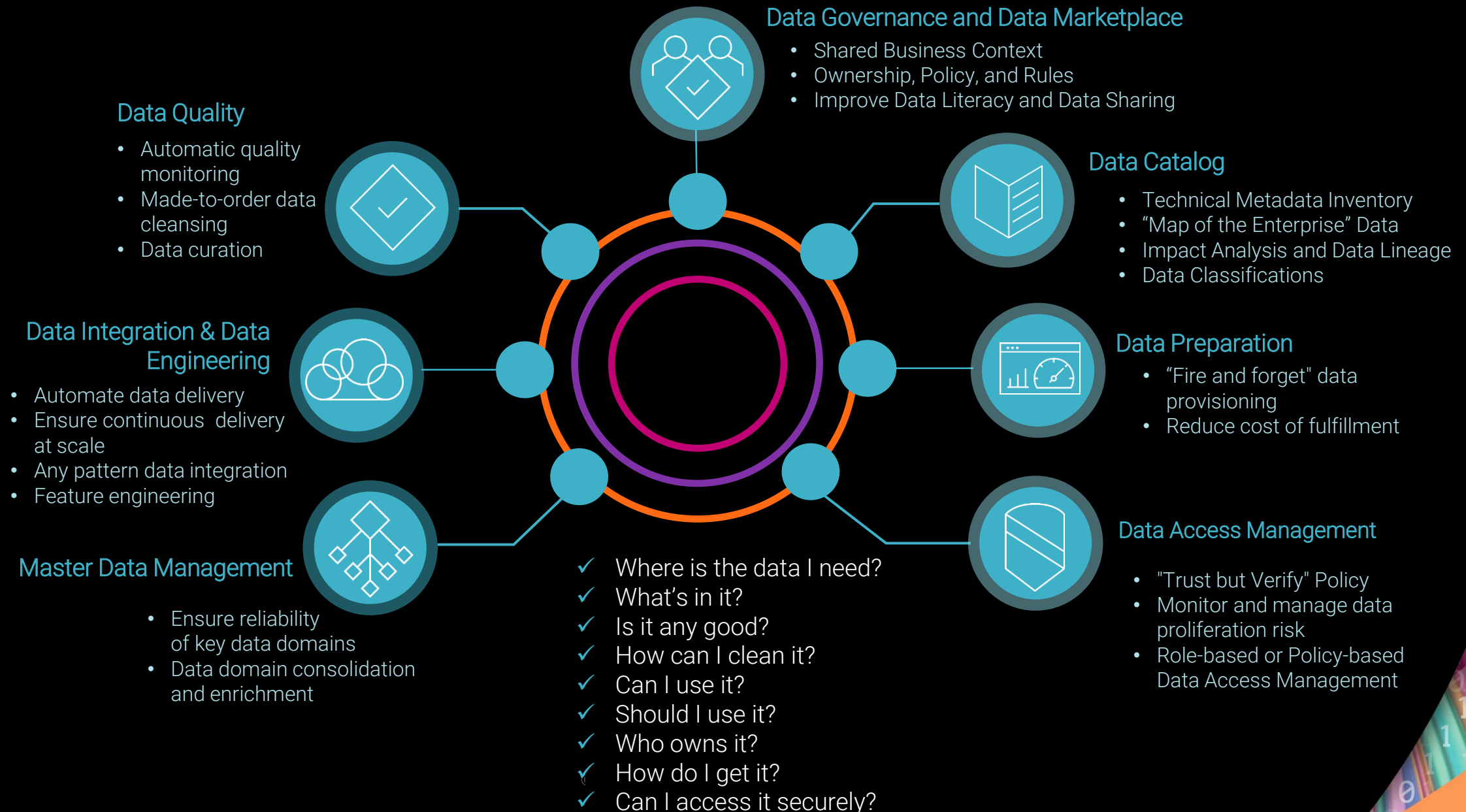
Accelerate time to value for data-driven outcomes, connecting data producers and consumers

