

TREDEENCE

 Microsoft  
Solutions Partner  
Data & AI  
Azure

Sancus →

# AI/ML-driven Data Quality Management (Sancus) Solution by Tredence

## Solution Deck

#BeyondPossible



# Tredence is a **Data & AI** consulting firm, driving value realization by enabling the **last mile adoption of insights**

**2000+** AI & Tech specialists supporting businesses across functions by delivering contextualized, high impact solutions

**50%** Accelerator led solutions that **improve time to value by +50%** vs. traditional consulting and technology service companies

**94%** **Best in class 94% NPS** customer satisfaction score driven by our laser focus on empowering decisions to help our clients win

 **6 of the top 10 Retailers and CPGs in the World Trust Tredence to Power their AI Strategy**



“Leader” in customer analytics



Market guide for data analytics & intelligence platforms in supply chain



Certified as Great Place to Work by GPTW India



Fastest growing US companies 6 years in a row

**Servicing industry leading clients** across verticals from 14 offices across North America, Europe, Asia, India and Australia with a Center of Excellence in Chicago

## Customers

Retail & CPG	Industrial	Telecom & tech	BFSI & publishing

# Tredence delivers value through full-stack capabilities & white box solutions that drive the last mile adoption of insights

## Data Engineering



Data Cleansing & Harmonization

Technology Platform Modernization

Cloud Performance & Cost Optimization



databricks



## Machine Learning & AI Enabled Analytics



### Supply Chain & Procurement

Demand Forecasting & Planning

Supply Chain Control Tower

Cost Negotiations Platform

Out of Stock Alerting



### Merchandising Strategies

Price and Promotion Optimization

Assortment and Space Management

Store Location Optimizer

Next Gen Product Recommender



### Marketing & Personalization

Personalization Engine

Test & Learn Platform

Marketing Mix Modeling & MTA



### Customer Analytics & Experience

Unified Customer Data Platform

Customer Explorer Visualizations

Customer Journey Mapping

Customer Behavior Modeling



### Cost Optimization & Productivity

HR Analytics Platform

Store labor optimization

Associate training management

## DATA SCIENCE R&D CENTER



Computer Vision



Internet of Things



Natural Language Processing



Proprietary Frameworks



Edge Deployment Capabilities

## Last Mile Adoption



### ML WORKS

White box model deployment and monitoring, to ensure complete transparency and explainability of models

### TREDENCE STUDIO

Custom Application development & deployment to deliver last mile insights & impact

# AI/ML based data quality tool that aims to deliver reliable data to your business

## Data Validation

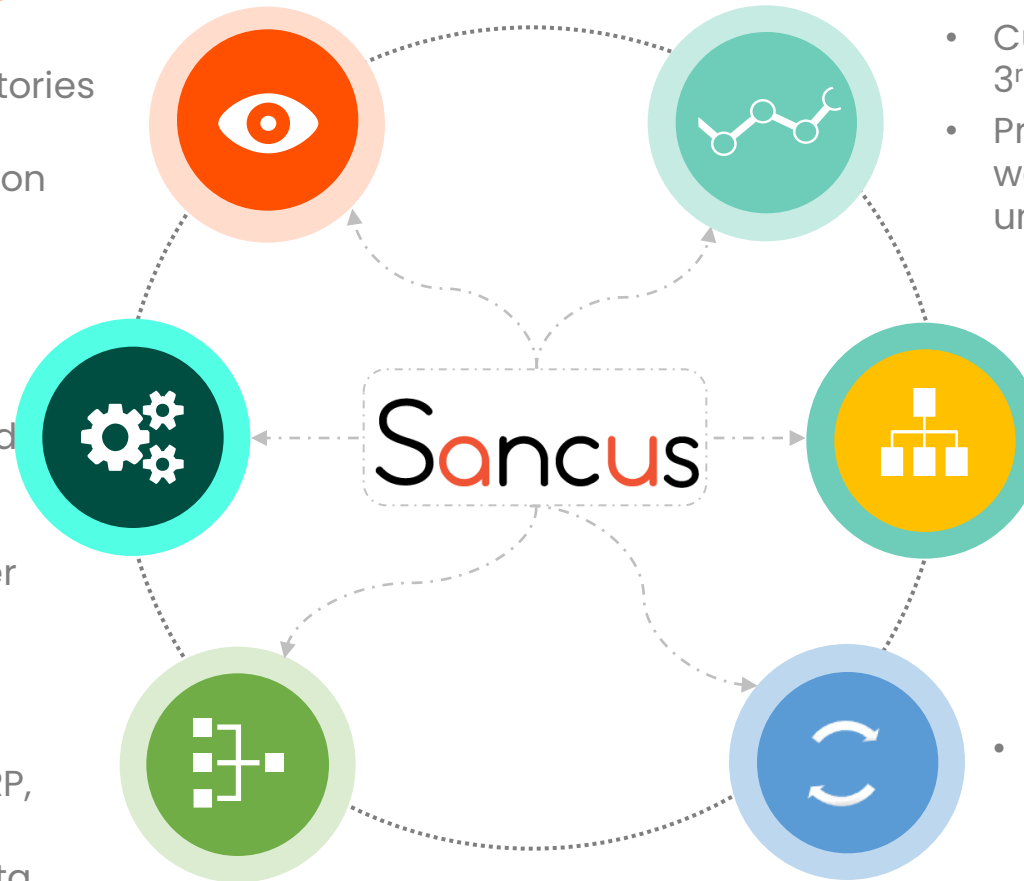
- Global address validation & correction using postal directories and 3<sup>rd</sup> party APIs
- Contact/lead email verification

