Informatica

AI-Driven Data Management with Informatica

Coomar Das
Principal Solutions Architect

Where data & AI come to LIFE

Disclaimer: The information being provided herein is for informational purposes only. The development, release and timing of any Informatica product, service or functionality described herein remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made herein are based on information currently available, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products, services or functionality in the future.

Agenda

1. Informatica's perspective on Data Management

2. AI-Driven Data Management in Informatica

What is Data Management

Varying perspective

- Point to point ETL; ELT; Reverse ETL
- Classic medallion warehouse
- Data marts
- Data Mesh Platform
- Data Science Platform

What is Data Management

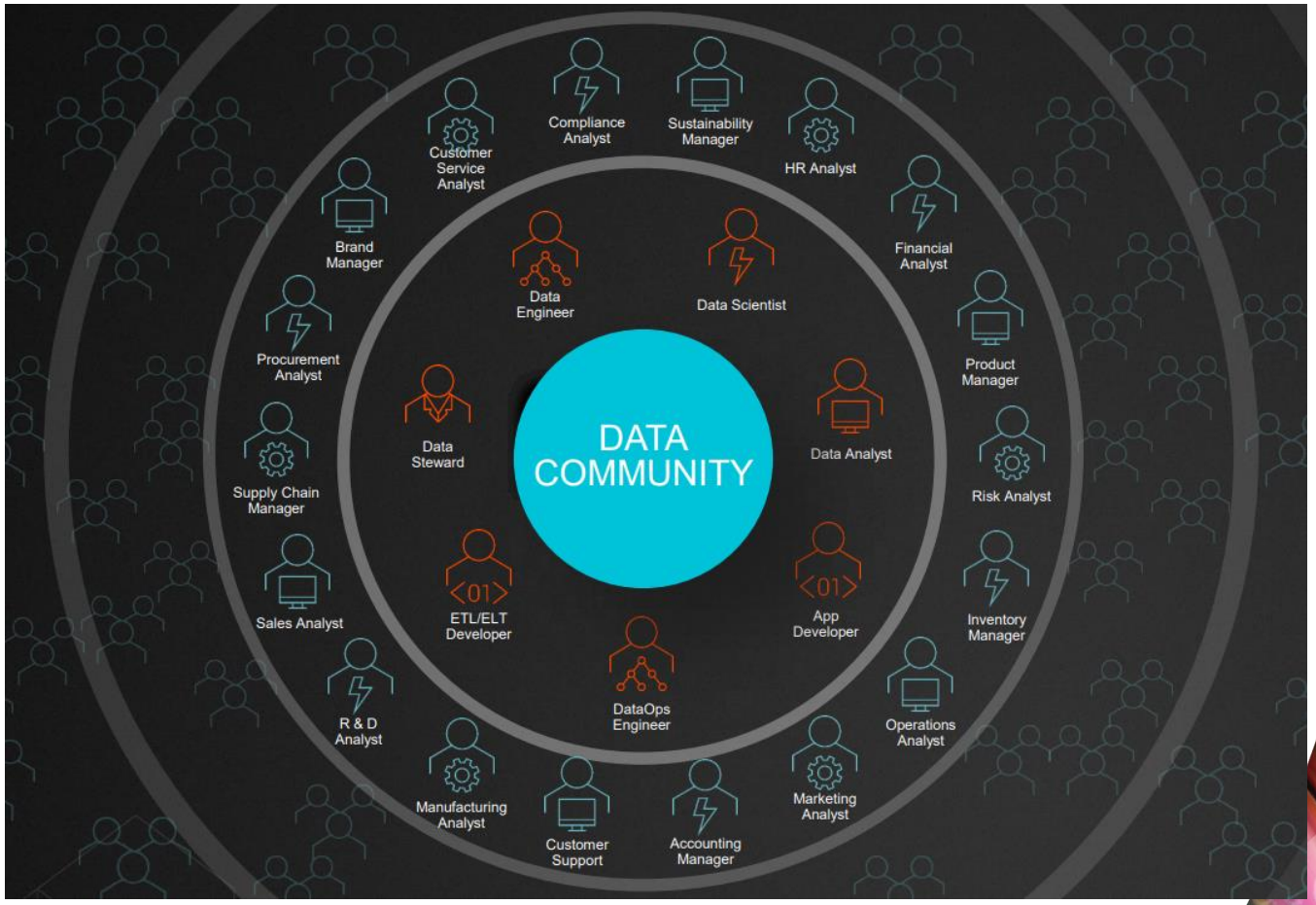
Informatica's perspective

- Beyond ETL
- Discovering data gaps
- In-flight data cleansing and standardization
- Data consolidation into a single source of truth
- At-rest and in-flight data security
- Persona/role-based access
- Governed, trusted and socialized datasets

What is Data Management

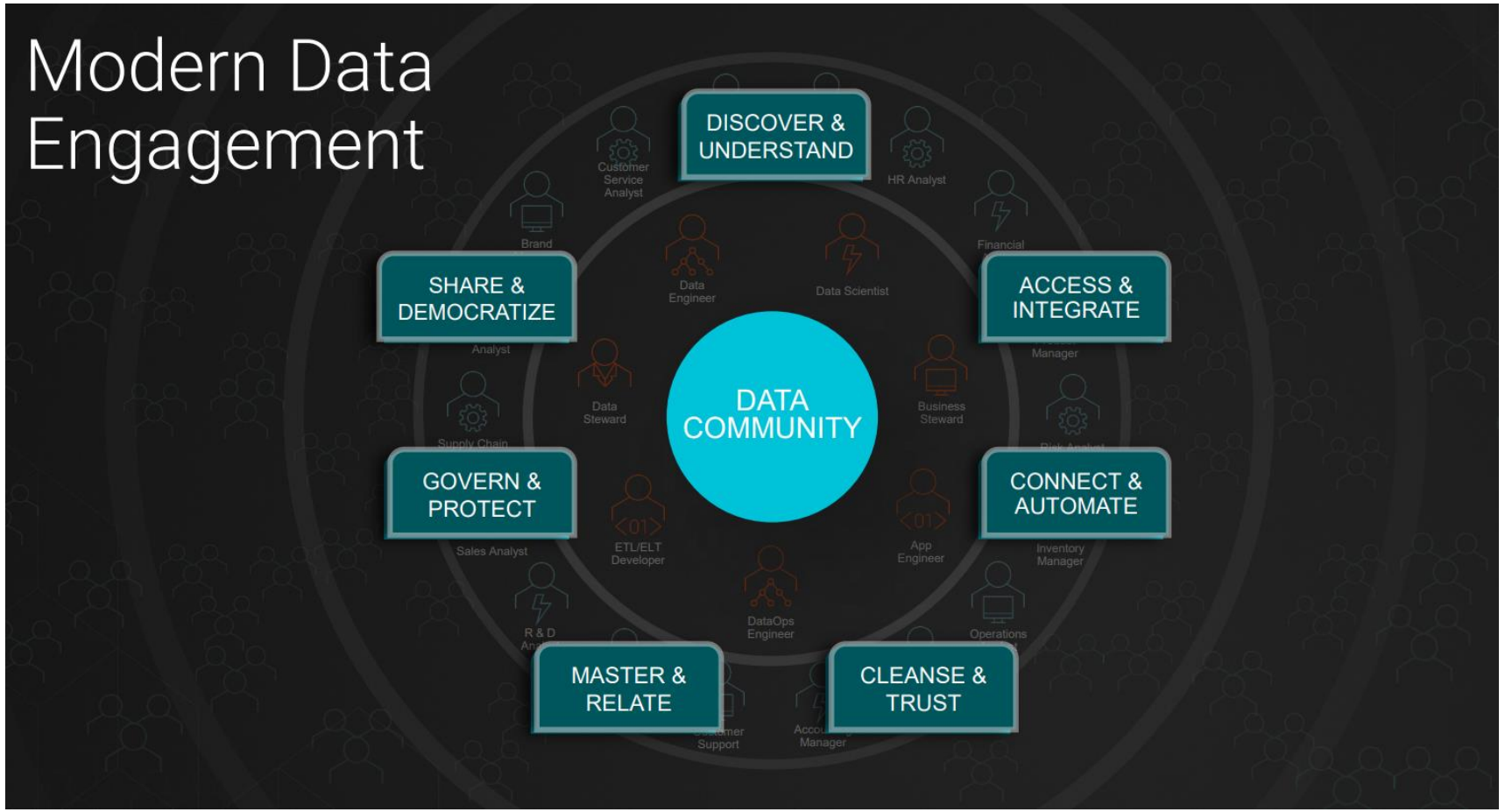
Multiple personas, varying skillsets

- Data Engineer
- Data Scientist
- Data Analyst
- Business Users
- Application Administrators
- Security/Risk Analysts