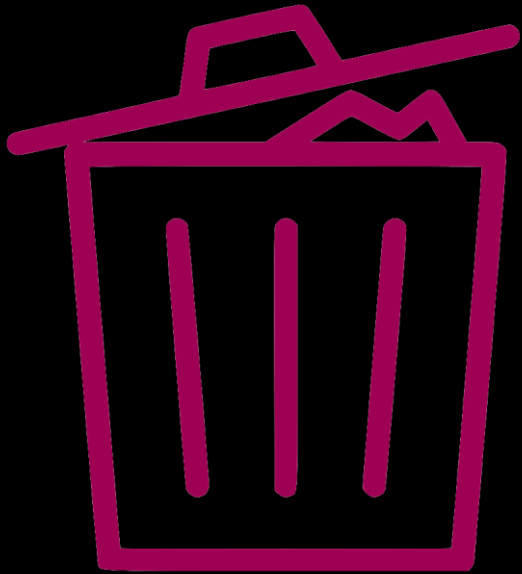
Observability



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 

Challenges that AI paired with Data Management will address

Scale & Agility



Policy & Trust



Skills & Literacy



AI Generated Classifications



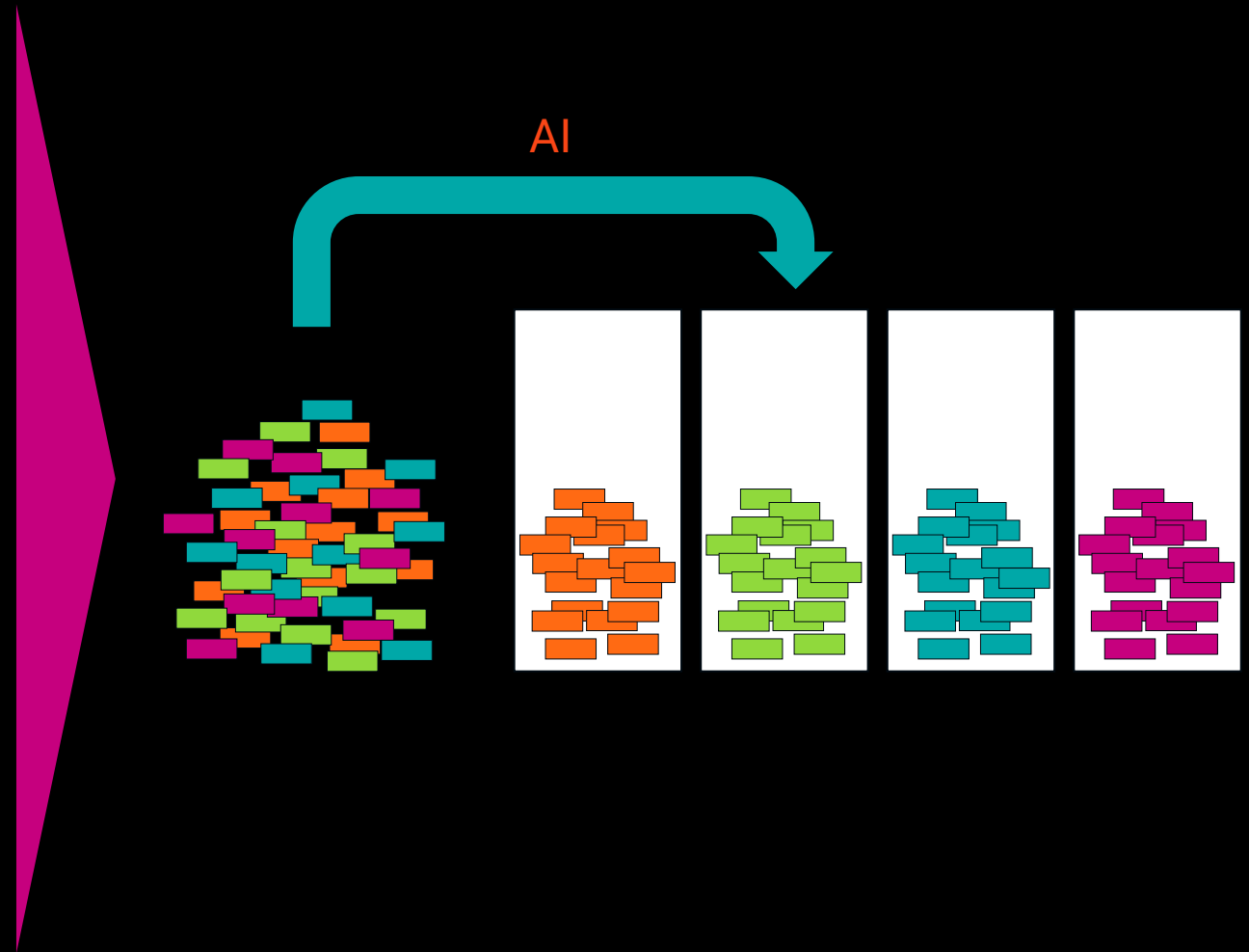
Key Highlights

- 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

The image displays two screenshots of the Informatica QLABS interface. The top screenshot shows a visual lineage diagram with nodes for 'Order Entry', 'Data Set 1', 'Data Set 2', and 'Sales Analysis'. A modal window is open over the diagram, showing a connection from 'Address' to 'Customer Address' with a status of 'Pending Review' and a transformation of 'Join'. The bottom screenshot shows a table of inferred connections with a filter overlay. The table has columns for 'CONNECTION ORIGIN', 'CONNECTION TARGET', and 'CONNECTION TYPE'. The filter overlay includes options for 'Lineage Type' (Inferred Lineage, Derived Lineage), 'Data Sources' (Order entry, Sales Analysis, Sales Report), 'Recommendation Status' (Pending review, Accepted, Declined), and a 'Confidence score' slider set to 40.

CONNECTION ORIGIN	CONNECTION TARGET	CONNECTION TYPE	STATUS
<input type="checkbox"/>			
<input checked="" type="checkbox"/>	Order Entry → Sales Analysis	Inferred	Pending Review
<input checked="" type="checkbox"/>	Data Set 1 → Data Set 2	Derived (2), Inferred (6)	Pending Review
<input type="checkbox"/>	Order number → Order Details	Derived	-
<input type="checkbox"/>	Customer Name → Customer Info	Derived	-
<input checked="" type="checkbox"/>	Customer Name → Transaction ID	Inferred	Accepted
<input checked="" type="checkbox"/>	Customer Name → Sales Region	Inferred	Accepted
<input checked="" type="checkbox"/>	Customer Name → Processing Status	Inferred	Accepted
<input checked="" type="checkbox"/>	Address → Address Pin Code	Inferred	Accepted

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_FIRSTNAME
DATA ELEMENT CLASSIFICATION

Classification Rule

NAME LIKE 'FIRSTNAME'

CLAIRE Data Classification

Associations

6 230

CDAM Access Control Policy

Rule Names: Ofuscate_PersonName

Conditions: Trigger this rule if:
→ User Group is any of [Data_Consumers]

Transformations: 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	78446	United Kingdom	Mechanical Engineer
Chaim	Ford	61428	Ireland	Compliance Officer
Alexandra	Sloan	49395	Ireland	Compliance Officer
Oliver	Combs	73320	United Kingdom	Compliance Officer

User without Privileges

FIRST_NAME	SECOND_NAME	EMPLOYEE_NUMBER	COUNTRY	JOB_TITLE
Osksbladshifv	*****	936042276038	Germany	Court Reporter
Xwlasxrcfcjip	*****	444607627894	United Kingdom	Marketing Research Analyst
Pkxqxbrtcjaj	*****	5484484031640	Germany	Project Manager
Mduoqilcgest	*****	6872587489830	United States	Copywriter
Xqmuozqjenn	*****	2201799735622	Germany	Cloud Solutions Architect
Lmonphcbeepzc	*****	7088887775853	Germany	Quality Control Analyst
Zrxqmyvaadga	*****	3309725977403	Germany	Marketing Research Analyst
Ypmxmpukmbad	*****	332215463410	France	Web Developer
Bpwjiqcyyi	*****	569455264405	France	Cybersecurity Analyst
flwvzmtabonw	*****	933206451159	United Kingdom	Compliance Officer
Nuzzrodiezaffi	*****	8854298161402	Germany	Media Buyer
Mpiczwpjxmi	*****	6649959732424	United Kingdom	Civil Engineer

