

DE-DUPLICATION ENGINE

Streamlined solution to ensure centralized de-duplication of doctor's database

CLIENT

The client is one of the fastest growing pharmaceutical companies in India. The company owns over 400 pharmaceutical brands; a variety of nutritional products, medical devices and a range of diagnostics solutions.

CHALLENGE

As an integral part of its Sales Force Management, all Pharmaceutical companies collate doctor information in a doctor database. However, since each division creates its own doctor's records, there are

multiple records created in the doctor's database for a number of doctors.

There was a need to clean up this data by removing such duplicates so that every doctor has a single record in the

doctors' database.

The objective was to get accurate reports on doctor spend and resulting business from those doctors across different divisions.

SOLUTION

MindCraft's De-duplication Engine was used to clean duplicate information in the database, resulting in accurate information for any given record.

- The full dump of doctor's database was provided by the company in a pre-defined format. The engine segregated this data based on District / Taluka level. For eg. West Zone > Maharashtra > Thane.
- Processing for de-dup was run on the segregated data in batches and the segregated data was processed for duplicates.
- Fields such as First Name, Last

Name, DOB, Registration Number, Zip Code were assigned as key values for finding duplicates.

- The filtered data was processed again for duplicates using more specific criteria such as Mobile Number, Email ID etc.
- A weighted score was calculated for every record to help identify a primary record from the duplicates found.
- The list of duplicates was displayed in a web based front end application.
- A merge option was provided which allowed users to select the entries to be merged.

- One entry was created based on the most current data of the selected entries. The new entry was assigned with a new Doctor ID.
- The mapping of old Doctor ID to the new Doctor ID was made available for future reference.
- Doctor to division assignment (Orthopedic, Cardiologist, and Pediatrician) was also mapped to the New Doctor ID in a separate table in the system.

BENEFITS



Consistency



Minimal Errors



Reliable Reports



Operational Efficiency

FEATURES



- Single view of doctor's data across the organization
- Tracking of spend on customers/ agents
- Fewer information errors, leading to improved sales and reduced losses
- Multichannel initiatives built upon a common, trusted source of information
- Access to accurate information

CONCLUSION

Data de-duplication or single instancing essentially ensures elimination of redundant data. In the de-duplication process, duplicate data is deleted, leaving only one copy, that is the single instance of the data leading to a single view of doctor's data that is to be stored. This greatly helped the client to bring about operational efficiency. The client could generate accurate reports by spending less time in manually deleting duplicates.

ABOUT US



MindCraft is a Software Products, Solutions and Services organization. We enable customers to transform their businesses through products for Banking, Insurance, Mutual Fund & Financial Services and a cross industry product suite.

We help our customers get competitive edge through technology services that include Development & Integration, Business Intelligence & Analytics and Systems Software Management. Incorporated in India in 2002, we have a team of over 500. We also have a presence in Singapore, USA & UAE.

CONTACT US



+91 22 4221 6200

www.mindcraft.in



+65 6652 3215

www.mindcraft.com.sg



+1 (908) 750 6647

www.mindcraftamerica.com



+91 22 4221 6200

www.adeptus-tech.com