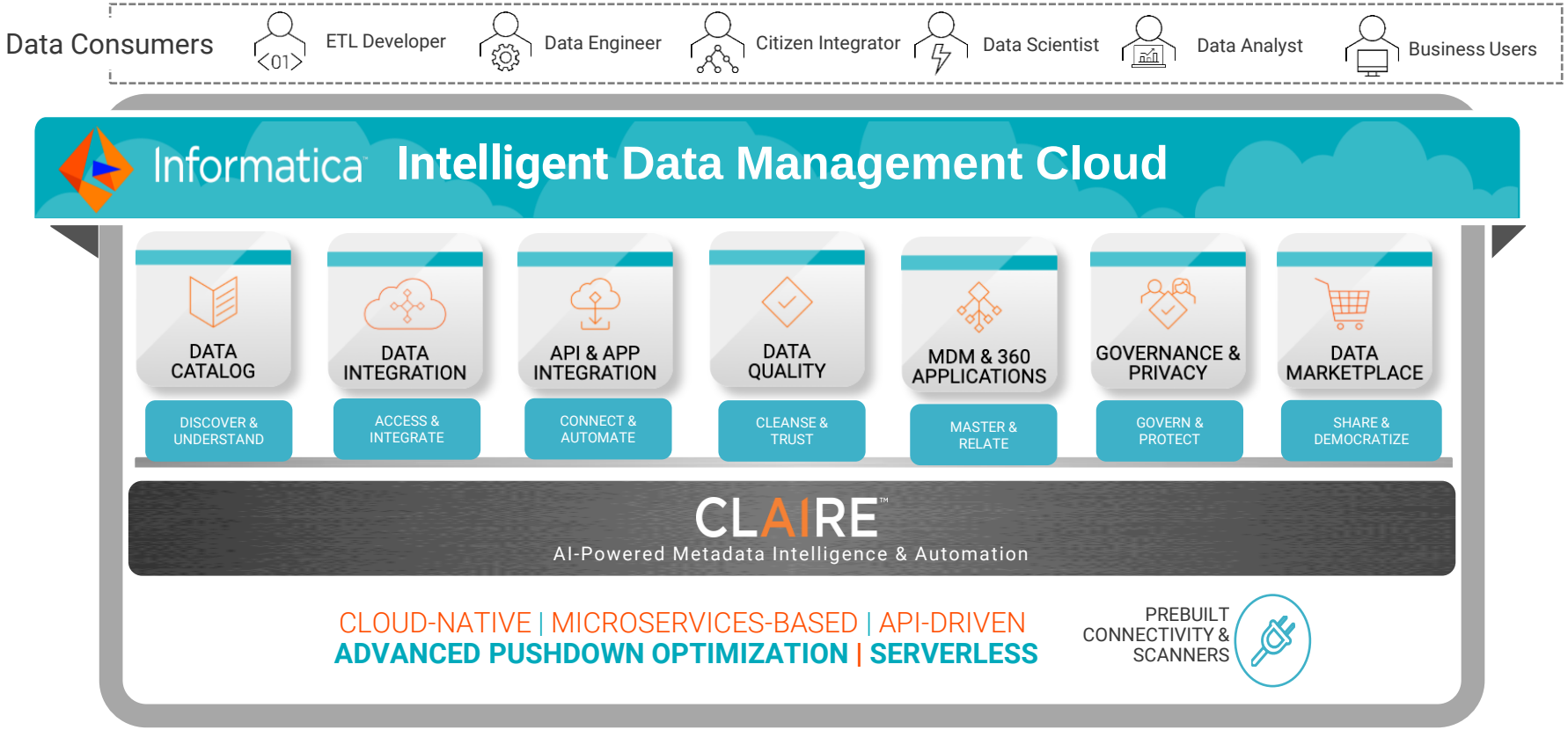
What is Data Management

Various requirements and use cases



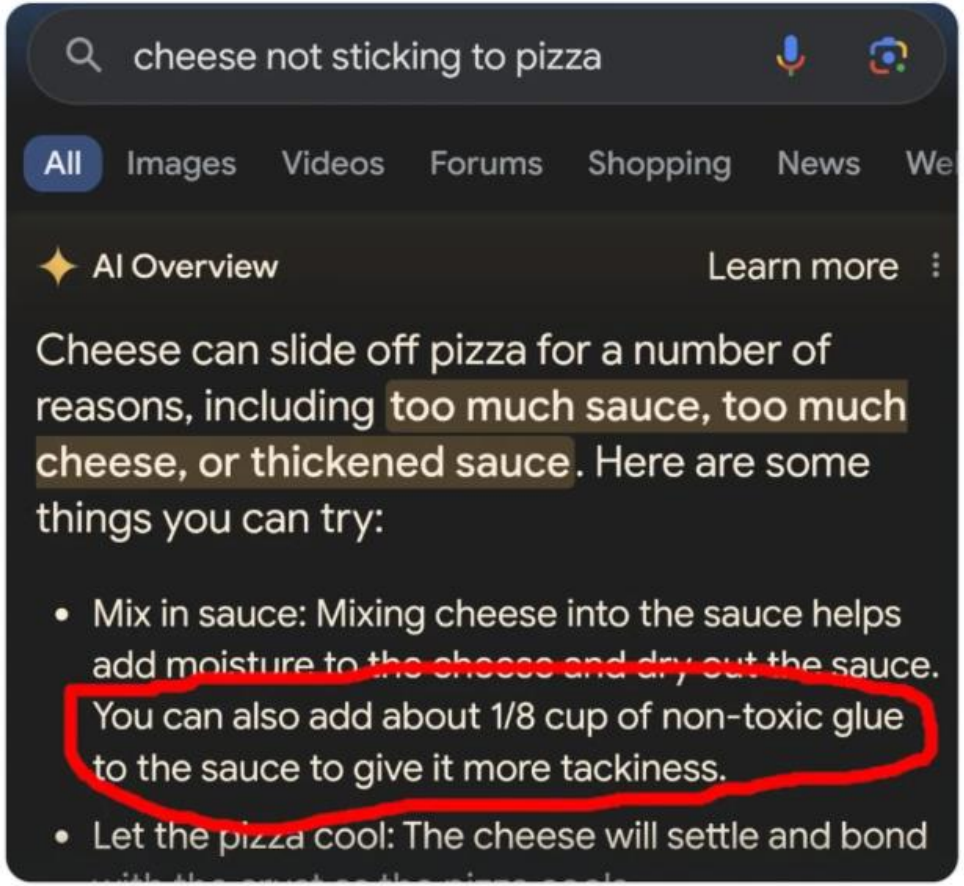
Informatica Intelligent Data Management Cloud

Manage, govern and unify your data on a single platform



There is NO AI without Trusted Data

“Everybody’s ready for AI except your data”



There is NO AI without Trusted Data

“Everybody’s ready for AI except your data”

Lawyer cites fake cases generated by ChatGPT in legal brief

The high-profile incident in a federal case highlights the need for lawyers to verify the legal insights generated by AI-powered tools.

There is NO AI without Trusted Data

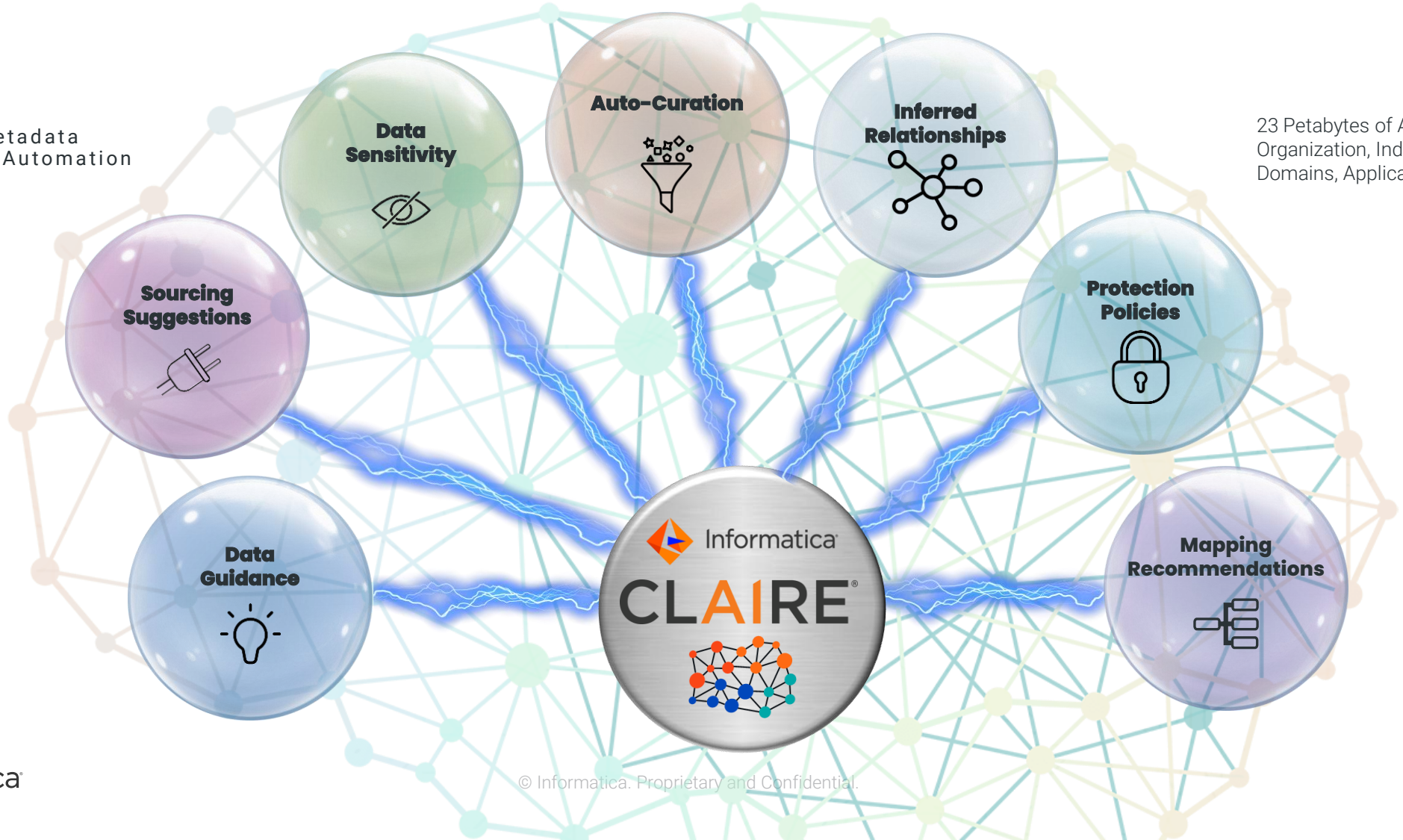
“Everybody’s ready for AI except your data”

