

Modern Data Management in the era of AI

Daniel Hein

Chief Architect - Asia, Pacific & Japan



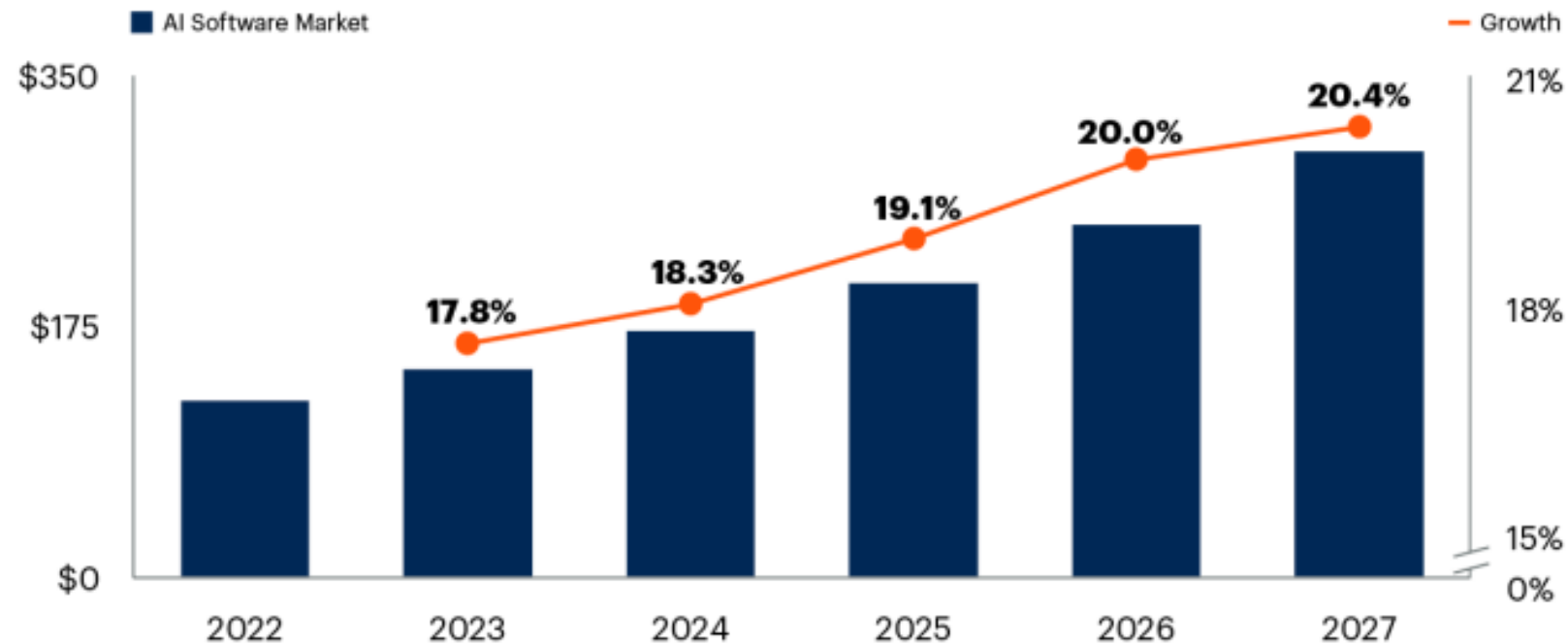
WE ARE NOW LIVING IN
A GENERATIVE AI WORLD

AI Market

According to Gartner

AI Software Forecast and Growth

In Millions of U.S. Dollars

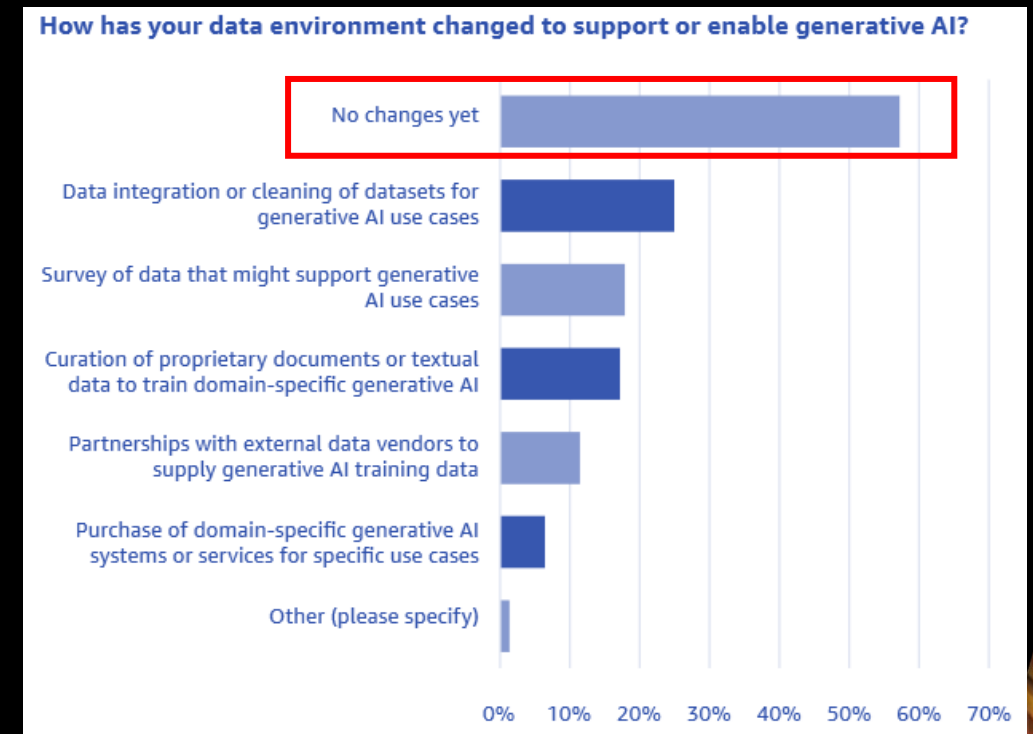


- By 2027, spending on AI software will grow to **\$297.9 billion** with a CAGR of **19.1%**.
- Over the next five years, the market growth will accelerate from 17.8% to reach **20.4%** in 2027.
- Generative AI software spend will rise from 8% of AI software in 2023 to **35%** by 2027.