## Data Cleansing

- Business rule configuration, Standardization, De-dupe using **AI/ML algorithm** & golden record creation
- Deployed on customer, contact, product, material, vendor master data

## Data Ingestion

- Integrates seamlessly with tools/applications like Oracle ERP, Salesforce, etc.
- Leverage OCR tools to scrap data from documents to be consumed by downstream layers for Cleansing



## Data Enrichment

- Customer/contact enrichment through 3<sup>rd</sup> party partnerships
- Product/material enrichment through web scraping, image processing, unstructured data analysis

## Hierarchy Management

- Customer/vendor hierarchy validation, creation and management through 3<sup>rd</sup> party partnerships
- **ML algorithms** for product/material hierarchy assignment at scale (including GPC/GSI migration)

## Feedback Loop

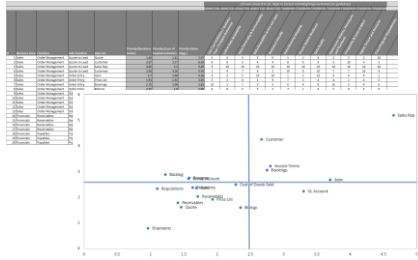
- **Active Learning** based feedback module allows business users to pass feedback and override the algorithm's results
- Enables a reinforcement loop to keep the machine learning models accurate with time

# Tredence Differentiation: We have structured process templates, frameworks and runbooks

1

## Data Assets Classification Framework

We have developed a framework to classify data assets into different categories based on different parameters – scope of usage, sensitivity, compliance, functional impact and type of data. Our classification framework helps to establish policies based on the parameters captured against individual entities.

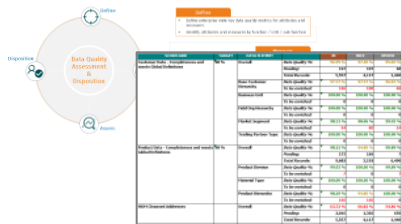


2

## Data Quality Scorecards & Dashboards

At Tredence, we built a AI based Data Quality solution to master and manage hierarchy data. As part of the data quality solution, we have delivered comprehensive quality scorecards that analysts, data engineers and data domain owners can leverage to measure the quality and take appropriate actions to continually improve the quality.

Data Quality Prioritization Methodology



3

## Standardised Operational Metadata for GCP Pipelines

Having done a number of GCP data lake implementations, we have standardised the data models to capture operational metadata in data pipelines that can be used for measuring the key operational metrics over time - volume growth, failures / errors, performance etc.



### Operational Metadata Definition

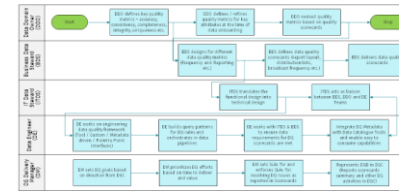
- Job Master
- Batch Master
- Batch - Job Bridge
- Job Statistics
- Batch Statistics

4

## Data Governance Policies for Quality and Security

We have pre-defined data quality and security policies that can be adopted by any of our clients with minimal tweaks. Policies are written leveraging our experience implementing data governance for not just retail clients but for some of the more heavily regulated clients as well.