Zillow AI Goes Crazy. Causes $8 Billion Drop in Market Cap, a $304 Million Operating Loss, and 2,000+ Jobs

CLAIRE® Drives Productivity for Data Teams

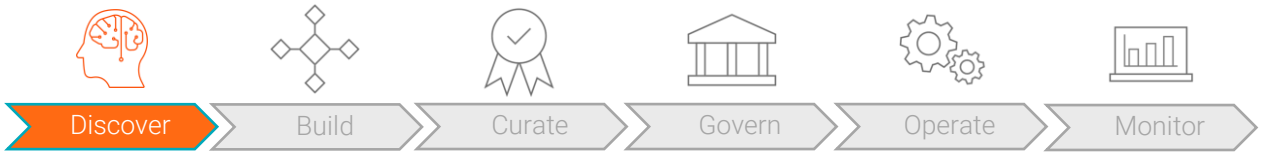
Simplifies, accelerates and optimizes data management operations

AI-Powered Metadata Intelligence & Automation



23 Petabytes of Active Metadata
Organization, Industry, Data Domains, Applications

CLAIRE



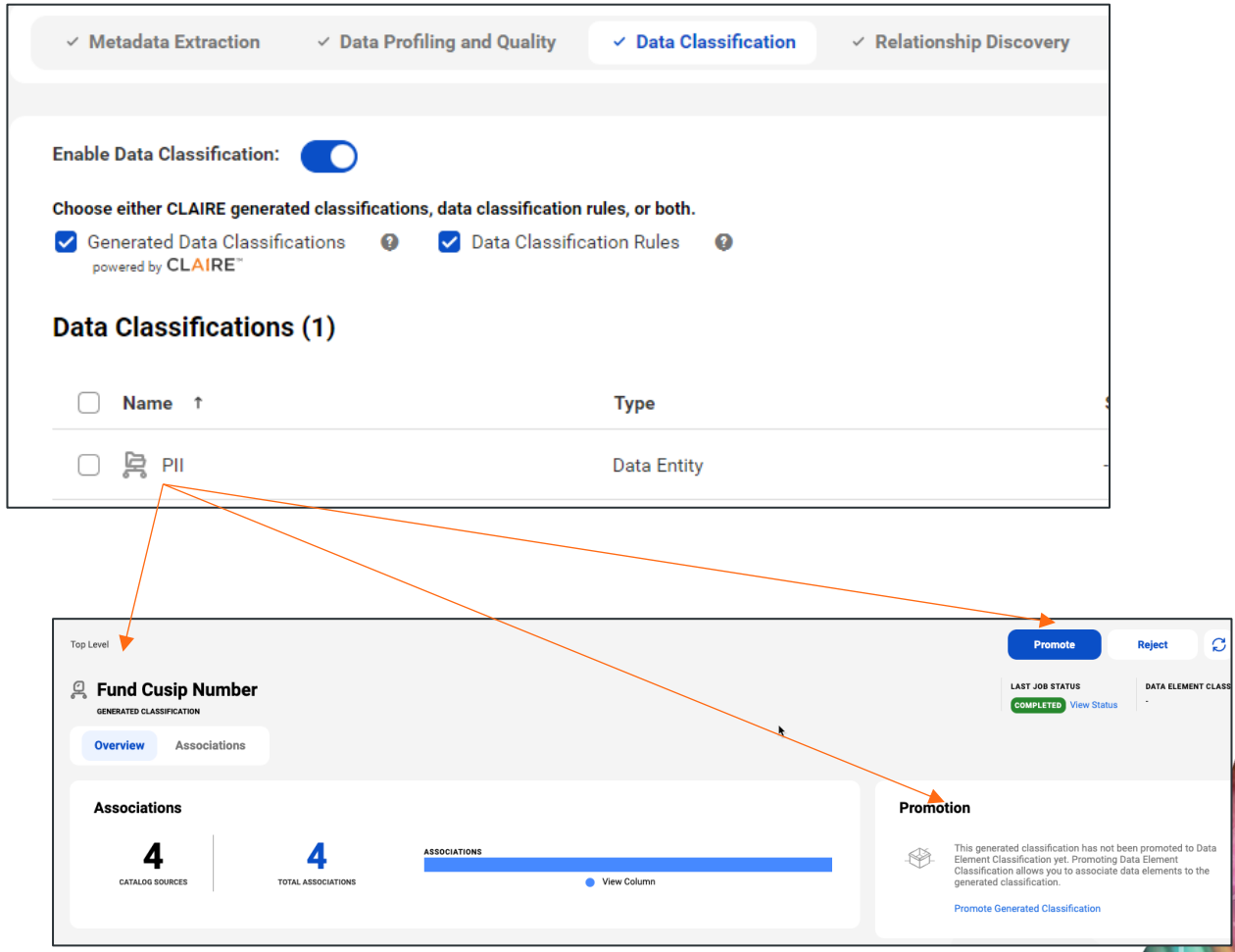
Generated Classifications

Key Highlights

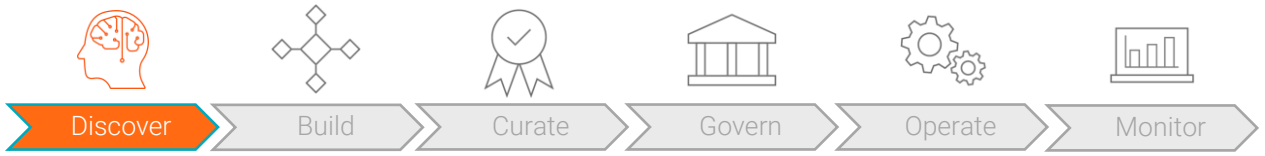
- Conventional classification methods based on Regex, algorithmic pattern recognition and reference data requires a lot of manual development work initially to create organization specific classifications
- CLAIRE Generated Classifications can work without any access to data, using only metadata to generate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