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

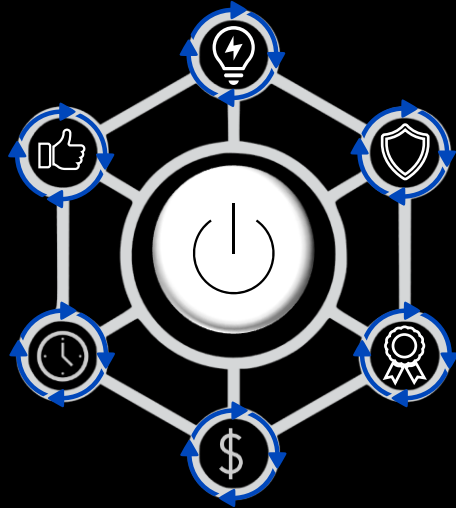
SOURCE: "CDO Agenda 2024: Navigating Data and Generative AI Frontiers." (2023), [AWSstatic.com](https://awsstatic.com)

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

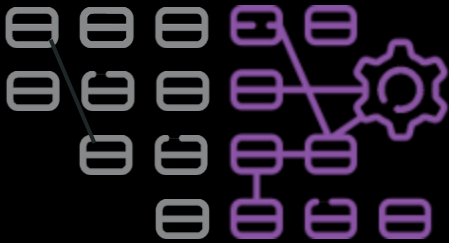
Key points to consider



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), [WPROMOTE.com](https://www.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?

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

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"

Bad AI Leads to Attorney Sanctions²

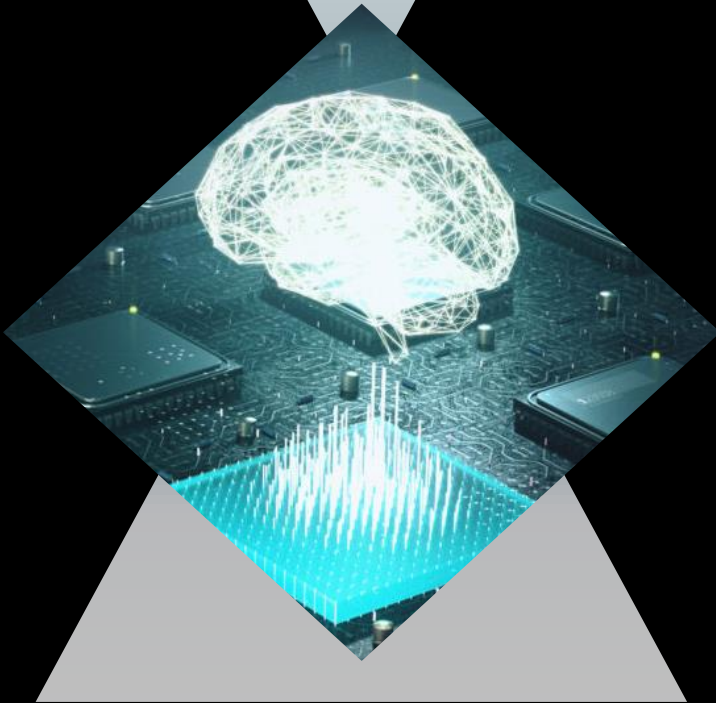
- Used generative AI to find prior cases to support a case
- At least 6 of the AI findings did NOT exist
- As of June 2023, Schwartz was facing possible sanctions by the court.

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

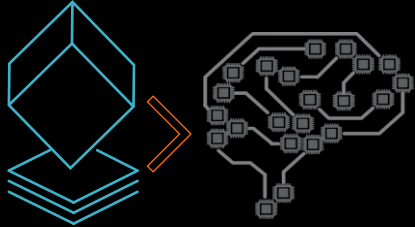
Decentralised Data Products Dilemma

- 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



• DM 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 DM

AI Powered Data Management Solutions

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



• GenAI DM

Data to All

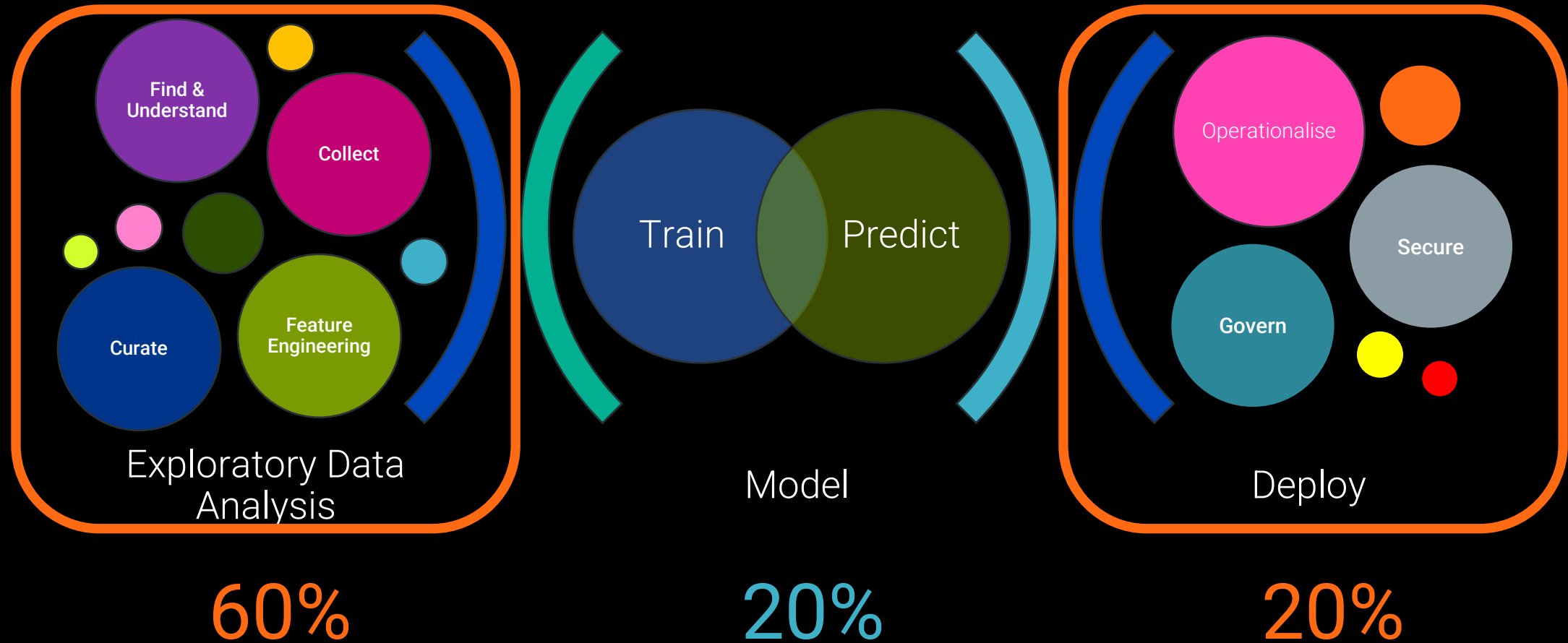
- Dramatically simplifies how organisations access and manage data
- Turbocharges data democratisation
- Exponentially increases the productivity of data teams