Data Quality Sustainance Model

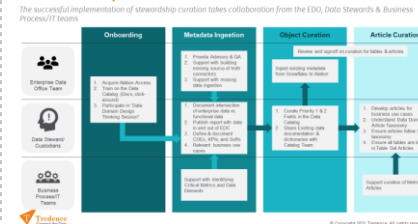


5

## Data Catalogue Operationalization

Tredence has created templates for operationalizing a data-catalog & standardizing Data stewardship across enterprise Organizations. Our framework helps us move from onboarding to meta data ingestion to object curation at double quick-time.

Data Stewardship Process Flow



6

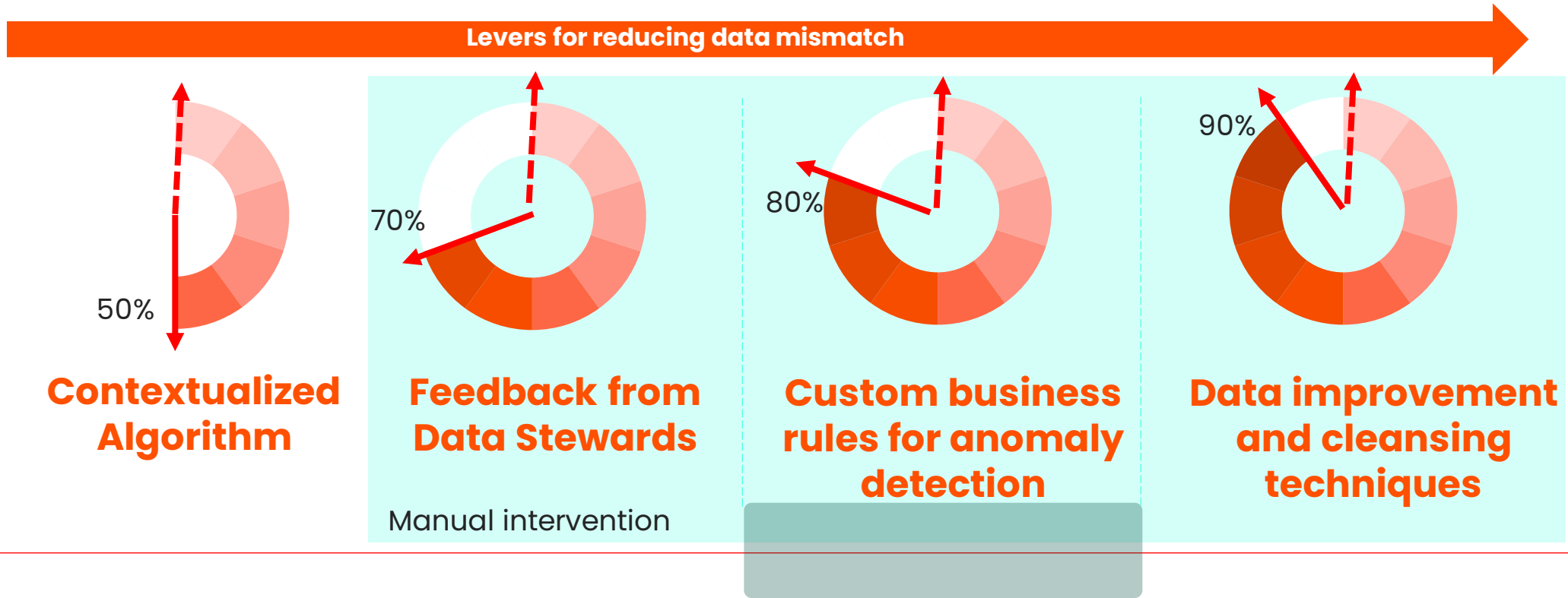
## Taxonomy Definer

We have developed an accelerator that runs through the curated object-metadata and helps to classify tables objects and determine domains. The accelerator uses best in classification techniques & ensemble models to accurately classify the information.