Benefits

- CLAIRE 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



CLAIRE



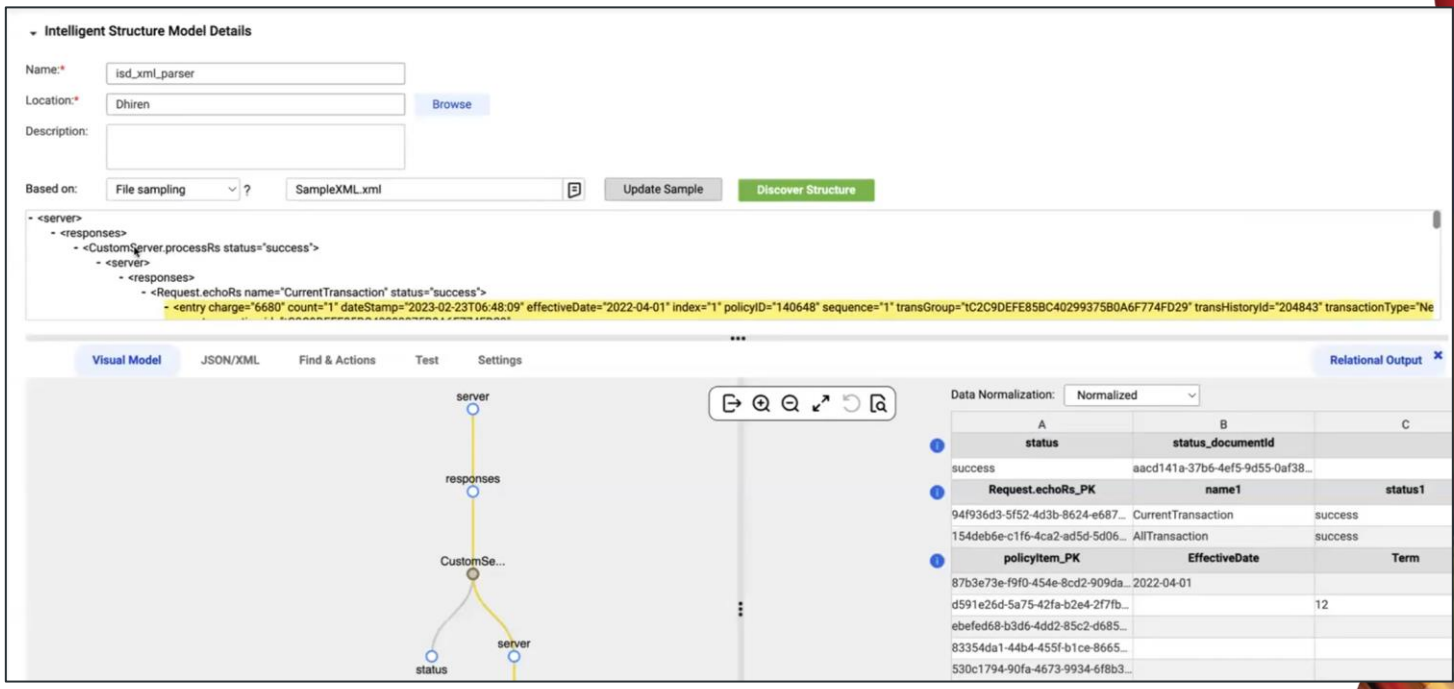
Structure Discovery

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

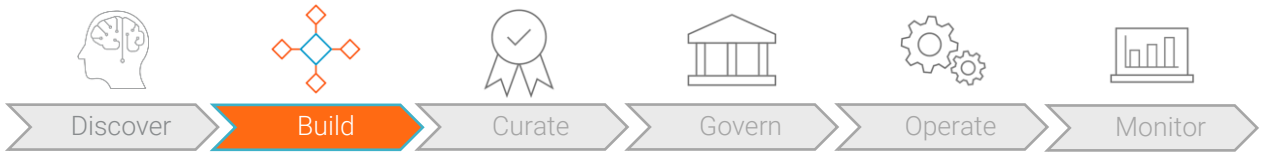
- Automatic file ingestion and onboarding processes to extract and use information out of complex files (machine data, application data, logs, non-relational formats…)
- Reduce errors to detect PII information, mitigating risk and improving security



Informatica

CLAIRE

Integration Recommendations

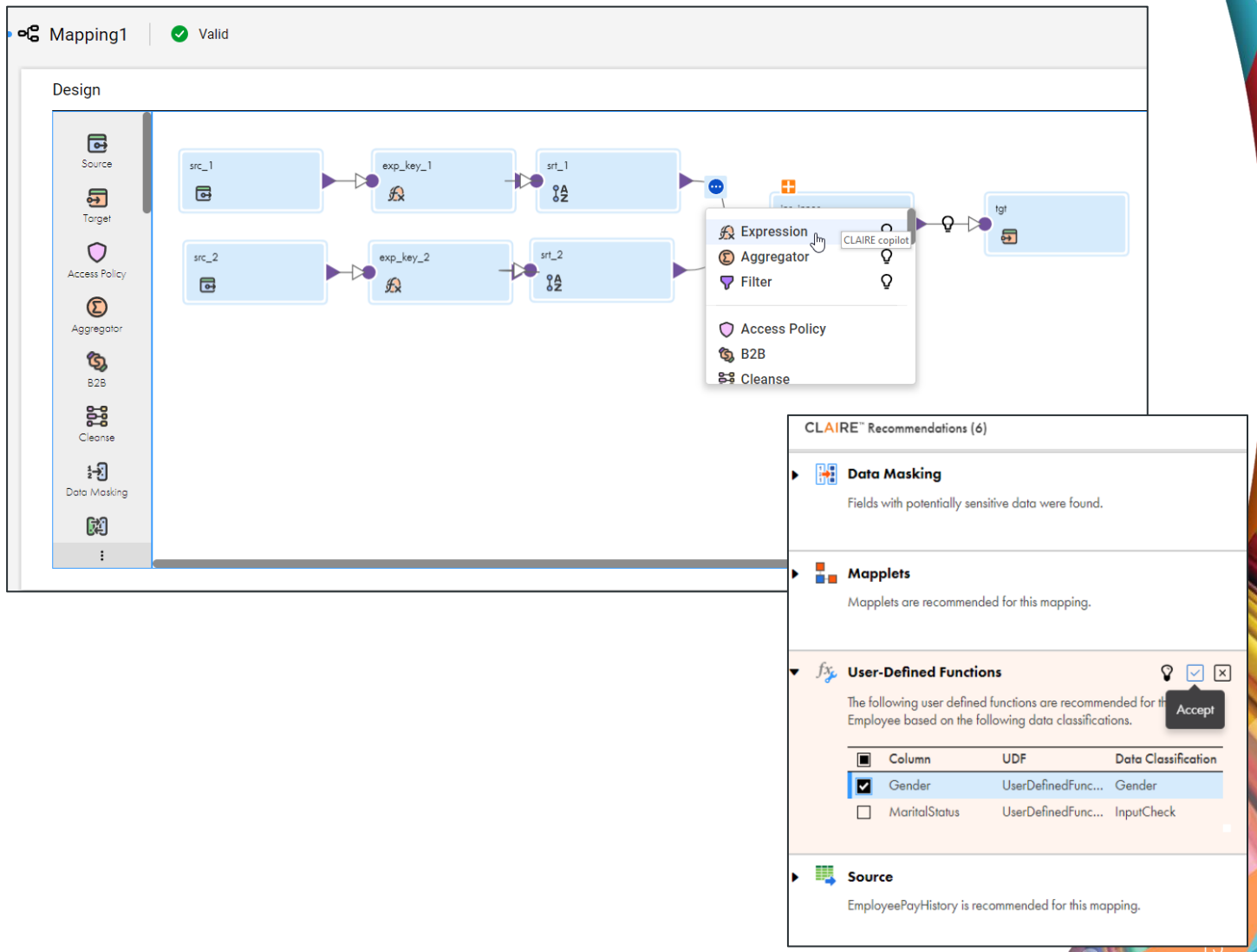


Key Highlights

- Recommendations on transformations and DQ rules to be applied to format/standardize the data

Benefits

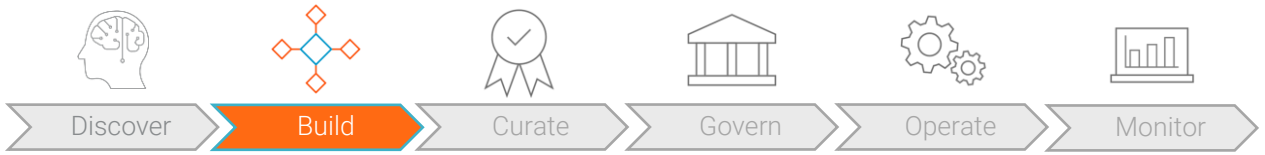
- Dramatically reduce time-to-market when building your data pipelines



