

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

Introduction

The client is a leading health & fitness club with a worldwide presence having 5000+ franchised locations across different continents and almost 50+ countries including countries like the USA, Canada, Belgium, Netherlands, Poland, Luxembourg, UK, Qatar, India, Singapore, Australia, Japan & more.

Need

Essentially our client wants an application that can enhance the accuracy of MATCH between existing customers from the database and generate the golden key for them. Successful management and protection of data require a lot of planning and collaboration between teams to move data from one stage to the next. When it comes to most customer-centric organizations it will be the biggest challenge to get a unique list of customers not because of storage cost reduction, but rather since understanding their customers is the biggest asset for them.

Solution

Contata accepts the challenge and analyzes the whole process of dirty data comprising unclear addresses, transliteration problems, contact information complications, most repeated values, etc. After all, contata creates the application by the approach of Deduplication. But before going to the process the concept of data normalization included through which the data is structured, aligned, and has a consistent format comprising all factors such as lowercase, abbreviation, missing values, address standardization, validation of phone numbers and email address. Now the process of Deduplication starts with moving the client data onto a separate database which is responsible for generating the Golden Keys. As part of pre-processing, checksum columns are created with a combination of :

- Name and Address
- Name and Email
- Name and Phone
- Name and Date of Birth

The data is then joined with the golden key tables based on these checksum columns. If a record matches particular records, it is assigned a golden key, otherwise, a new golden key would be generated and assigned to the record.

Advantage

- Data duplication is minimized at maximum scale.
- More structured representation of data to a user.
- Separate identity in the form of a Golden key.
- In-depth business insights that bring transparency to the process.

Industry

Fitness

Franchise

Our Services

Data warehouse

Reporting/Self-Service BI

Data Discovery

De-duplication

Data Governance

Technology Stack

Azure SQL

C#

Azure Data Factory

Power BI

Data Lake

Service Bus

Azure alerts