First AI Theme

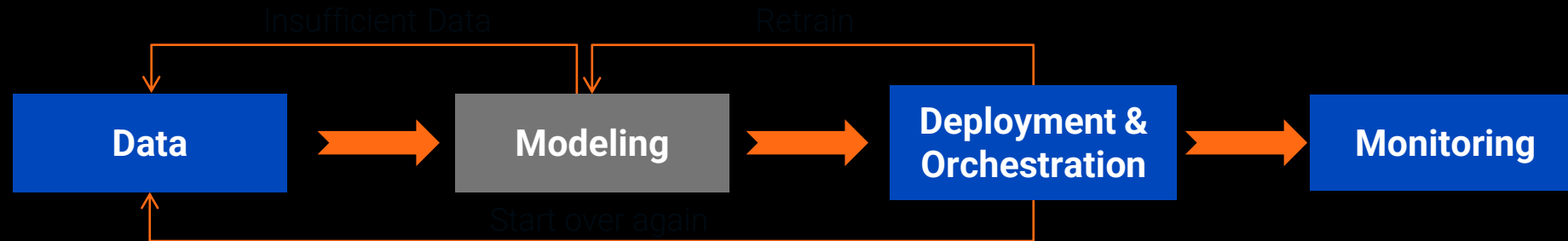
Data Management for AI

The Data Science Workflow

Streamlined Data Management is the Data Science Supercharger



Data Management is integral to AI



- Define data
- Procure data
- Data cleansing
- Data preparation
- Metadata management
- Data Governance