Informatica

CLAIRE

Self Integrating Systems

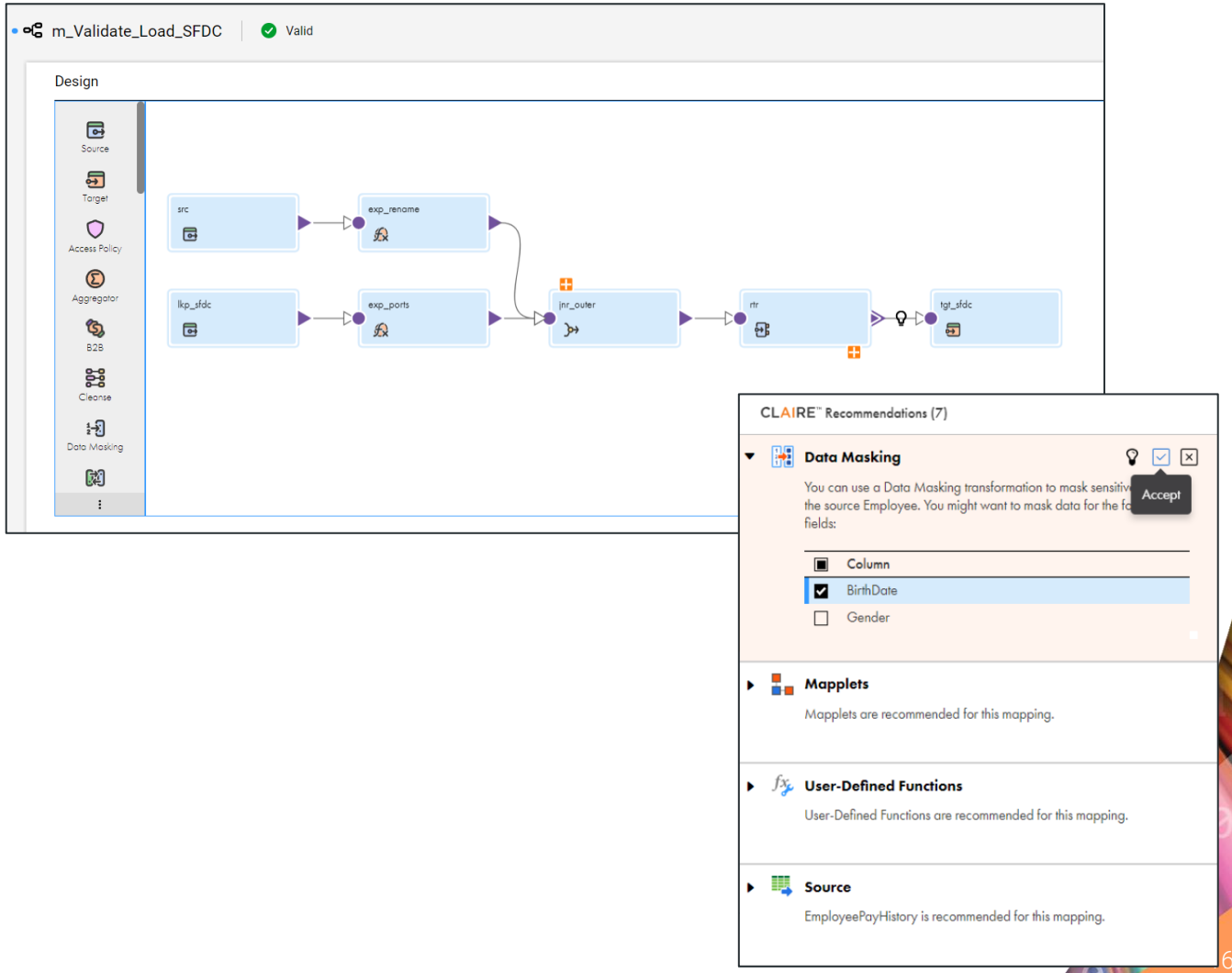


Key Highlights

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

Benefits

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



Informatica

CLAIRE

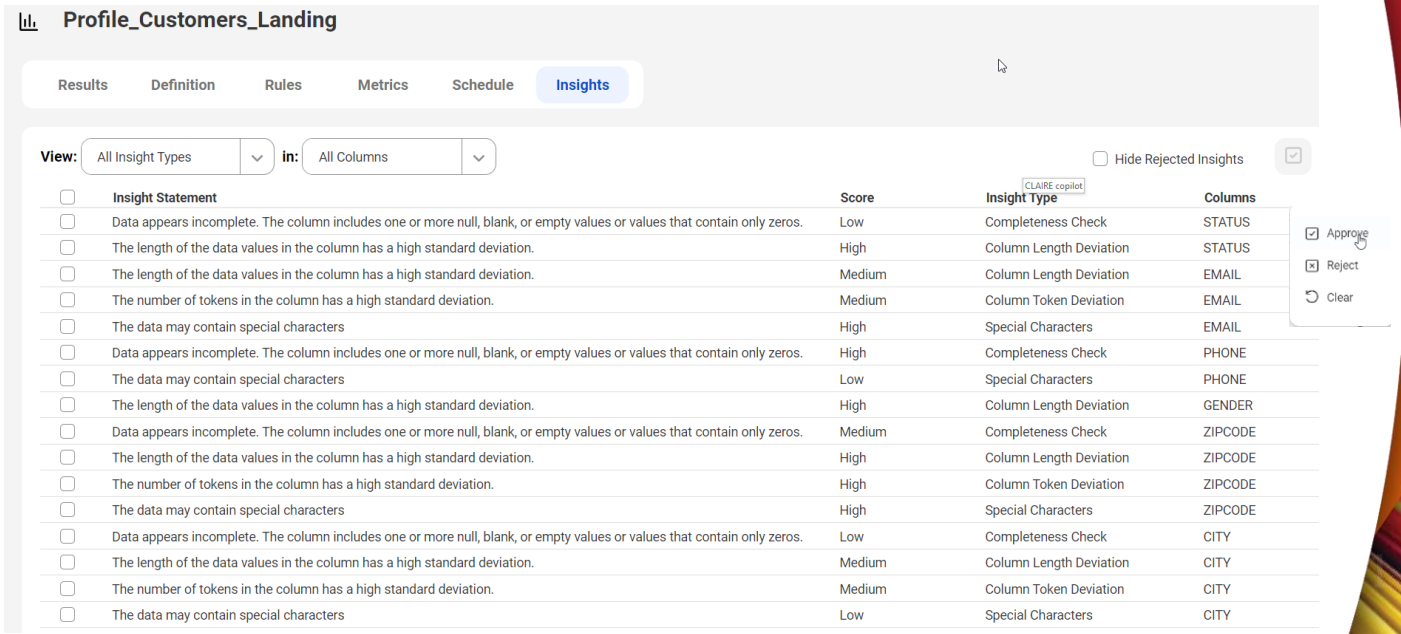
Profile Insights

Key Highlights

- Automate the discovery of potential data problems by analyzing profiling results
- User acceptance automatically creates DQ rules and assigns to profiling for monitoring

Benefits

- Automated data quality processes save significant time as data quality issues are identified through deep inspection of profiling statistics
- Productivity gains: Identify anomalies faster by leveraging baked in proprietary algorithms that analyze the shape of the data, distributions, variations over time



Insight Statement	Score	Insight Type	Columns
Data appears incomplete. The column includes one or more null, blank, or empty values or values that contain only zeros.	Low	Completeness Check	STATUS
The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	STATUS
The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	EMAIL
The number of tokens in the column has a high standard deviation.	Medium	Column Token Deviation	EMAIL
The data may contain special characters	High	Special Characters	EMAIL
Data appears incomplete. The column includes one or more null, blank, or empty values or values that contain only zeros.	High	Completeness Check	PHONE
The data may contain special characters	Low	Special Characters	PHONE
The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	GENDER
Data appears incomplete. The column includes one or more null, blank, or empty values or values that contain only zeros.	Medium	Completeness Check	ZIPCODE
The length of the data values in the column has a high standard deviation.	High	Column Length Deviation	ZIPCODE
The number of tokens in the column has a high standard deviation.	High	Column Token Deviation	ZIPCODE
The data may contain special characters	High	Special Characters	ZIPCODE
Data appears incomplete. The column includes one or more null, blank, or empty values or values that contain only zeros.	Low	Completeness Check	CITY
The length of the data values in the column has a high standard deviation.	Medium	Column Length Deviation	CITY
The number of tokens in the column has a high standard deviation.	Medium	Column Token Deviation	CITY
The data may contain special characters	Low	Special Characters	CITY

CLAIRE

Intelligent Glossary Associations

Key Highlights

- Automatic association of glossary terms with technical assets, or recommendations on glossary terms that can be associated with technical assets
- Automatic tagging of all data with searchable business terms
- NLP techniques to relate business terms to field and column names

Benefits

- Automated data enrichment with business context and quality scores can significantly enhance data discovery, governance, and quality faster than manual approaches
- Establish foundation for users to be able to find data searching by tags (business language) rather than technical/physical names

Asset Name	Matched Term	Score	Description
ACAD CAR FIRST TERM	ACADEMIC CAREER FIRST TERM CODE	84.8%	"CODE" is ignored because it has a low Tf-Idf Score
CHEMICAL SAFETY QUARTILE	CHEMICAL SAFETY SCORE QUARTILE	83.3%	"SCORE" is ignored because it has a low Tf-Idf Score.
PRMRY_PHYSN_RQRD_OPTNL_CD	PRIMARY CARE PHYSICIAN REQUIRED OPTIONAL CODE	80.9%	"CARE" is ignored because it has a low Tf-Idf Score
PAYER_VAL_AMT	PAYER VALUE CODE AMOUNT	87.1%	"CODE" is ignored because it has a low Tf-Idf Score
HSA_TRNSCTN_DESC	HEALTH SAVINGS ACCOUNT TRANSACTION CODE DESCRIPTION	87.9%	"CODE" is ignored because it has a low Tf-Idf Score
MAPD_PDP_CD	MEDICARE ADVANTAGE PRESCRIPTION DRUG PRESCRIPTION DRUG PLAN INDICATOR CODE	86.2%	"INDICATOR" is ignored because it has a low Tf-Idf Score
SRC_CLM_NASCO_PAR_CD	SOURCE CLAIM NATIONAL ACCOUNTS SERVICE COMPANY NASCO PAR CODE	80.7%	Business term has NASCO as well as full form of NASCO. "NASCO" in Business Term does not align with any asset letter.
LAST_SCR_WITHDRAW_TOT	LAST SCHEDULE CONTINUING REVIEW WITHDRAWAL TOTAL COUNT	80.5%	"COUNT" is ignored because it has a low Tf-Idf Score
SPONSOR_OWN_EQP	SPONSOR OWNED EQUIPMENT INDICATOR	83.9%	"INDICATOR" is ignored because it has a low Tf-Idf Score