AI Generated Data Observability



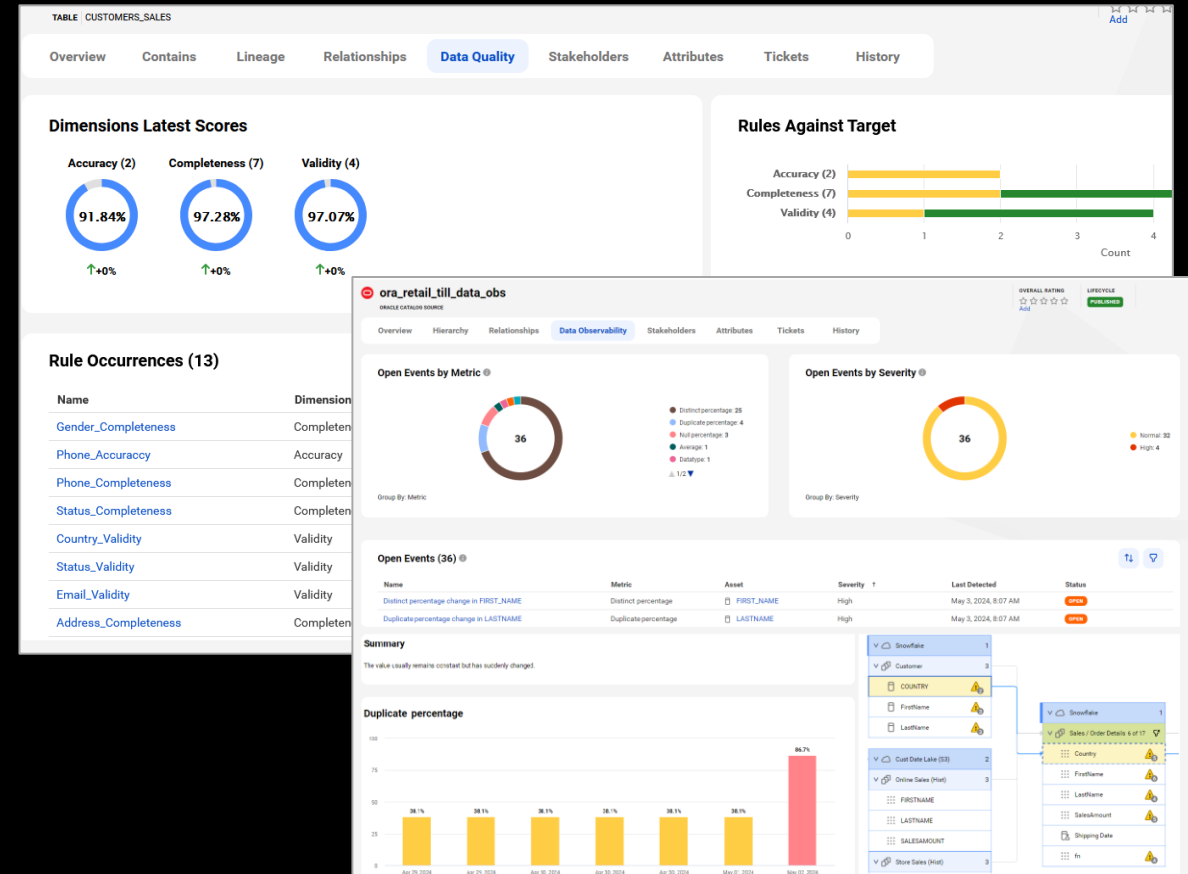
Key Highlights

- 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. It includes a form for 'Intelligent Structure Details' with fields for Name (Acme_Log), Location (Default), and Description (Claire on BDM demo). Below the form is a 'Sample File' section with a file named 'Acme_log.txt' and a 'Discover Structure' button. The interface also shows a 'Visual Model' tab with a tree diagram of the data structure, including nodes for 'table', 'element', 'IP', 'value', 'token', 'statusCode', 'number', 'datetime', 'HTTPMethod', 'URI', and 'protocol'. To the right, there is a 'Table' view showing a data table with columns A through F, including 'IP', 'value', 'token', 'datetime', 'HTTPMethod', and 'URI'. The table contains multiple rows of data, with some cells highlighted in green.

AI Self Integrating Systems



Key Highlights

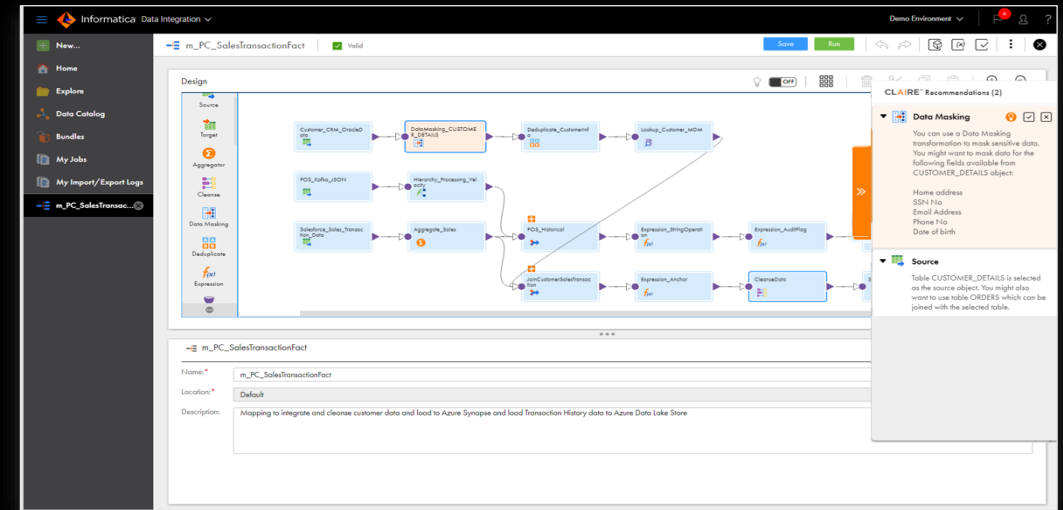
- **Fully automated data pipeline** generation
- Support for unstructured, semi-structured and hierarchical data sources
- Automated creation data masking, cleansing and standardization rules
- Support for join, union, normalization and denormalization of data sources



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**

Profile_PERSON_... Profile run 16 of 16 | 12 of 13 Columns | 11 of 11 Rules | 142 Rows (All rows) | Nov 24, 2023, 6:27:08 PM

Results Definition Rules Metrics Schedule Insights

View: Columns and Rules with: All Statistics

Columns	Value Distribution	% Null	# Null	% Distinct	# Distinct	% Non-dist...	# Non-dist...	# Patterns	% of Top
CUST_ID		0%	0	95.77%	136	4.23%	6	2	98.59%
COMPANY		1.41%	2	67.61%	96	30.98%	44	4	81.69%
FULLNAME		0%	0	97.89%	139	2.11%	3	5	78.87%
TITLE		9.86%	14	36.62%	52	53.52%	76	7	64.79%
ADDR_ID		0%	0	82.39%	117	17.61%	25	3	80.99%
EMAILADDRESSID		0%	0	82.39%	117	17.61%	25	3	80.99%
BUSINESSENTITYID		0%	0	82.39%	117	17.61%	25	3	80.99%
COUNTRY		6.34%	9	11.97%	17	81.69%	116	7	36.62%
EMAIL_OPT_IN		5.63%	8	0.7%	1	93.67%	133	2	94.37%
MOBILE_OPT_IN		0%	0	1.41%	2	98.59%	140	1	100%
GENDER		1.41%	2	2.82%	4	95.77%	136	4	50.7%
SSN		4.93%	7	93.66%	133	1.41%	2	4	84.51%

Details Rules

Trend

Type	Trend	% Change
Null		0%
Distinct		0%
Non-distinct		0%

Data Types (6)