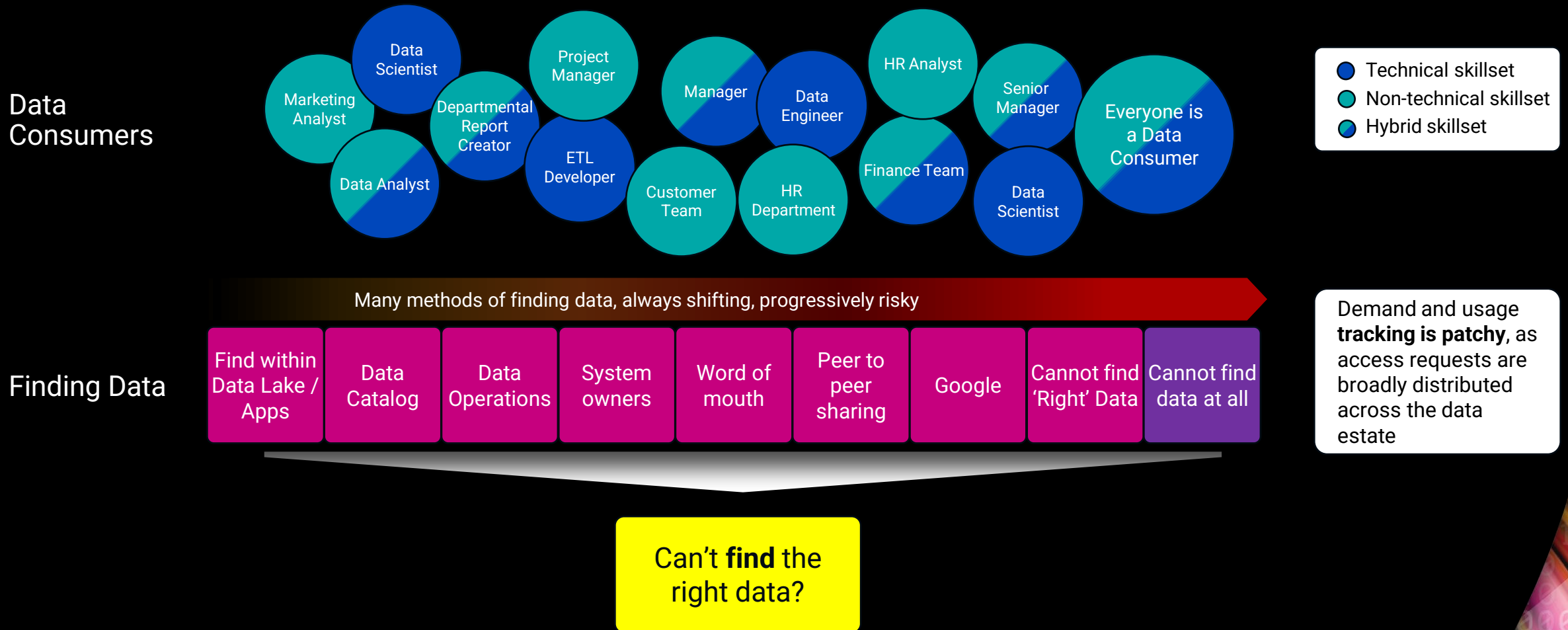
- Model selection
- Model training

- Model deployment
- Model consumption

- Model performance

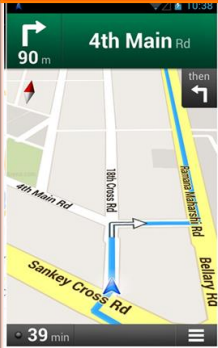
Key Challenges

1. Challenges Discovering and Understanding Data



2. Challenges with Data Quality

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

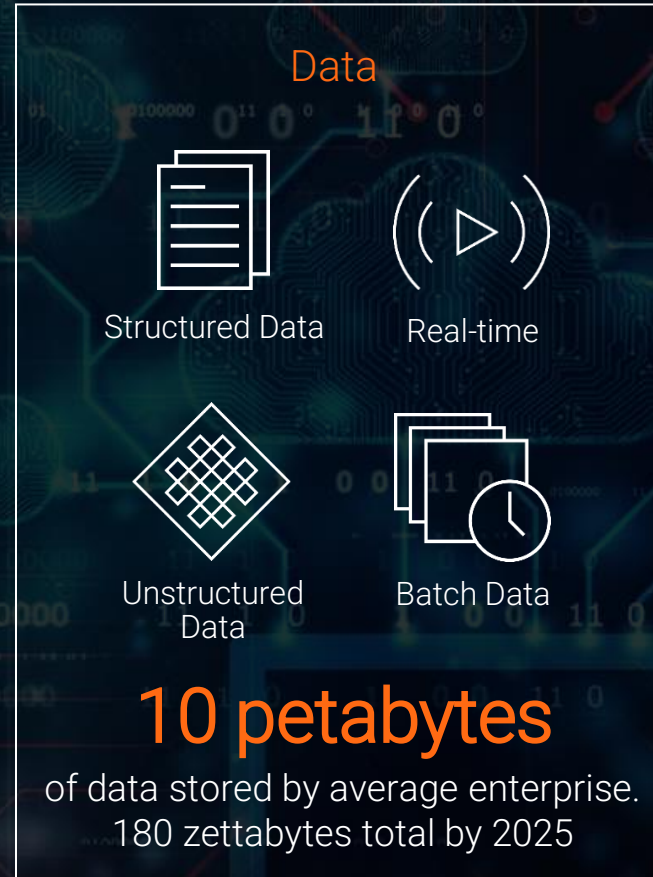
- ✓ Currency code must be consistent with country code
- ✓ Employee ID must be unique
- ✓ If customer tier is bronze then max credit is 1000
- ✓ All ICD10 codes must have a verified description

Enrich

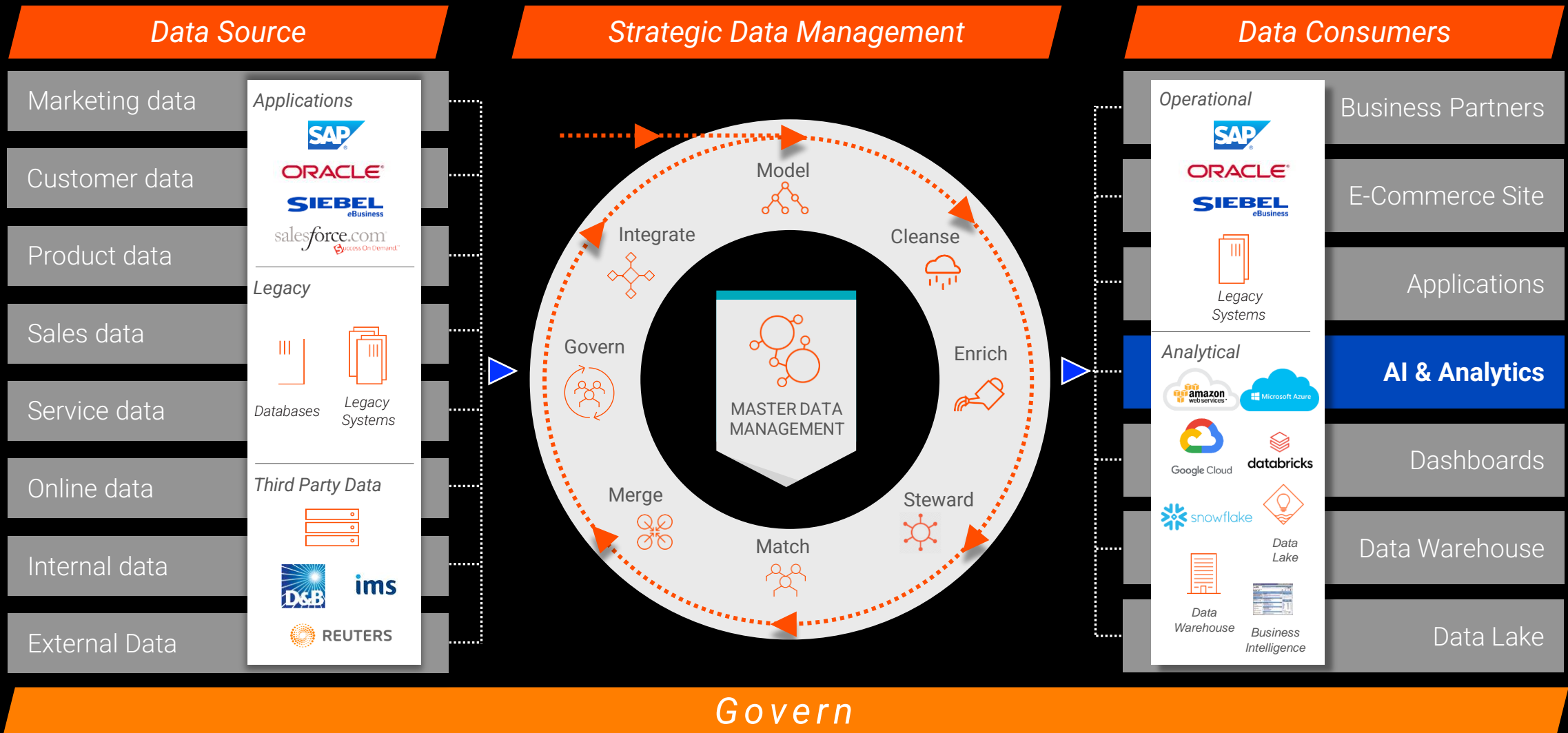
AddressL1: 1008 Avenue of the Americas
AddressL2: Suite 7
City: New York
State: NY
Zip Code: 10018-5402
Longitude: 40.7325525
Latitude: -74.004970

3. Challenges with Data Ingestion & Integration

Fragmented Landscape of Applications, Data, People...



4. Challenges with Unconsolidated Data Domains



Single View of Your Critical Data Domains Enabling Better AI Outcomes

5. Challenges Automating Modern Data Governance at Scale and AI Models 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

ARE YOU READY FOR THE AI REVOLUTION?