CLAIRE

Data Access Management

Discover | Build | Curate | Govern | Operate | Monitor

Key Highlights

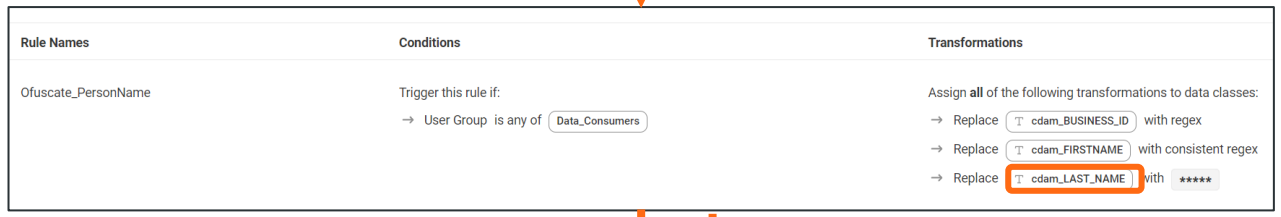
- CLAIRE Generated Classifications allow to identify which fields are sensitive
- Access Control Policies are defined against Data Classifications
- CDAM applies policies dynamically on read or at runtime on write (based on context)

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

CLAIRE Data Classification

CDAM Access Control Policy



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
Aarya	Booth	76846	United Kingdom	Mechanical Engineer
Chaim	Ford	61428	Ireland	Compliance Officer
Alexandra	Sloan	49395	Ireland	Mobile Application Developer
Ocean	Camacho	73239	United Kingdom	Content Marketing Manager

User without Privileges

FIRST_NAME	SECOND_NAME	EMPLOYEE_NUMBER	COUNTRY	JOB_TITLE
Jxkbladdhinfv	******	936042276038	Germany	Court Reporter
Xwlqsxcrfcsjp	******	444607627894	United Kingdom	Marketing Research Analyst
Pkvxqxbrtcajn	******	5484484031640	Germany	Project Manager
Msduoqilcgext	******	6872587489830	United States	Copywriter
Xqmzuogzjemn	******	2201799735622	Germany	Cloud Solutions Architect
Lmonphcbeepzc	******	7088887775853	Germany	Quality Control Analyst
Zrvxqmyvaadga	******	3309725977403	Germany	Marketing Research Analyst
Ypmxmpujkmbad	******	332215463410	France	Web Developer
Bpwjiqccyvi	******	569455264405	France	Cybersecurity Analyst
Bpvcxmtnbcqw	******	2132995421123	United Kingdom	Software Developer
Nuzzrodlezafl	******	3854298161402	Germany	Media Buyer
Xpiczswgxrnnl	******	6649559732424	United Kingdom	Civil Engineer

Informatica

CLAIRE

Intelligent Continuous Tuning

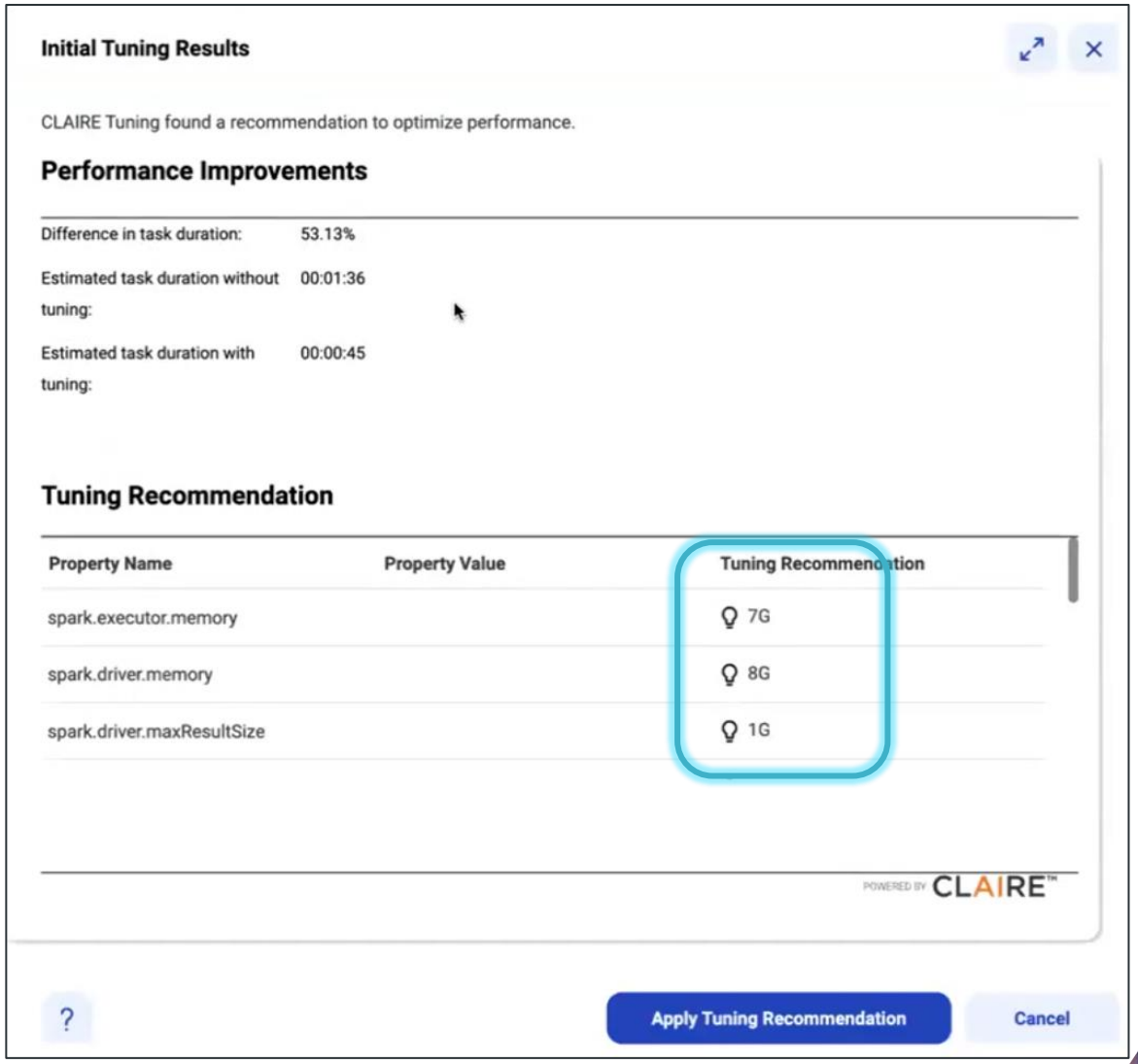
Discover | Build | Curate | Govern | **Operate** | Monitor

Key Highlights

- ML algorithms to learn from previous executions on how to optimize configuration to deliver better performance
- Tuning recommendation for the set of Spark properties that optimizes task performance
- Silent monitoring on job executions to adjust the Spark properties over time

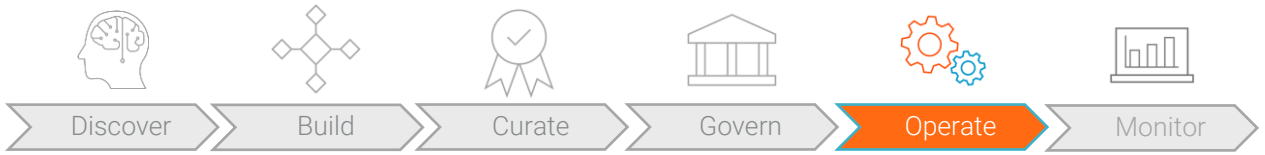
Benefits

- Operators are provided with Spark properties (which can be very difficult to govern and optimize as there are many of them) for each run to fine tune in order to boost performance
- Reduced burden on developers to build optimal jobs (no more trial and error)