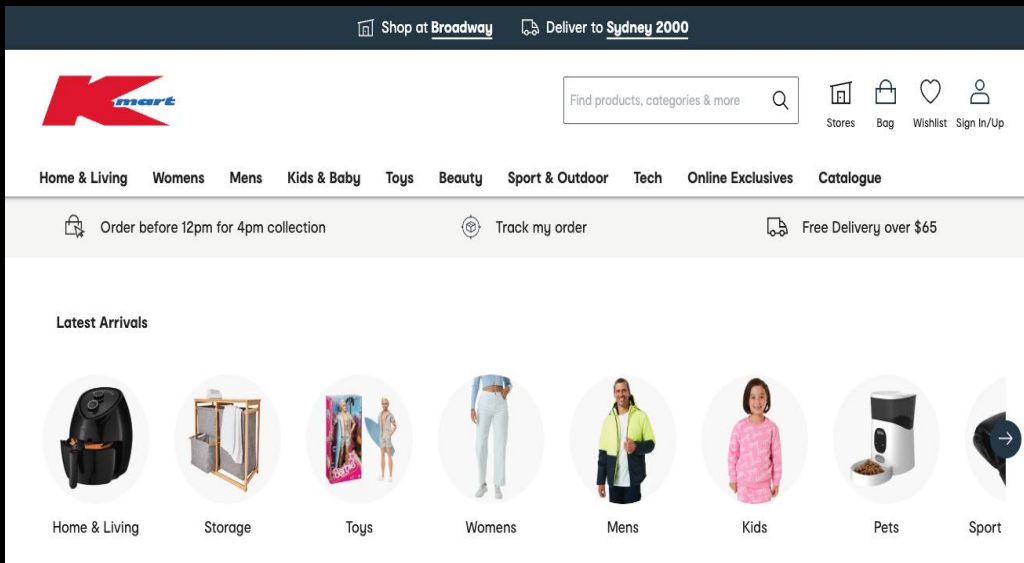
- varchar(10) Documented Data Type
- Decimal(6) 100%
- Integer(6) 100%
- String(6) 100%

Profile_PERSON_...

Results Definition Rules Metrics Schedule Insights

View: All Insight Types in: All Columns

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	Approved
<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	Approved
<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	Approved
<input type="checkbox"/> The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	EMAILADDRESSID	Approved
<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	Approved
<input type="checkbox"/> Numeric values found outside the 95% standard deviation range.	Low	Number Value Distribution	CUST_ID	Approved
<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	Approved
<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	Approved
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or valu...	High	Completeness Check	COUNTRY	Approved
<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	Approved
<input type="checkbox"/> Data appears incomplete. The column includes one or more null, blank, or empty values or valu...	Medium	Completeness Check	SSN	Approved

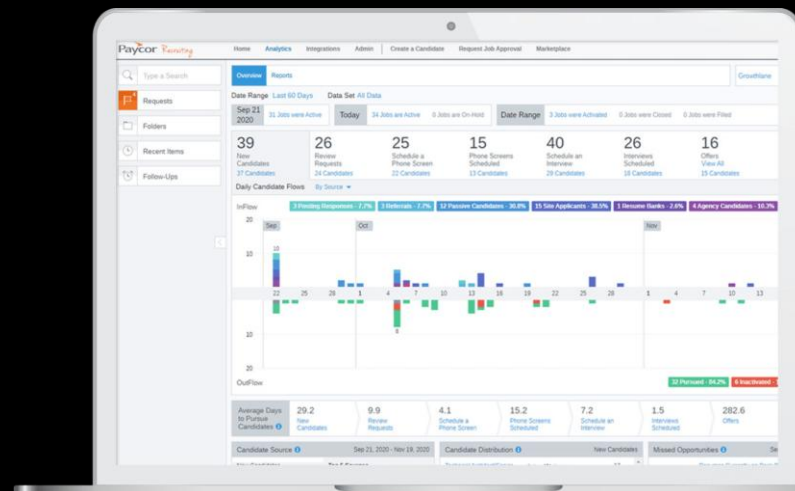


Latrobe University

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





Integration

Intelligent Structure 1

Intelligent Structure Details

Name: Acme_Log

Location: Default

Description: Claire on BDM demo

Sample File: Acme_log.txt

```

0:0:0:0:0:1 - 96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar [18/Sep/2014:14:54:24 +0300] GET /dx-console/ HTTP/1.1 302 -
0:0:0:0:0:1 - 96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar [18/Sep/2014:14:54:25 +0300] GET /dx-console/com.informatica.b2b.de.Main/main.jsp HTTP/1.1 302 -
0:0:0:0:0:1 - 96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar [18/Sep/2014:14:54:25 +0300] GET /dx-console/login.jsp HTTP/1.1 200 4472
0:0:0:0:0:1 - 96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar [18/Sep/2014:14:55:47 +0300] POST /dx-console/_spring_security_check HTTP/1.1 302 -
0:0:0:0:0:1 - 96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar [18/Sep/2014:14:55:47 +0300] GET /dx-console/com.informatica.b2b.de.Main/main.jsp HTTP/1.1 300 147344
  
```

Visual Model

```

graph TD
    table --> element
    element --> IP
    element --> value
    element --> token
    element --> element2
    element --> element3
    element --> statusCode
    element --> number
    element2 --> HTTPMethod
    element2 --> URI
    element3 --> HTTPMethod
    element3 --> URI
    element3 --> protocol
  
```

A	B	C	D	E	F
IP	value	token	datetime	HTTPMethod	URI
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/login.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		POST	/dx-console/_spring_security_check
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp
0:0:0:0:0:1	-	96E8FA61FD3802CA0349ACAF0C5152EC.d8fomcar		GET	/dx-console/com.informatica.b2b.de.Main/main.jsp

- 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

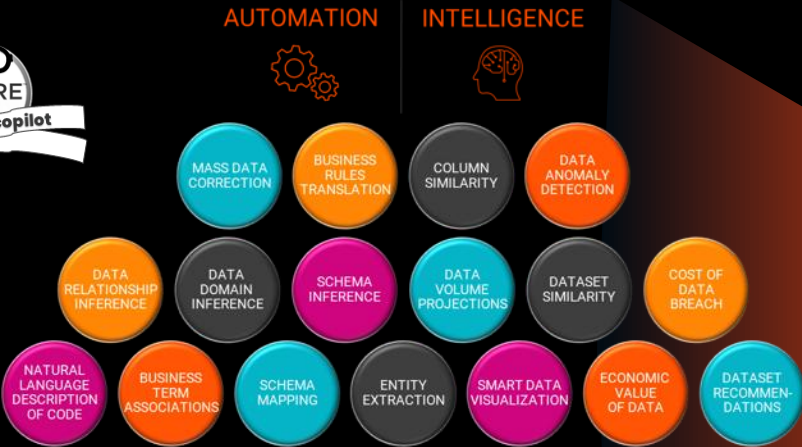
Third AI Theme

Gen AI Data Management

CLAIRE[®]: AI-Powered Data Management

CLAIRE AI copilot

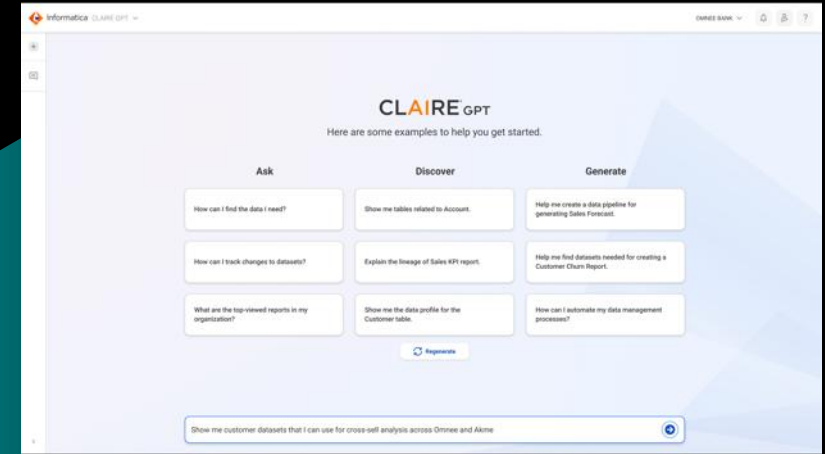
IN PRODUCT AI-POWERED METADATA INTELLIGENCE AND AUTOMATION



CLAIRE[®]
UNIFIED METADATA INTELLIGENCE

CLAIRE GPT

NATURAL LANGUAGE INTERFACE TO DATA



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

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-BASED DISCOVERY	NL-BASED EXPLORATION	NL-BASED TESTING AND DOCUMENTATION
CROSS PRODUCT EXP.	AI COPILOT	IN-CONTEXT DATA INTELLIGENCE
		NL TO GOVERNANCE
		AUTOMATED FINOPS
		OPEN, POLYGLOT