*Can You Ensure Non-sensitive,
private and Bias-free Data Is
Used for AI?*

*Is Your Organisation
Data Conversant and have you
Democratised Data?*

*Can Your Data Producers
Efficiently Feed AI Pipelines?*

*Are you able to cleanse, curate
and preprocess the data for
training?*

*Do you have appropriate policy
frameworks, processes and
tools to ensure governance and
compliance?*

*Will the Quality of Your Data
Ensure There Are No
Cascading Failures?*

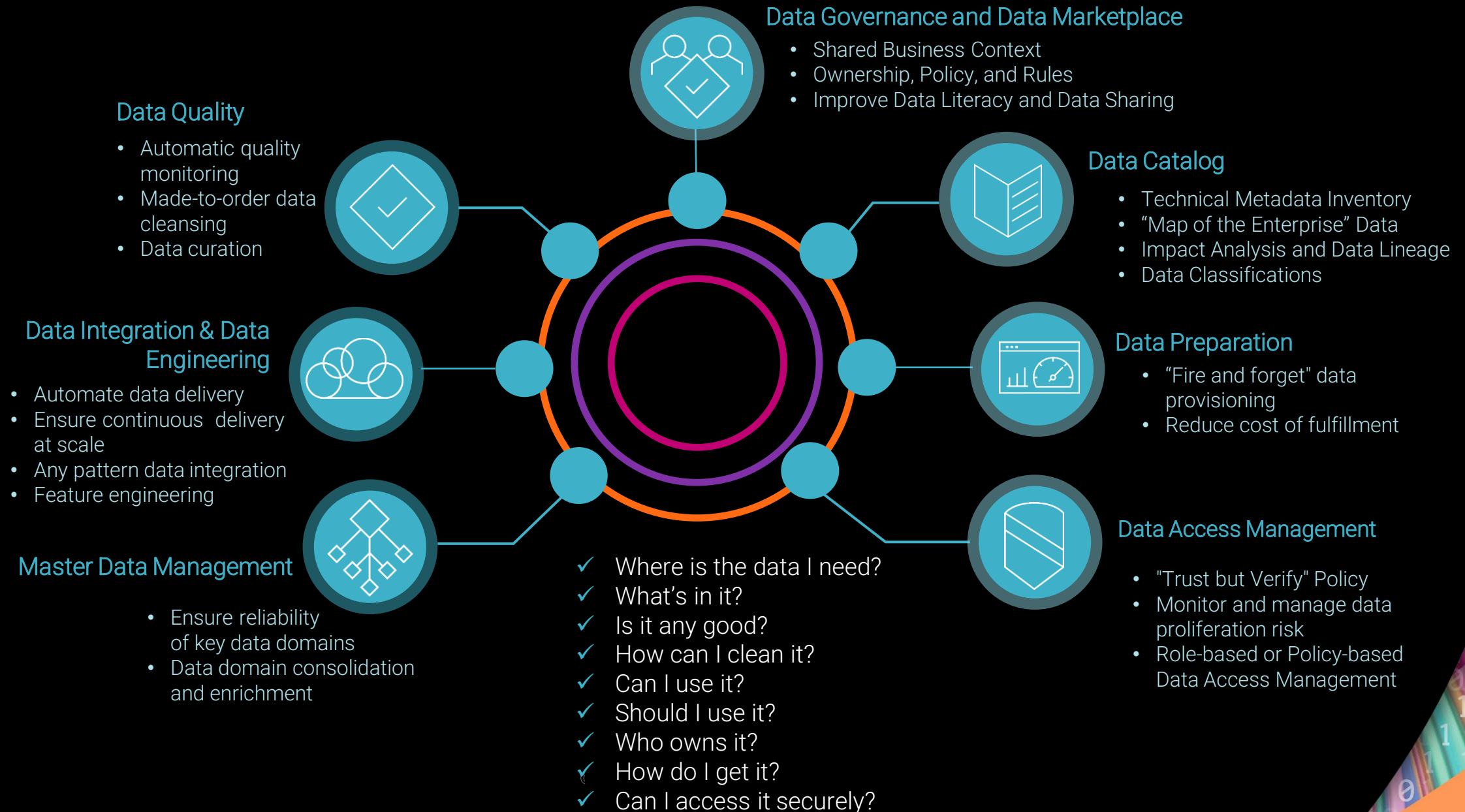
*Will It Be Easy to Find,
understand and access Data
Required for AI?*

*Are You Spending Your Time
Stitching Tools or Doing Data
Management?*

IS YOUR DATA MANAGEMENT IN SHAPE?



Data Management Capabilities to drive AI Initiatives



AI Needs Data Management

The success of AI is dependent on holistic, governed, & trusted data

- The success of AI models is dependent on the availability of **trusted** and **timely** data.
 - If data is missing, incomplete, or inaccurate, the model's behavior will be adversely affected during both training and deployment, which could lead to **incorrect** or **biased** predictions and **reduce** the **value** of the entire effort.
- AI also needs **AI-powered data management** to quickly find all the features for the model
 - Transform and prepare data to meet the needs of the AI model
 - Feature engineering, feature scaling, standardisation, etc..
 - Deduplicate data
 - Provide trusted master data about customers, partners, and services; and
 - Provide end-to-end lineage of the data
 - Including within the model and its operations.



Second AI Theme

AI for Data Management

Challenges that AI paired with Data Management will address

Scale & Agility



Policy & Trust



Skills & Literacy



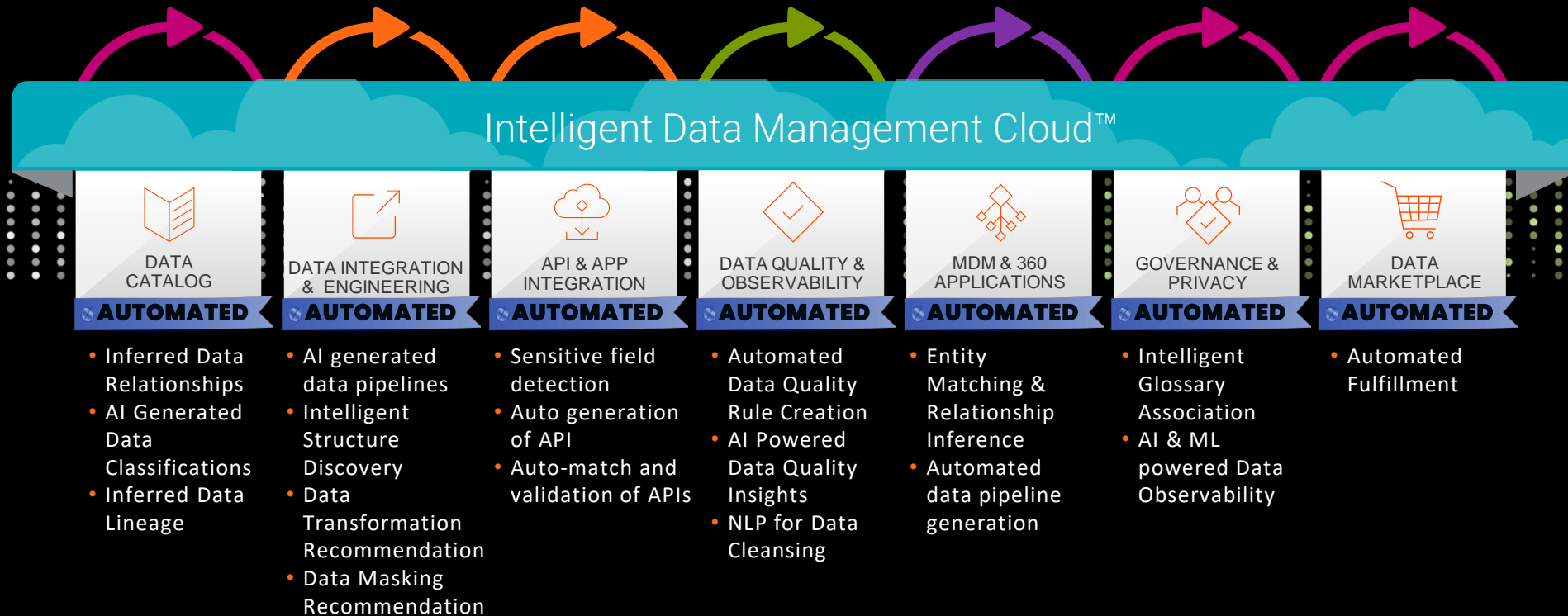
What is CLAIRE?