CLAIRE

Smart Engine Optimizer

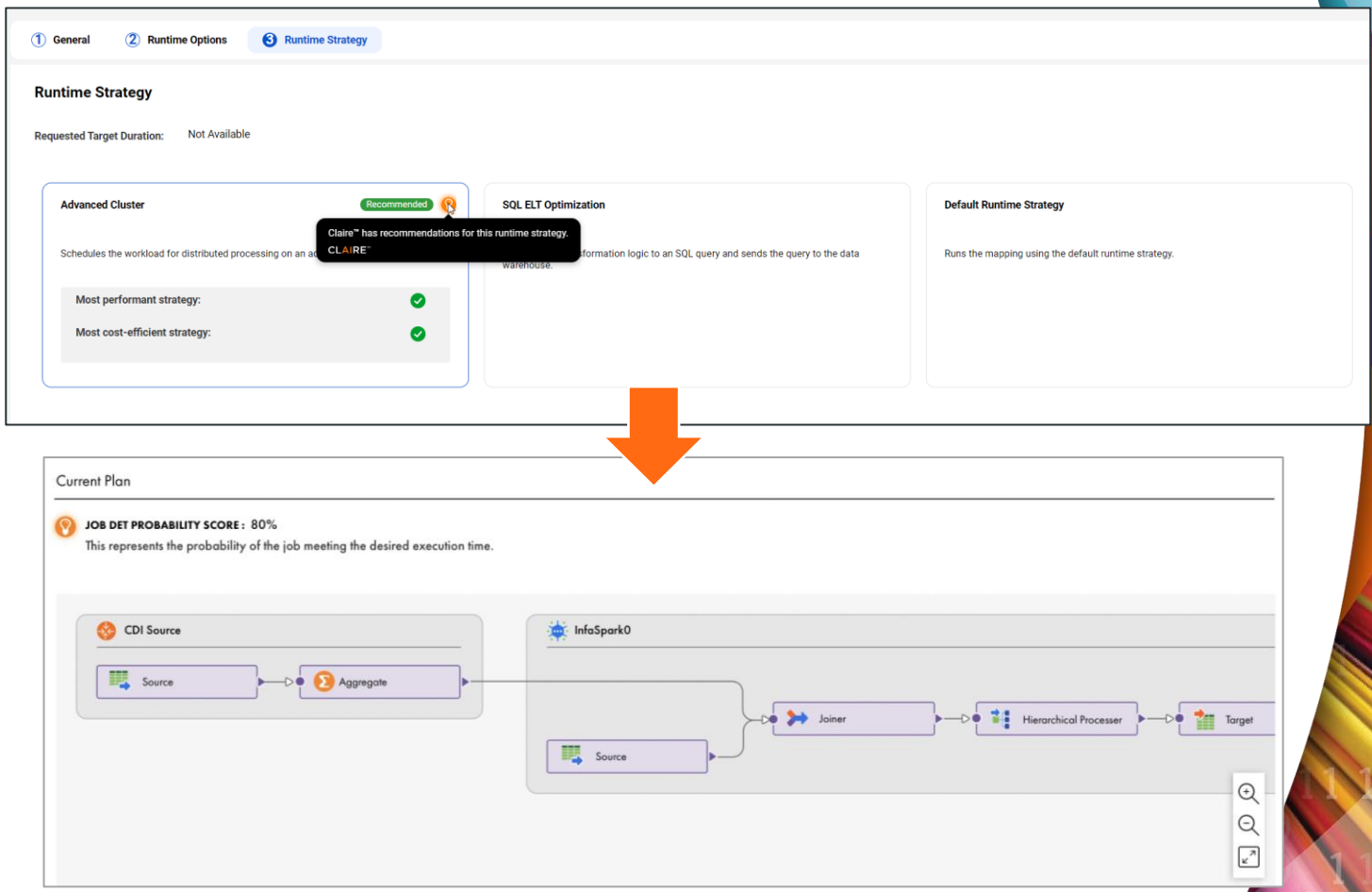


Key Highlights

- Runtime analysis of job logic, using the most optimal engine for each section of the process
- Execution using Spark, SQL ELT and native engines
- Optimization on performance or cost

Benefits

- Optimal execution leveraging multiple engines
- Abstract developers from runtime, accelerating time-to-market and improving productivity



CLAIRE

Data Observability

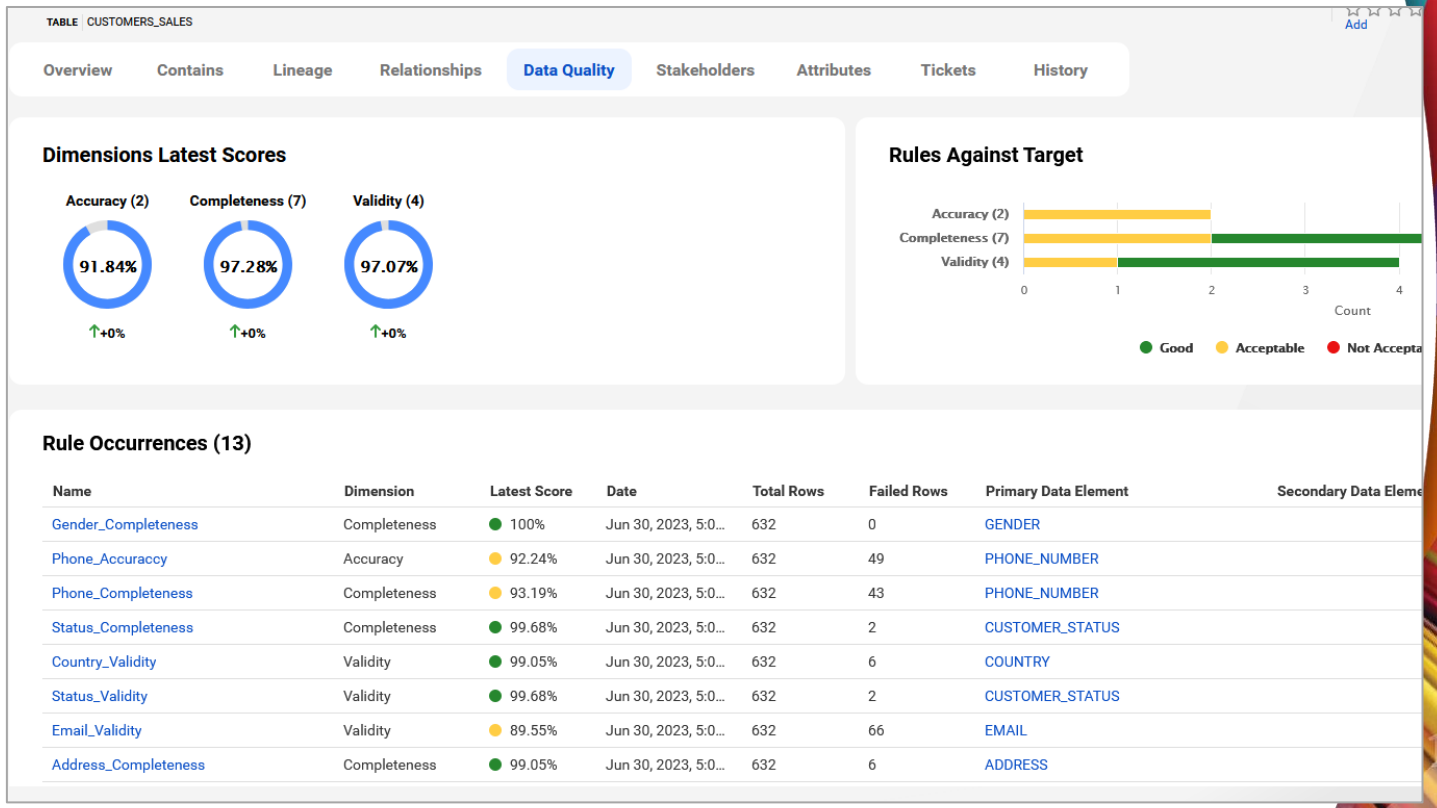
Discover | Build | Curate | Govern | Operate | **Monitor**

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 many hours of work
- Provides a more complete Data Quality view



Name	Dimension	Latest Score	Date	Total Rows	Failed Rows	Primary Data Element	Secondary Data Eleme
Gender_Completeness	Completeness	100%	Jun 30, 2023, 5:0...	632	0	GENDER	
Phone_Accuracy	Accuracy	92.24%	Jun 30, 2023, 5:0...	632	49	PHONE_NUMBER	
Phone_Completeness	Completeness	93.19%	Jun 30, 2023, 5:0...	632	43	PHONE_NUMBER	
Status_Completeness	Completeness	99.68%	Jun 30, 2023, 5:0...	632	2	CUSTOMER_STATUS	
Country_Validity	Validity	99.05%	Jun 30, 2023, 5:0...	632	6	COUNTRY	
Status_Validity	Validity	99.68%	Jun 30, 2023, 5:0...	632	2	CUSTOMER_STATUS	
Email_Validity	Validity	89.55%	Jun 30, 2023, 5:0...	632	66	EMAIL	
Address_Completeness	Completeness	99.05%	Jun 30, 2023, 5:0...	632	6	ADDRESS	