CLAIRE[®] GPT

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

Discover

Show me the assets controlled by

List critical data elements across domains

Show me a list of data sources available in the catalog

Explore

Show data preview of **PRODUCT**

Show data preview of **CUSTOMER_TRANSACTION**

Create a mapping for **PRODUCT**

Ask

How to delete any corrupt or non-responsive asset using IICS repository API in CAI?

Describe MDM SaaS architecture

What are the rule specifications that I can use to configure data quality in Business 360 console?

Ask CLAIRE... 

Show me datasets related to customer Last modified: 10/17/2024, 06:13 PM



Show me datasets related to customer



Searching for datasets related to customer

Found 6 of the available 10 records.

Here are the datasets related to **customer**. They are spread across different resources and databases.

1. Tables under **Snowflake_RetailScan** Snowflake resource in **PLATFORMQA** database and **SNOW_NEXGEN_RETAIL** schema:
 - **CUSTOMER**: This table contains customer data.
 - **CUSTOMER_TRANSACTION**
2. Datasets under **customer_segmentation** Microsoft Power BI resource in **Customer_Demo** workspace:
 - **customer_segmentation** (Dataset)
 - **customer_segmentation** (Report)
3. Table under **customer_staging** Oracle resource in **RND12CR2** database and **CUSTOMER_RETAIL** schema:
 - **CUSTOMER_PORTFOLIO**
4. FlatFile under **customer_retail** Amazon S3 resource in **hawkqe** bucket and **customer_reports** folder:
 - **customer_demographics.csv**

Sorry, I don't understand the question. Can you please rephrase the question and retry?



Ask CLAIRE...



Show me datasets related to customer Last modified: 10/17/2024, 06:14 PM

- Relational Owner: SNOW_NEXGEN_RETAIL
- Last Profiled State: Complete

CUSTOMER | TABLE

Customer

Snowflake_RetailScan / PLATFORMQA / SNOW_NEXGEN_RETAIL Data Elements: 11 Stakeholders: --

Last Updated: Jun 6, 2024

Data Quality Scores:



Data Elements

1	CUSTOMERID	6	CITY
2	FIRSTNAME	7	STATE
3	ADDRESS	8	COUNTRY
4	PHONE	9	GENDER
5	LASTNAME	10	EMAIL

Showing 10 of 11



Ask CLAIRE...



Show me datasets related to customer Last modified: 10/17/2024, 06:17 PM

- **Data Type:** Fixed Length String
- **Analysis of 1069 Entries:** Moderate duplicate rate (4.49%), high distinctiveness (95.51%).

LASTNAME

- **Data Type:** String
- **Analysis of 1069 Entries:** Low duplicate rate (2.81%), very high distinctiveness (97.20%).

CITY

- **Data Type:** String
- **Analysis of 1069 Entries:** High duplicate rate (71.28%), very low distinctiveness (28.72%).

STATE

- **Data Type:** Fixed Length String
- **Analysis of 1069 Entries:** Nearly all values are duplicates (95.32%) with no nulls or blanks.

COUNTRY

- **Data Type:** Fixed Length String
- **Analysis of 1069 Entries:** Almost all values are duplicates (99.90%), indicating lack of diversity.

GENDER

- **Data Type:** Fixed Length String
- **Analysis of 1069 Entries:** Extremely high duplicate rate (99.81%), suggesting limited gender diversity.

EMAIL

- **Data Type:** String
- **Analysis of 1069 Entries:** Very low duplicate rate (0.09%), almost complete uniqueness (99.91%).



[Open in Data Governance and Catalog](#)

Show me an overview of **CUSTOMER**

Show me data preview for **CUSTOMER**

Show me glossaries associated to **CUSTOMER**

Ask CLAIRE...



Show me datasets related to customer Last modified: 10/17/2024, 06:21 PM

Data Type: String

- **Analysis of 1069 Entries:** Very low duplicate rate (0.09%), almost complete uniqueness (99.91%).



Open in Data Governance and Catalog



Show me the data quality



Showing the data quality of CUSTOMER

The data quality for **CUSTOMER** is as follows:

- **Validity:** 100.0%
 - Indicates that there are no invalid data entries in this dataset.
- **Completeness:** 100.0%
 - Shows that all required fields have been filled out, ensuring full coverage of data.
- **Overall Data Quality:** 100.0%
 - Reflects a high level of confidence in the reliability and usefulness of the customer data. Improvements may still be possible but they would likely have minimal impact on the current high standard.

Data Quality Scores:



Open in Data Governance and Catalog

Ask CLAIRE...



Show me datasets related to customer Last modified: 10/17/2024, 06:25 PM



Showing the lineage of customer

- The lineage for **CUSTOMER** involves a single resource.
- Data flows into this resource from one source: **customer_retail**.
- Specifically, the data comes from the asset named customer_demographics.csv.
- There are no resources that the data flows out to (end sink).

Lineage Details

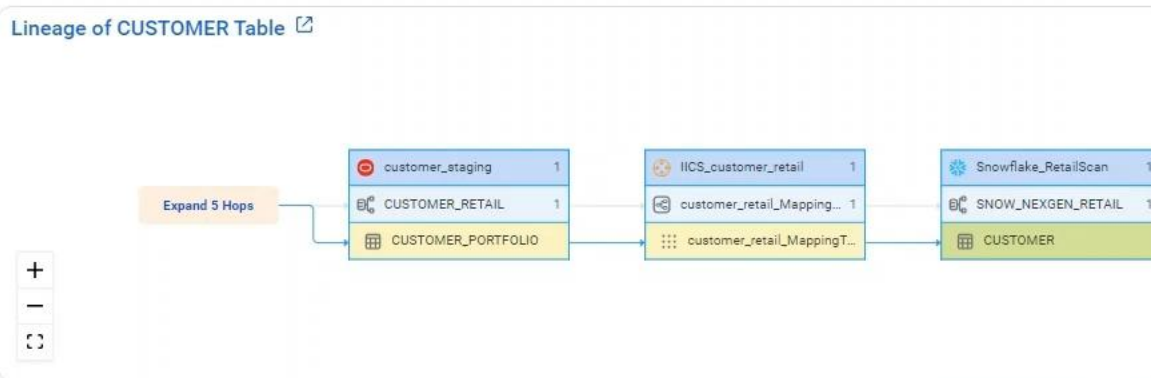
Input Resource

- **customer_retail:**
 - **Asset:** customer_demographics.csv

Output Resources

The data doesn't flow out to any target resource.