Bringing Data and AI to Life through AI-Powered Data Management



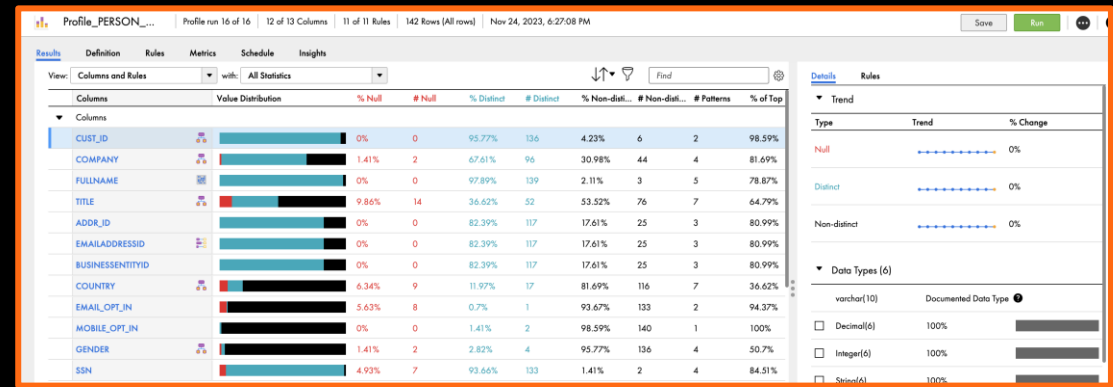
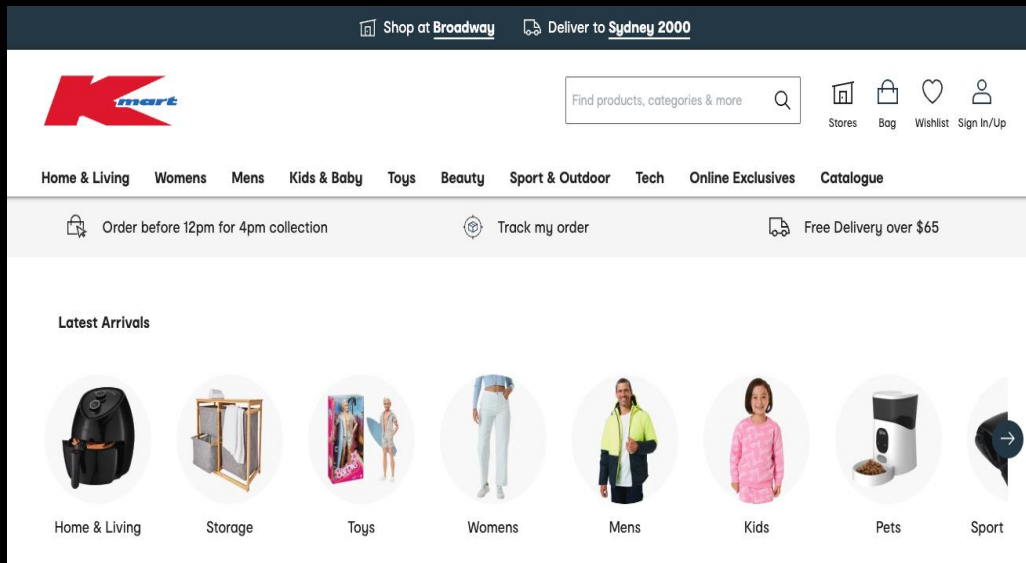
CLAIRE as an AI copilot