CLAIRE

Anomalies detection and alerting

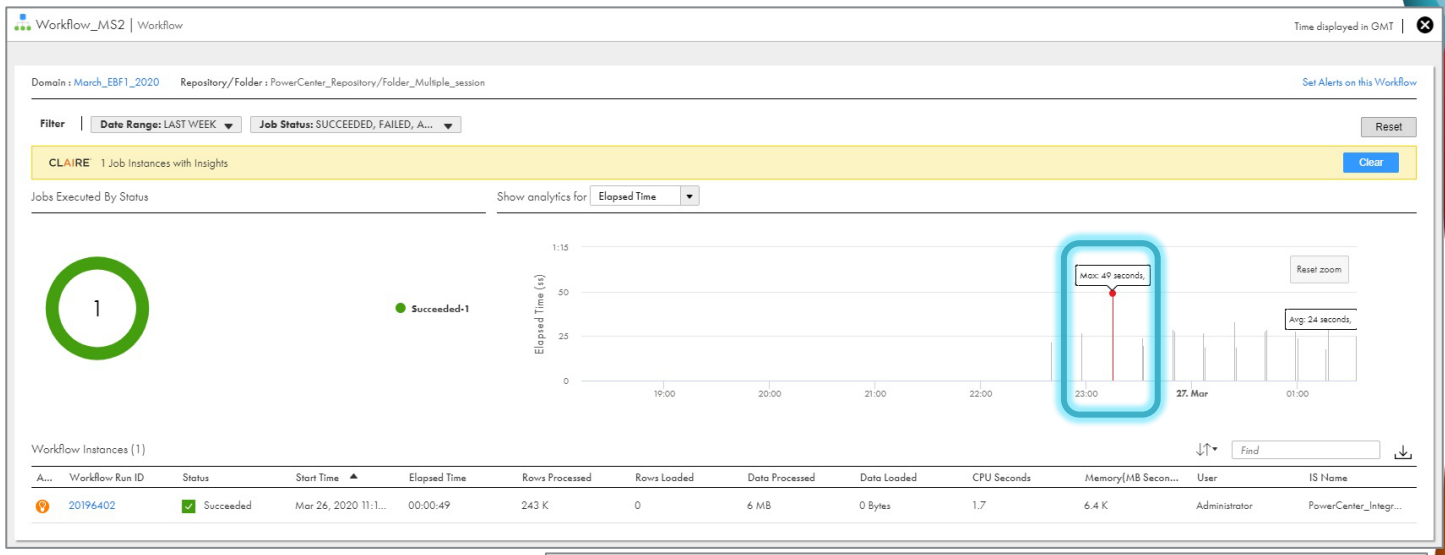
Discover | Build | Curate | Govern | Operate | Monitor

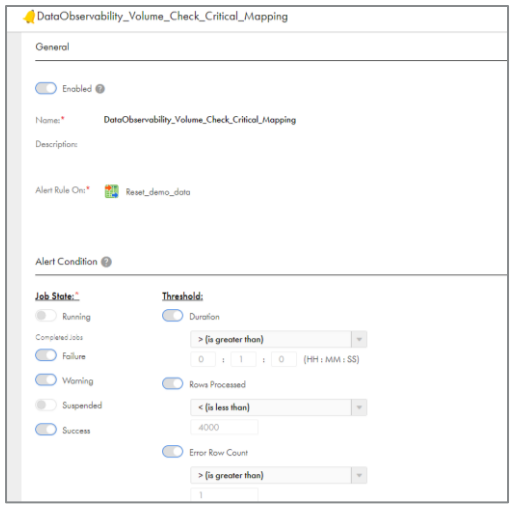
Key Highlights

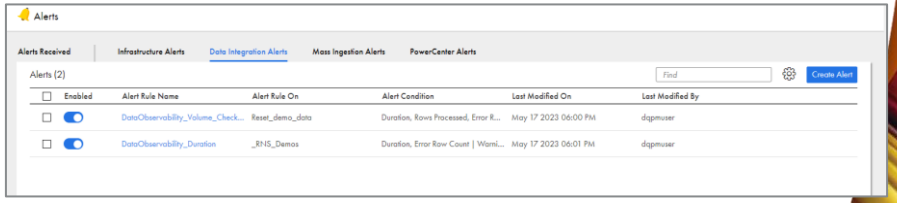
- Observe Infrastructure for Critical Jobs
- Connection observability
- CLAIRE powered self heal, Auto Tune, auto scale, smart shutdown
- Proactive detection of anomalies on job executions (idle time, long running jobs, # of rows processed…)

Benefits

- Proactive detection of potential issues
- Dramatic reduction on time spent troubleshooting and finding root causes



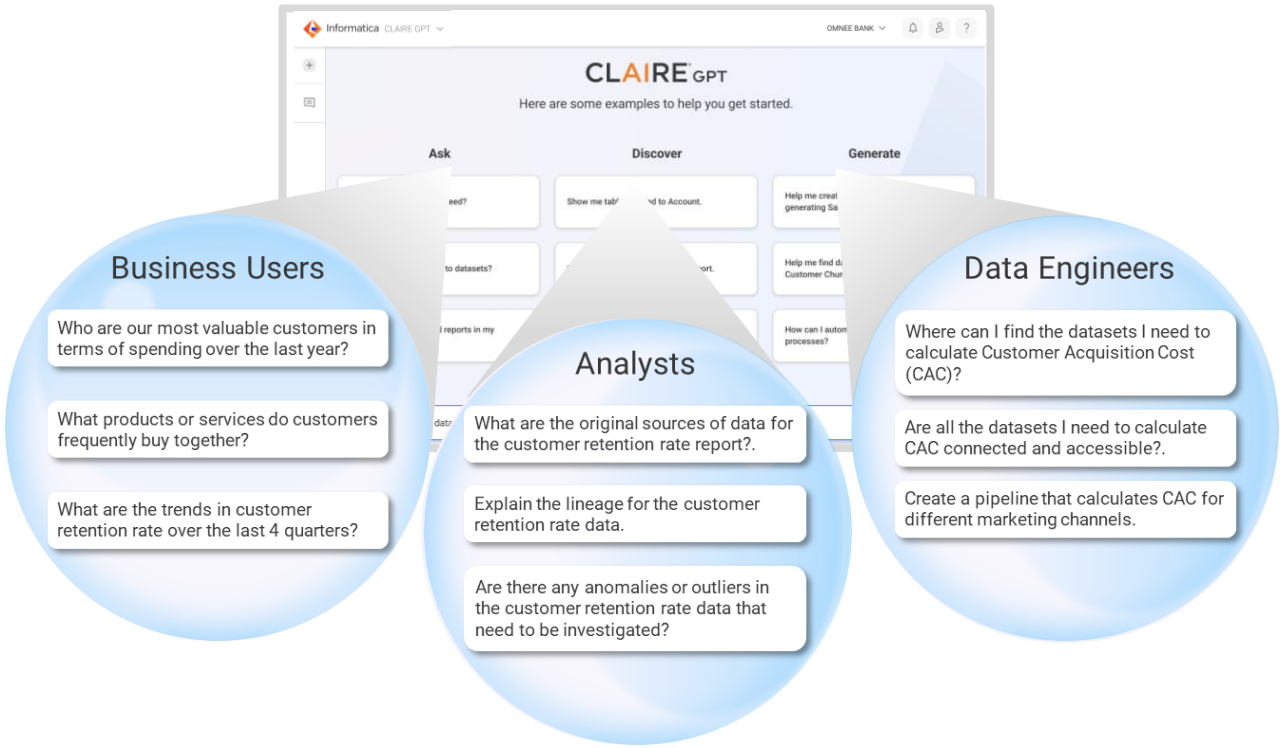




Informatica

CLAIRE® GPT Capabilities

Empowering Customers To Talk To Their Data Using Natural Language



CLAIRE® GPT

NATURAL LANGUAGE INTERFACE TO DATA

Generative AI-powered Data Management

Enhance productivity of experienced data management professionals with fully automated workflows	Reduce data management costs	Data democratization and management by empowering business users to create specifications and perform basic data management independently.

NATURAL LANGUAGE TO ETL/ELT	NATURAL LANGUAGE TO DATA QUALITY	NATURAL LANGUAGE TO PREPARATION	NATURAL LANGUAGE TO GOVERNANCE
NATURAL LANGUAGE BASED DISCOVERY	NATURAL LANGUAGE BASED EXPLORATION	NATURAL LANGUAGE BASED TESTING AND DOCUMENTATION	AUTOMATED FINANCIAL OPERATIONS (FINOPS)
CROSS PRODUCT EXPERIENCE	ARTIFICIAL INTELLIGENCE (AI) COPILOT	IN-CONTEXT DATA INTELLIGENCE	OPEN, POLYGLOT

Informatica

DEMO

Where data & AI come to LIFE


Informatica
Thank You
Where data
& AI come to LIFE

Where data & AI come to

Informatica