CASE STUDY

Data Deduplication for a Health & Fitness Club for Accurate & Reliable Business Insights

CONTATA

OVERVIEW

The client is a prominent health and fitness club franchise operating across 50 countries in over 5,000 locations.

CHALLENGES

The client sought an application to enhance the accuracy of matching existing customers from the database and generate a unique identifier, the golden key, for each. Managing and safeguarding data required meticulous planning and collaboration between teams to ensure seamless transitions between stages. Additionally, obtaining a list of customers was a big challenge, emphasizing the significance of understanding the users for a customer-centric organization.

INDUSTRY

- Fitness
- Franchise

SERVICES

- Data Warehouse
- Reporting/Self-Service BI
- Data Discovery
- Deduplication
- Data Governance

TECHNOLOGY

- Azure SQL
- C#
- Azure Data Factory
- Power BI
- Data Lake
- Service Bus
- Azure Alerts

SOLUTION

Contata addressed the challenges by thoroughly analyzing the client's process of handling "dirty data" including unclear addresses, transliteration issues, contact information complications, and frequently repeated values.

Our solution involved creating an application with a deduplication approach, preceded by data normalization. This normalization process ensures that the data is structured, aligned, and consistently formatted, considering factors such as lowercase, abbreviation, missing values, address standardization, and validation of phone numbers and email addresses.

The deduplication process begins by transferring client data to a separate database responsible for generating Golden Keys. As part of pre-processing, checksum columns are created by combining various factors such as:
- Name and Address
- Name and Email
- Name and Phone
- Name and Date of Birth.

These checksum columns are then used to join the data with the Golden Key tables. If a record matches specific criteria, it is assigned an existing Golden Key; otherwise, a new Golden Key is generated and assigned to the record.

BENEFITS

- Reduced data duplication significantly.
- More structured representation of data for the user.
- Provision of a separate identity in the form of a Golden Key.
- In-depth business insights that enhance transparency in the process.

About Contata

Contata Solutions is a trusted leader in technology and digital innovation. Through our work in data engineering, data analytics, machine learning, marketing automation and app development, we deliver solutions that address complex problems in ways that are simple, insightful and impactful.

Our promise and value proposition to our customers is simple: we leverage our deep technical expertise and global presence to bring software products and data-driven decision capabilities to life.

Founded in 2000, Contata is a privately-held company headquartered in Minneapolis that serves clients globally from offices in the United States and India.