Lineage of CUSTOMER Table



Download

Show me data preview for **CUSTOMER** Find the data quality scores of **CUSTOMER** Find dataset with data lineage to **CUSTOMER**

Ask CLAIRE...

Show me datasets related to customer Last modified: 10/17/2024, 06:27 PM

Placeholder for a previous dataset view, including thumbs up/down icons, a Download button, and a share icon.



Show me a data preview



Showing a data preview of customer

CUSTOMER | TABLE

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.

Party_ID	First_Name	Address	Phone_Num...	Last_Name	City
C927	Yale	92 Dovetail Terra...	202-143-1112	Von Salzberg	Washington
C943	Virginie	20141 Kingsford...	216-309-7318	Yankin	Cleveland
C888	Vonni	64 Debra Plaza	704-253-6655	Hayden	Charlotte
C984	Trumann	56579 Annamar...	619-969-1163	Shugg	San Diego
C972	Tiena	77086 Warbler C...	727-325-7662	Mourgue	Clearwater
C928	Timothee	04 Rutledge Jun...	209-730-5478	Rickards	Modesto
C887	Tiff	1 Bartelt Trail	405-762-1019	McCoid	Edmond
C886	Tisha	79 John Wall Lane	501-108-8575	Spat	Little Rock
C936	Zeke	831 Wayridge P...	786-770-2333	Reap	Miami
C911	Zacharias	3538 Hallows Pa...	941-361-3279	Mowsdell	Sarasota

Viewing sample data: 100 rows and 1 data types

Interaction bar for the data preview, including thumbs up/down icons, a Show Code button, a Download button, and a share icon.

Combine CUSTOMER with ORDERS Create a mapping for the above

Ask CLAIRE...

Show me datasets related to customer Last modified: 10/17/2024, 06:48 PM

CUSTOMER | TABLE

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 Party_ID	A First_Name	A Address	A Phone_Num...	A Last_Name	A City
C927	Yale	92 Dovetail Terra...	202-143-1112	Von Salzberg	Washington
C943	Virginie	20141 Kingsford...	216-309-7318	Yankin	Cleveland
C888	Vonni	64 Debra Plaza	704-253-6655	Hayden	Charlotte
C984	Trumann	56579 Annamar...	619-969-1163	Shugg	San Diego
C972	Tiena	77086 Warbler C...	727-325-7662	Mourgue	Clearwater
C928	Timothee	04 Rutledge Jun...	209-730-5478	Rickards	Modesto
C887	Tiff	1 Bartelt Trail	405-762-1019	McCoid	Edmond
C886	Tisha	79 John Wall Lane	501-108-8575	Spat	Little Rock
C936	Zeke	831 Wayridge P...	786-770-2333	Reap	Miami
C911	Zacharias	3538 Hallows Pa...	941-361-3279	Mowsdell	Sarasota

Viewing sample data: 100 rows and 1 data types



Show Code

Download



NAN

how many customers do we have?



Finding out the number of customers in CUSTOMER

1069



Show Code

Combine CUSTOMER with ORDERS

Create a mapping for the above

Ask CLAIRE...



Show me datasets related to customer Last modified: 10/17/2024, 06:50 PM

1069



Show Code



How many per state?



Figuring out how many customers there are per state in CUSTOMER

CUSTOMER | TABLE

Snowflake_RetailScan / PLATFORMQA / SNOW_NEXGEN_RETAIL Columns: 2

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

# COUNT(*)	A STATE
35	DC
29	OH
22	NC
130	CA
93	FL
18	OK
10	AR
3	HI
30	LA
14	KY
31	GA

Viewing sample data: 50 rows and 2 data types



Show Code

Download



Combine CUSTOMER with ORDERS

Create a mapping for the above

Ask CLAIRE... [Send icon]

Show me datasets related to customer

Last modified: 10/17/2024, 06:53 PM

18	OK
10	AR
3	HI
30	LA
14	KY
31	GA

Viewing sample data: 50 rows and 2 data types



Show Code

Download



Which top 5 states have the most sales



Figuring out the top 5 states with the most sales in CUSTOMER

CUSTOMER | TABLE

Snowflake_RetailScan / PLATFORMQA / SNOW_NEXGEN_RETAIL Columns: 2

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

A STATE	# SUM('ORDERS'.TOTALAMOUNT)
CA	61104.670
TX	47007.270
FL	32477.710
NY	18010.420
DC	13198.530

Viewing sample data: 5 rows and 2 data types



Show Code

Download



Combine CUSTOMER with ORDERS

Create a mapping for the above

Ask CLAIRE...



Show me datasets related to customer Last modified: 10/17/2024, 06:55 PM



Show Code

Download



NAN

Show customers that live in CA, NY, or FL with the number of orders, average order amount, and number of products purchased with the revenue generated in each product category



Showing customers that live in ca, ny, or fl with the number of orders, average order amount, and number of products purchased with the revenue generated in each product category using customer

CUSTOMER | TABLE

Snowflake_RetailScan / PLATFORMQA / SNOW_NEXGEN_RETAIL Columns: 7

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

A First_Name	A Last_Name	# Number_of_...	# Average_Ord...	# Number_of_...	# Revenue
Kevon	Schreurs	1	799.000000000	1	799.000
Kevon	Schreurs	1	8.990000000	1	8.990
Cally	Redmain	1	29.990000000	1	29.990
Leonor	Kellar	1	19.990000000	1	19.990
Edna	Fielden	1	14.990000000	1	14.990
Oberon	Crosswaite	1	9.990000000	1	9.990
Erv	Kettlesting	1	6.990000000	1	6.990
Laurie	McAllester	1	7.990000000	1	7.990
Wylie	Ranger	1	4.990000000	1	4.990
Cello	Reah	1	9.990000000	1	9.990

Viewing sample data: 100 rows and 3 data types



Show Code

Download



Combine CUSTOMER with ORDERS

Create a mapping for the above

Ask CLAIRE...