Defining Taxonomy



# Evolution of the solution for a new market



Matching Accuracy

40%

Native Algorithm

50%

Contextualized Algorithm

70%

Feedback from Data Stewards

Manual intervention

80%

Custom business rules for anomaly detection

90%

Data improvement and cleansing techniques

**Note:** Depending on the available fields and level of structured data available in the market, the final matching accuracy for the pilot may vary by +/- 5%. Although, during scaling, cross geography learning available to the algorithm will result in overall improved accuracy

**Sancus team will provide documents explaining step-by-step process of onboarding a new market in detail**

**Success Criteria = 80% to 85% matching accuracy**  
**Subject to level of user interaction and feedback**

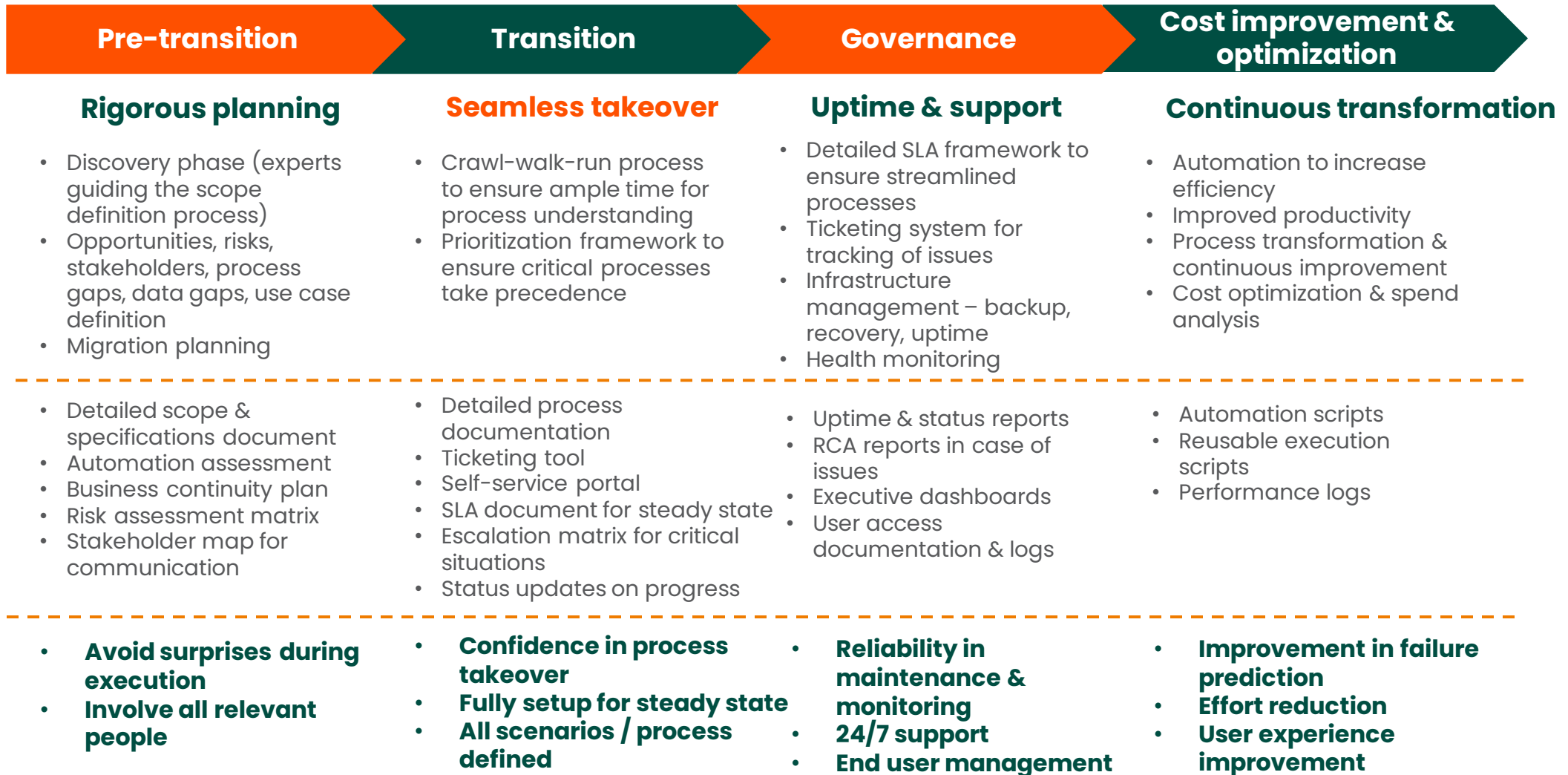
# Expected Deliverables through the engagement