Driven by CLAIRE® AI-embedded Data Management Intelligence and Automation





- 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**



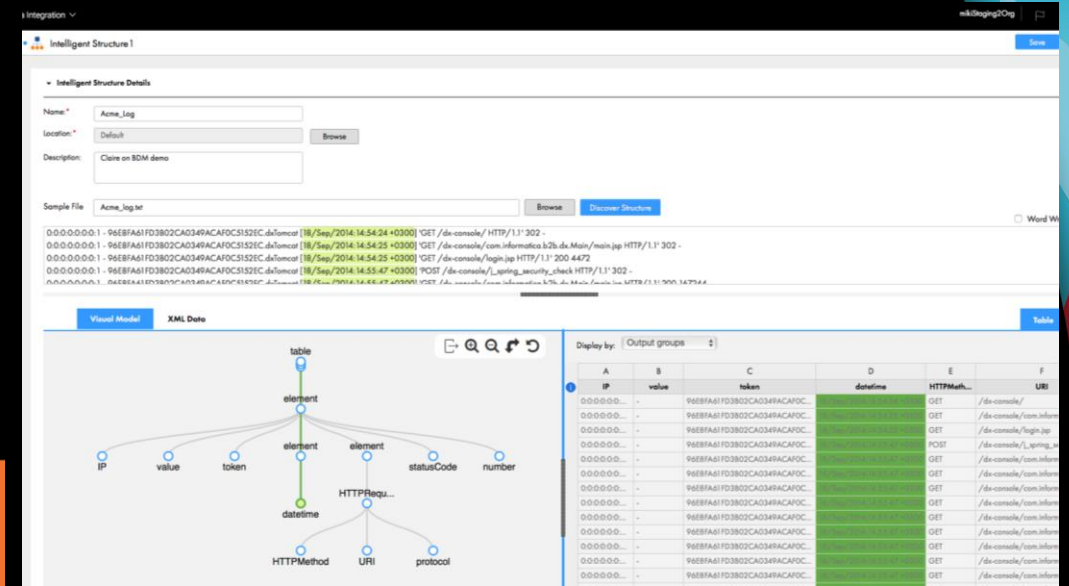
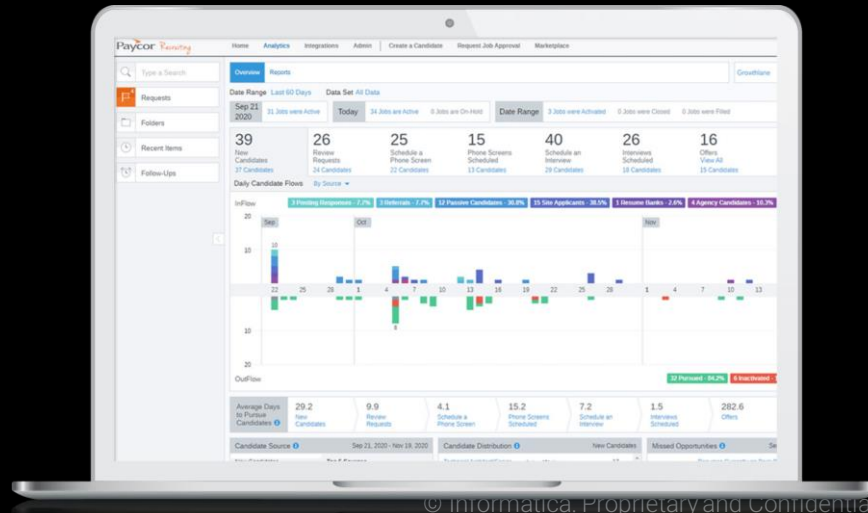
The screenshot shows a data insights tool interface. The main table lists insights:

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

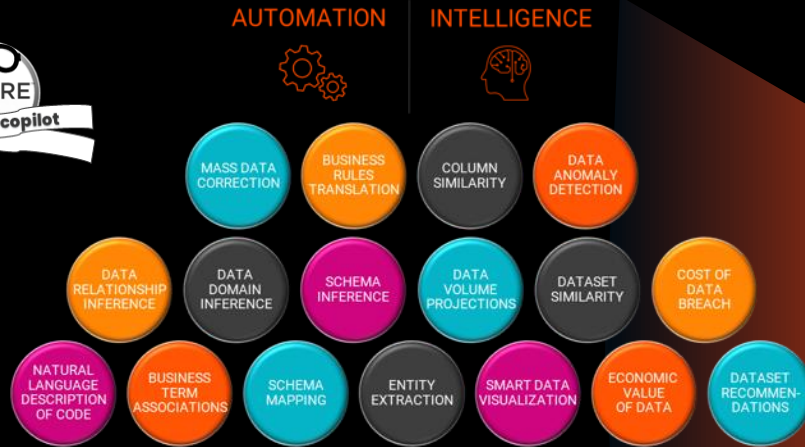
Third AI Theme

Gen AI Data Management

CLAIRE®: AI-Powered Data Management

CLAIRE AI copilot

IN PRODUCT AI-POWERED METADATA INTELLIGENCE AND AUTOMATION



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

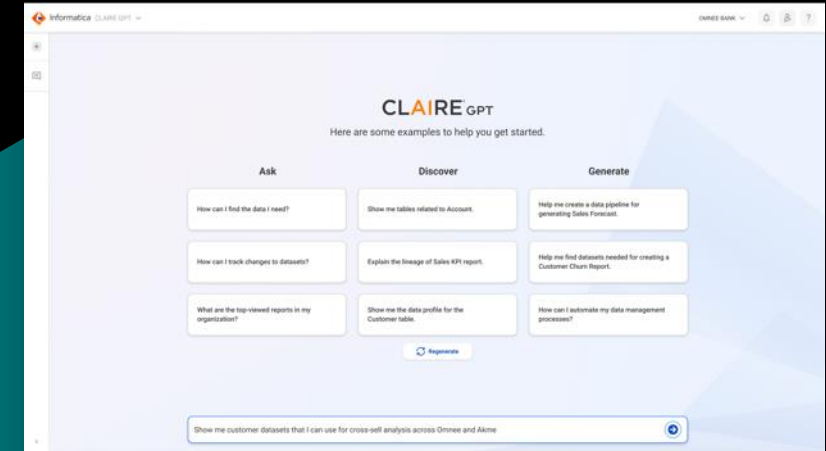
DATASET RECOMMENDATIONS

CLAIRE®
UNIFIED METADATA
INTELLIGENCE



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

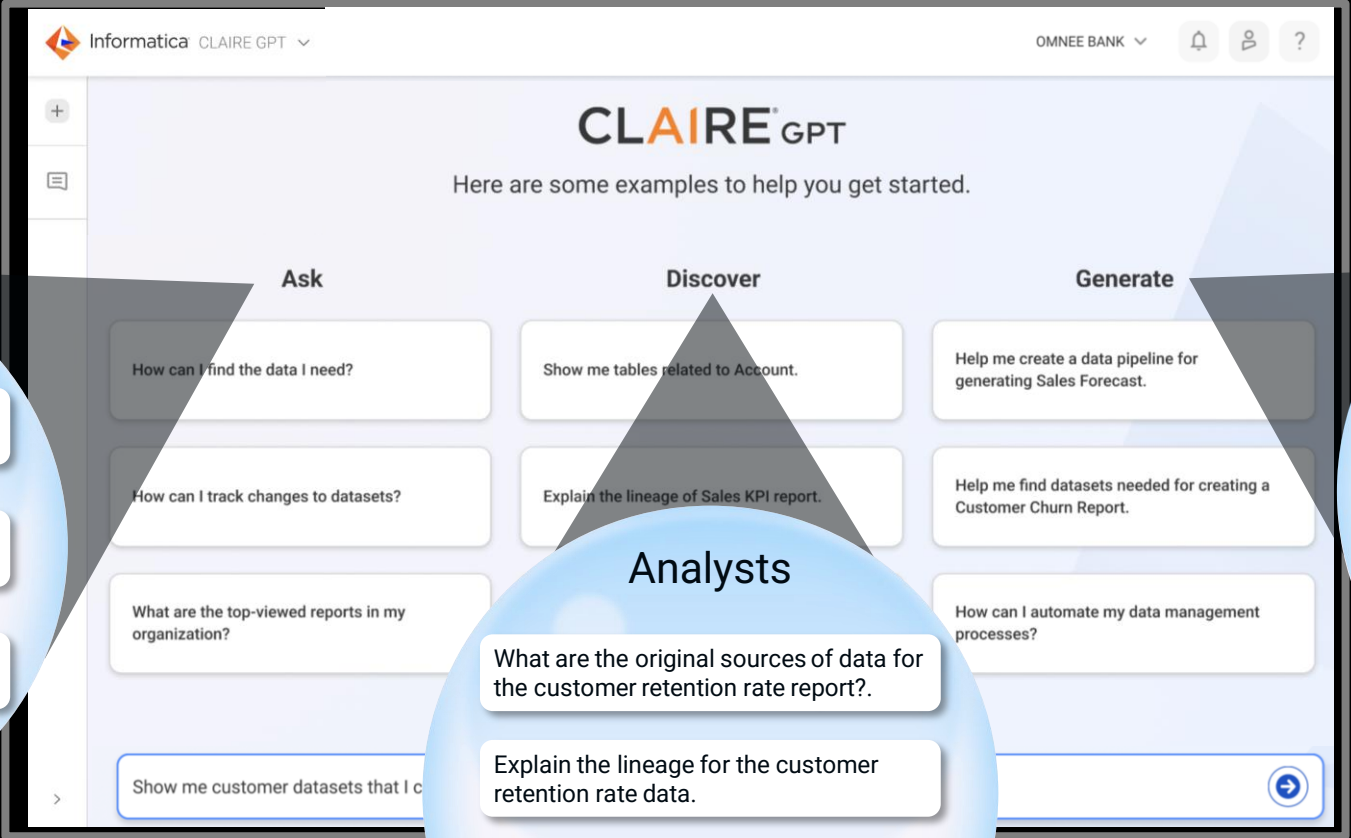
CLAIRE GPT

Empowers More Business Users Across the Organisation to Speak to their Data



CLAIRE[®] GPT Possibilities

Simplifying Data Management for All Users



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?

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?

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.

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



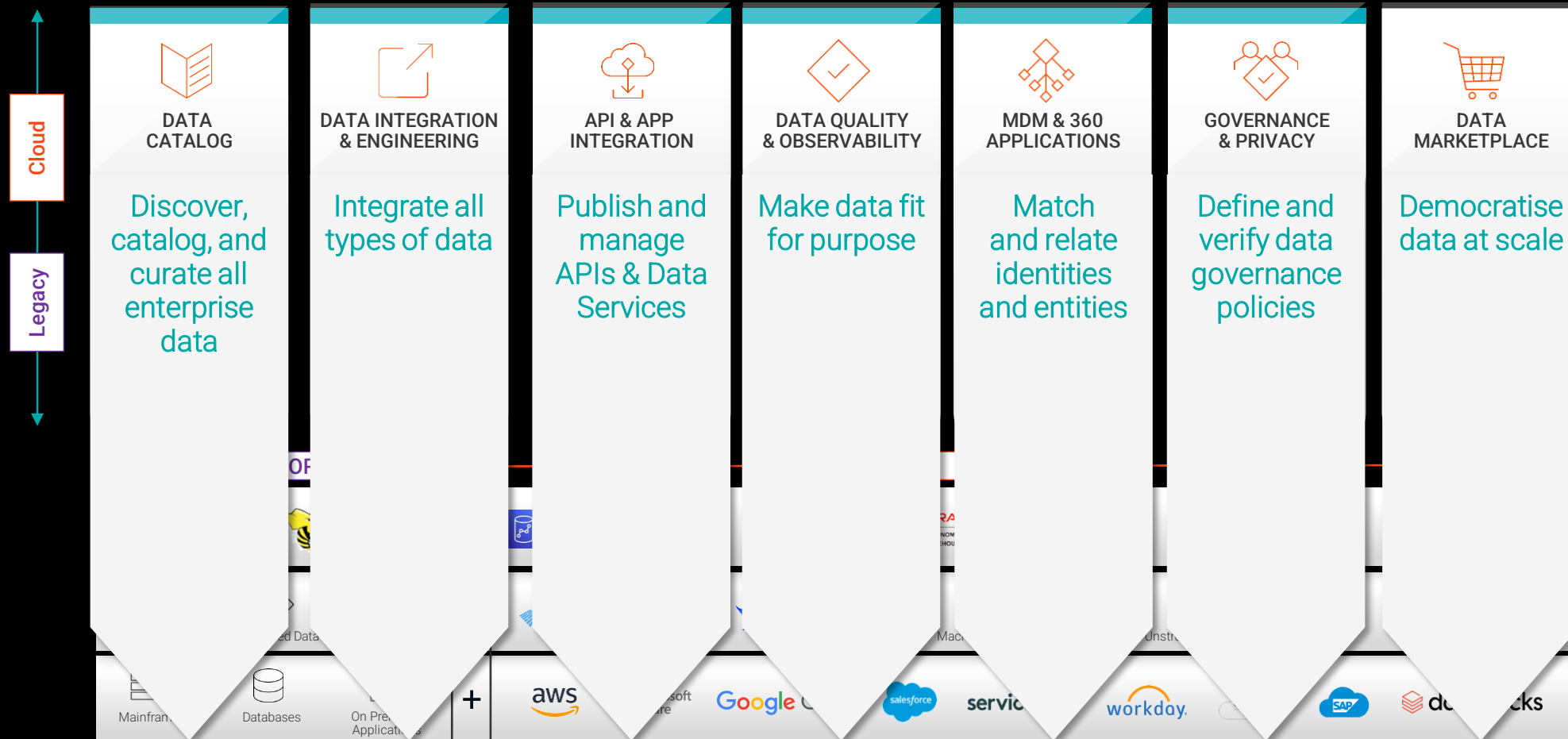
Great data for great AI

Informatica

Where data
& AI come to **LIFE**

Impact on Data Management

One Code-less Solution for all your Data Management Needs



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

workday

servicenow

salesforce



SaaS Apps
Sources

+



Mainframe



Applications



Databases

On-premises
Sources

+



IoT



Machine Data



Logs

Real-time /
Streaming
Sources

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

Your local Informatica Team



Jason Leith
QLD Account Manager



Fian Clark
Territory Account Manager



Alan Yves
Solution Architect



Thank You

Where data
& AI come to **LIFE**