Show me datasets related to customer Last modified: 10/17/2024, 06:55 PM

Cellid	Keatn		9.990000000		9.990
--------	-------	--	-------------	--	-------

Viewing sample data: 100 rows and 3 data types

```
SQL
SELECT
  "CUSTOMER"."FIRSTNAME",
  "CUSTOMER"."LASTNAME",
  COUNT("ORDERS"."ORDERID") AS "Number_of_Orders",
  AVG("ORDERS"."TOTALAMOUNT") AS "Average_Order_Amount",
  COUNT(DISTINCT "ORDERS"."PRODUCTID") AS "Number_of_Products_Purchased",
  SUM("ORDERS"."TOTALAMOUNT") AS "Revenue",
  "PRODUCT"."CATEGORY"
FROM
  "PLATFORMQA"."SNOW_NEXGEN_RETAIL"."CUSTOMER" AS "CUSTOMER"
  INNER JOIN "PLATFORMQA"."SNOW_NEXGEN_RETAIL"."ORDERS" AS "ORDERS" ON "CUSTOMER"."CUSTOMERID" =
"ORDERS"."CUSTOMERID"
  INNER JOIN "PLATFORMQA"."SNOW_NEXGEN_RETAIL"."PRODUCT" AS "PRODUCT" ON "ORDERS"."PRODUCTID" =
"PRODUCT"."PRODUCTID"
WHERE
  LOWER("CUSTOMER"."STATE") LIKE '%ca%'
  OR LOWER("CUSTOMER"."STATE") LIKE '%ny%'
  OR LOWER("CUSTOMER"."STATE") LIKE '%fl%'
GROUP BY
  "CUSTOMER"."FIRSTNAME",
  "CUSTOMER"."LASTNAME",
  "PRODUCT"."CATEGORY"
LIMIT
  100
```



Hide Code Download ↗

Combine CUSTOMER with ORDERS Create a mapping for the above

Ask CLAIRE... ➤

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**

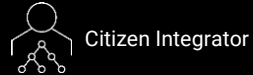
DATA CONSUMERS



ETL Developer



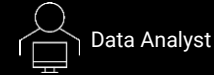
Data Engineer



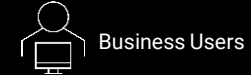
Citizen Integrator



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

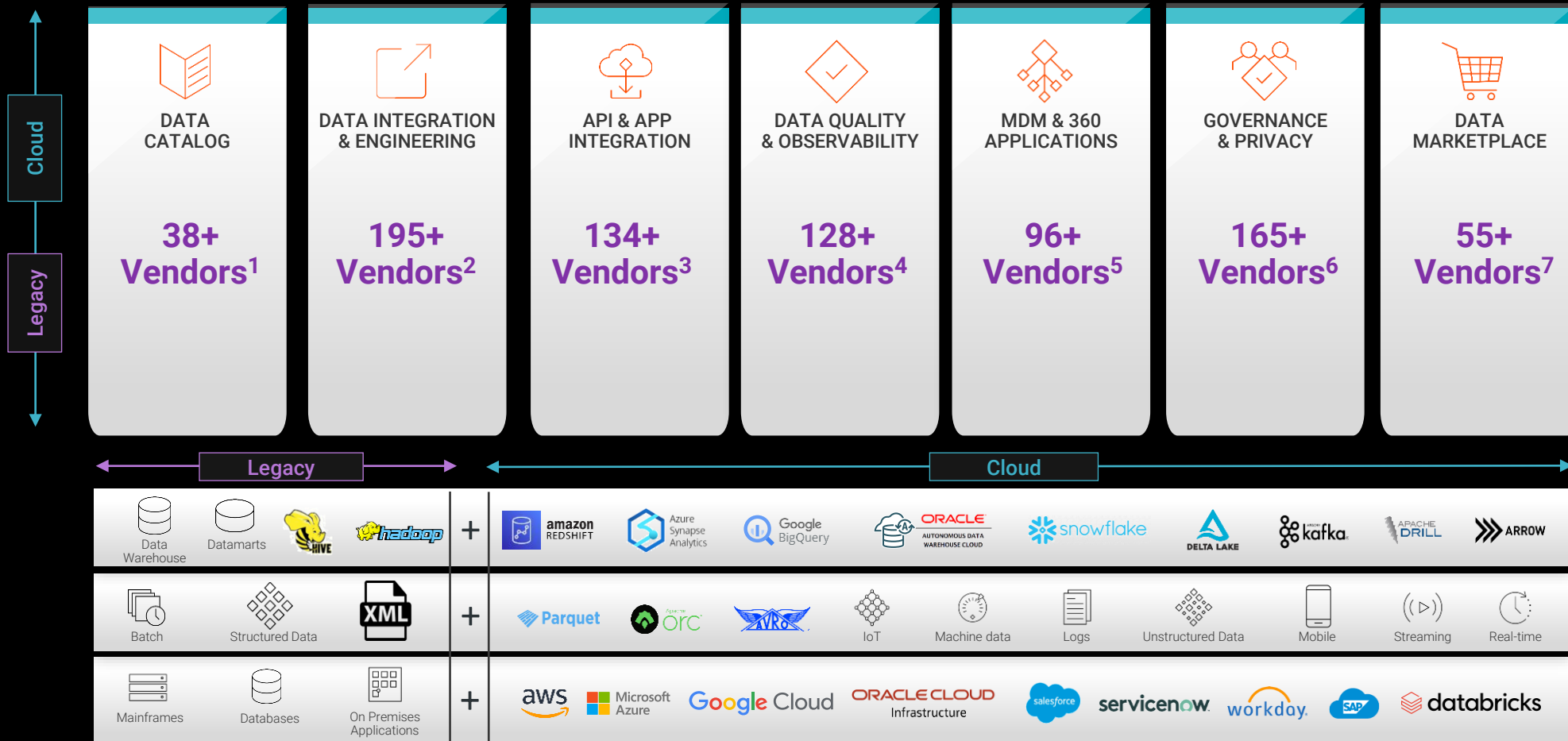


Logs

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](#)