**On-demand Consulting, Guidance and Product Enhancements**

# We follow a robust process to ensure smooth **process transition and maintenance**

Maintenance cost depends on the level of involvement expected from the solution support team





# Sancus – Our AI/ML based data cleansing & consolidation solution has a **proven track record of success**



**Standardized & Consolidated Customer Master data for a leading CPG company**

- Standardized data across **40+** sources across **7** countries
- Improved the customer-channel mapping by **26%** and created various hierarchies for Pricing, Local Reporting, Global, etc.

**Standardized PO receipts for a large CPG company**

- Realized **200k** of receivables per year by consolidating POC receipts across partner and retailer source systems
- Reduce PO reconciliation time from **2+ weeks to 2 days**

**Cleansed and created a customer master for a leading automobile manufacturer**

- Reduced time for master data creation by **50%** over the existing solution
- Reduced Marketing campaign costs due to duplicate mails sent to the same customers
- Enriched customer address data through third party APIs

# Automation of Purchase Order to Invoice Reconciliation



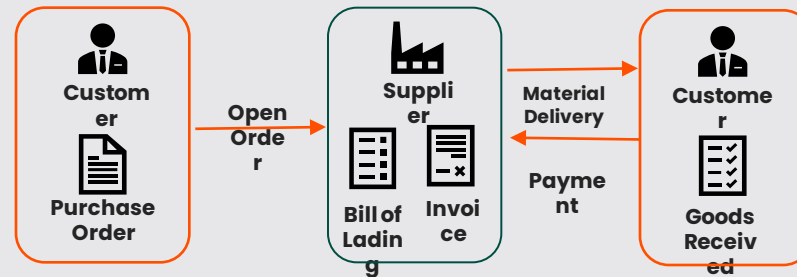
## Situation

- Purchase order to invoice three-way match is transaction reconciliation process of approving payment to a supplier upon completion of an order
- The manual process of reconciliation requires to check products, quantities and costs across purchase order, bill of lading, invoices and contract/MSA
- Reconciling the documents from multiple sources in different formats for transactions takes valuable time and resources



## Need

- Objective is to reduce margin leakages caused by incorrect or inefficient PO reconciliation.
- Reduce scope of human error and eliminate nonvalue added activities
- Automate the reconciliation process of POs against the invoices for payment processing



## Proposed Solution

- **Data Gathering & Cleansing:**  
Use image processing techniques to convert PDF and JPEG format Invoices to structured CSV or text format for processing
- **Data Tally:**  
Use Sancus to match invoice details. Validate address anomalies to ensure correct carrier charges
- **Automation of PO Reconciliation:**  
Improve the efficiency and error proof the process by automating PO reconciliation to close POs

# Enabled a large CPG player with Sancus' AI-ML matching & need-based customizations



## BACKGROUND

- The CPG Client Food Solutions is looking to enhance the data quality of their Operator and Contact data objects to enable downstream analytics and BI.
- Tredence's ML based matching engine is used to cleanse and consolidate Operator and Contact data, identifying and flagging duplicates and creating a golden record that is propagated forward



## BUSINESS NEED

Client wanted an intelligent automated solution to tackle multiple Data Quality challenges that exists, including duplicate and inconsistent data entries. Scope – 75 markets; 12-17 objects per market (including Operator and Contacts)

## ★ SOLUTION APPROACH

Tredence's solution created immediate ROI for the client

- Design, build and deliver data pipelines for 75 markets
- Implement data quality, measure and optimize quality for 15 objects for each of the markets
- Read Operator and Contact datasets from UDL ,perform DQ check , process and harmonize the datasets for Sancus consumption
- Implement better data profiling, quality and enrichment in the pipeline to improve adoption
- Orchestrate end to end pipelines on ADF. Optimize performance & cost. Operationalize logging across pipelines.



## OUTCOME

Sancus acts on **100%** of the source data and adapts to unique scenarios within each market  
**21%** increase in accuracy over existing engine

Results provided within **24 hours** of dataset being made available

Proven the effectiveness of Sancus' AI ML matching model for scaling up across all **75 markets** with need-based customizations





**#beyondpossible**