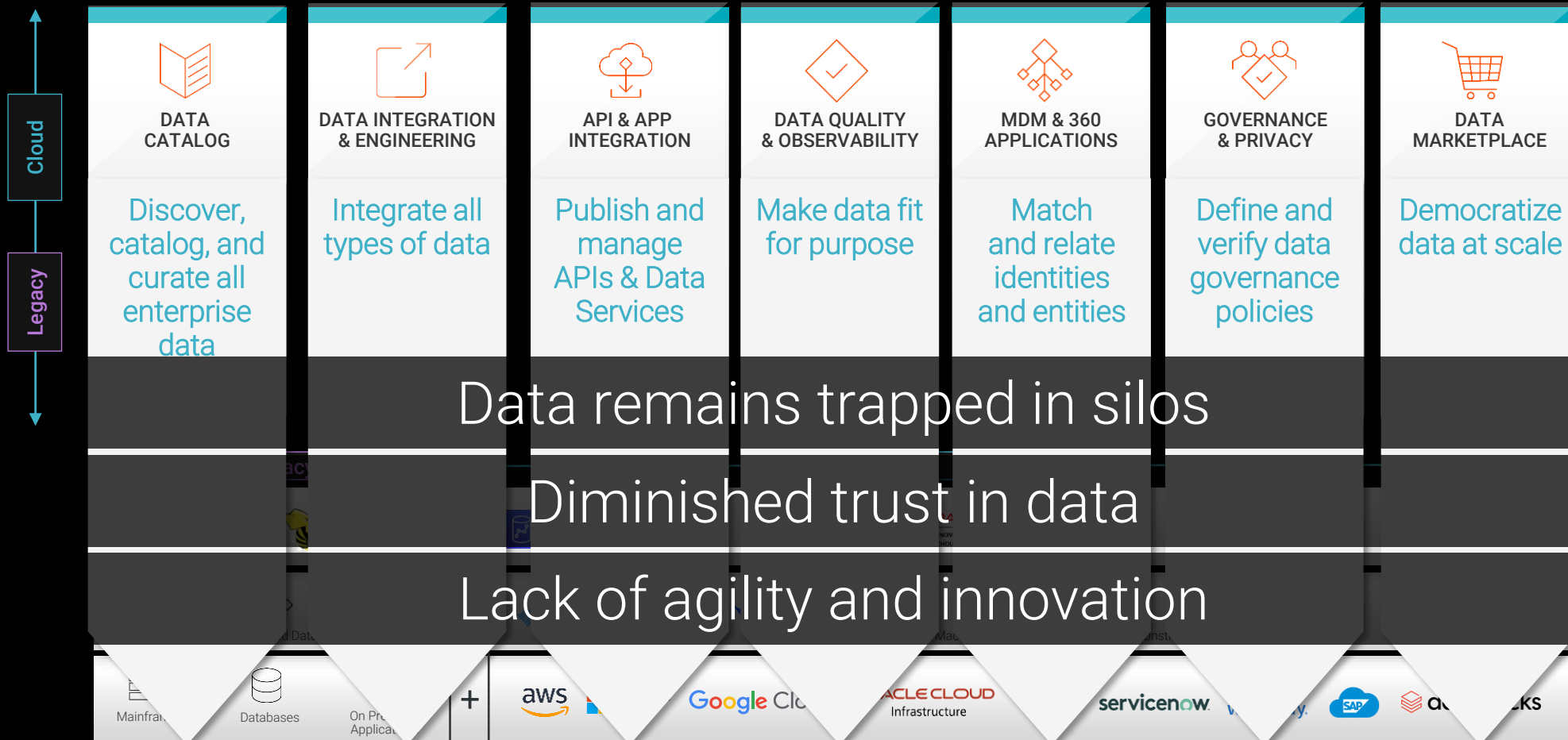
⁵Source: [Trust Radius, Master Data Management \(MDM\) Tools, May 2023](#)

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

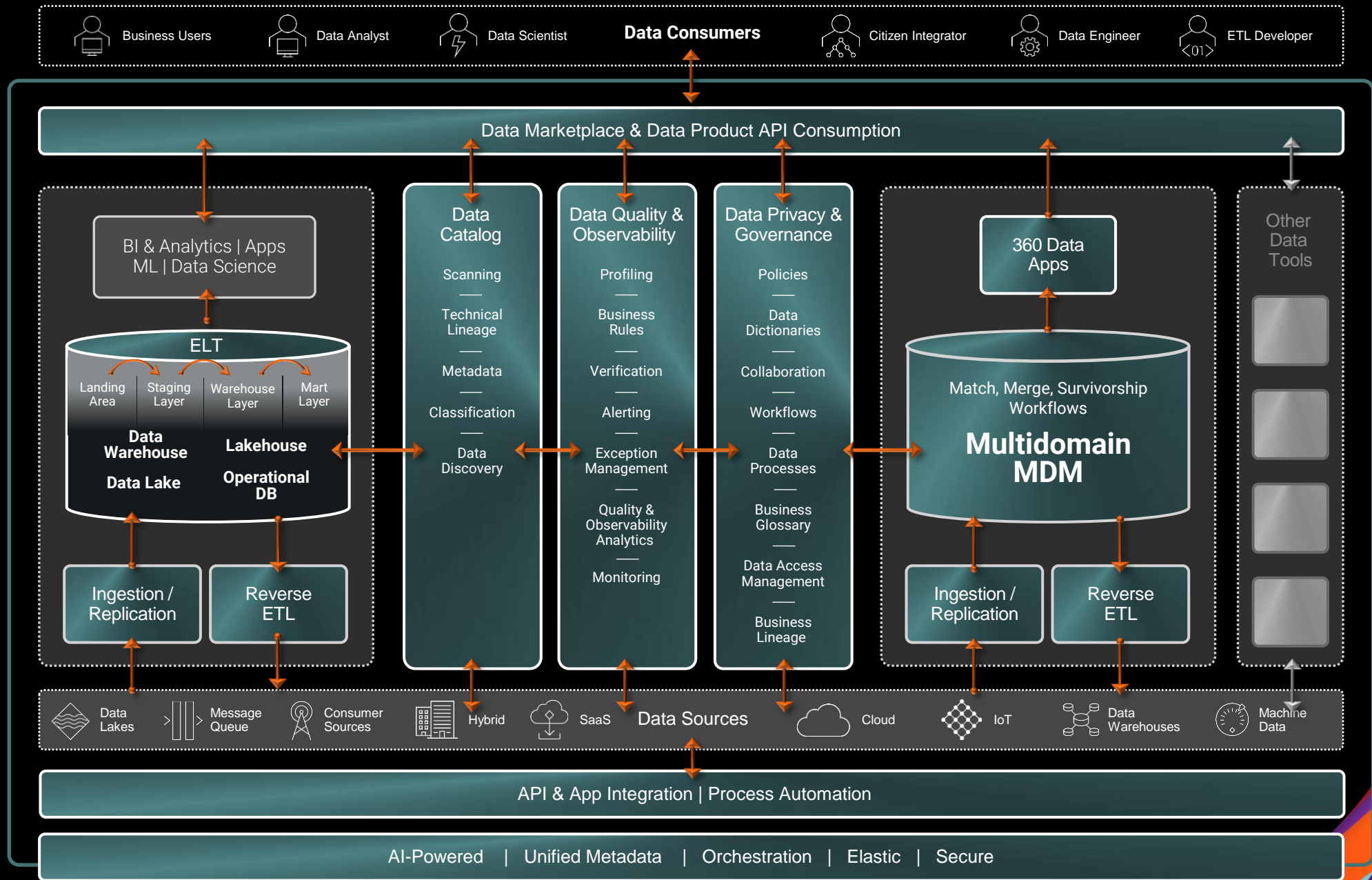


```

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

```

Modern Data Architecture for Analytics and AI



Customer Leadership—Global and Across All Key Verticals

AUTO / TRANSPORTATION



BANKING / INSURANCE



RETAIL



TECHNOLOGY / SERVICES



GOVT. / PUBLIC SECTOR / EDU.



HEALTHCARE / PHARMA



MANUFACTURING



ENERGY / UTILITIES



10,000+
Customers

9

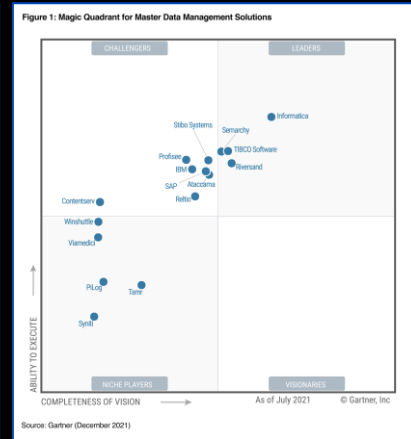
of Fortune 10

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



Jonathon Bowring
Field Engineer at Informatica



Thank You!