

MANAGEMENT OF ORGANIZATIONAL DATA

Group Number 7

Client: ASP OL Media

Team Members:

Alexander Scarcelli

Archita Ray

Ashvin Raj

Vamsi Krishna Seemakurthi

Priyam Sarkar

Sai Mona Duvvapu

Background and Data Set Overview

The client we serve is ASP OL Media, a marketing company that provides email services to Forbes 100 companies. We assist these companies in their marketing campaigns by sending a promotional email campaign on their behalf. ASP OL Media boasts a workforce of over 170 professionals and maintains a presence in multiple locations across India. On a daily basis, our client is tasked with sending out more than 8 million emails. This necessitates the implementation of meticulous assessments and stringent quality control measures to ensure the effectiveness of their marketing campaigns.

We initially had 8 tables, we created 2 associative tables and 1 more table after normalization

Our primary data set comprises of 8 tables, each with its own unique description and purpose. Here's an overview of these tables:

1. Customers Table: This table contains information about the customers, including their customer_id and email_id. There are two categories of customers: active customers and unsubscribed customers, each of which has a separate table.
2. Active Customers Table: This table stores information about customers who have active emails.
3. Unsubscribed Customers Table: This table stores information about customers who are not active or have unsubscribed.
4. Email Customer Table: This table is the associative entity table which has the Customer ID and the Group ID information
5. Email Group Table: This table provides information about companies and their various campaigns associated with a group ID.
6. Campaign Email Table: This table is the associative entity table which has the Campaign ID and the Group ID information
7. Company Campaign Table: This table has the information about the company and the campaigns it is running
8. Offer Details Table: This table contains a unique ID (offer_id) for each offer and corresponding information about that offer. This includes details such as which company is offering it, the offer link, offer type, start and end dates, the number of emails sent, the number successfully sent, and the number of customers who clicked on the offer link.
9. Offer Send Details Table: This table uses the send_id as a primary key, representing a unique key for each company's campaign sent at different times.
10. Offer Email Subject Table: This table uses an auto-incremented ID as a primary key and associates an email subject with each unique ID.

Objective

Our objective in this project is to create and optimize the database management system for the company and derive insights which will help the company improve its operations. This entails rigorous data wrangling and cleaning procedures to ensure data quality and consistency. We then created an insightful Entity-Relationship Diagram (ERD) to visually represent the database's structure and table relationships. Subsequently, we construct a well-defined relational schema, specifying tables, attributes, and data types. For the purpose of optimization, we also applied normalization techniques to minimize redundancy and improve efficiency. To maintain data integrity, we establish appropriate primary and foreign keys, demonstrating a meticulous approach to database management throughout the ASP OL Media project.

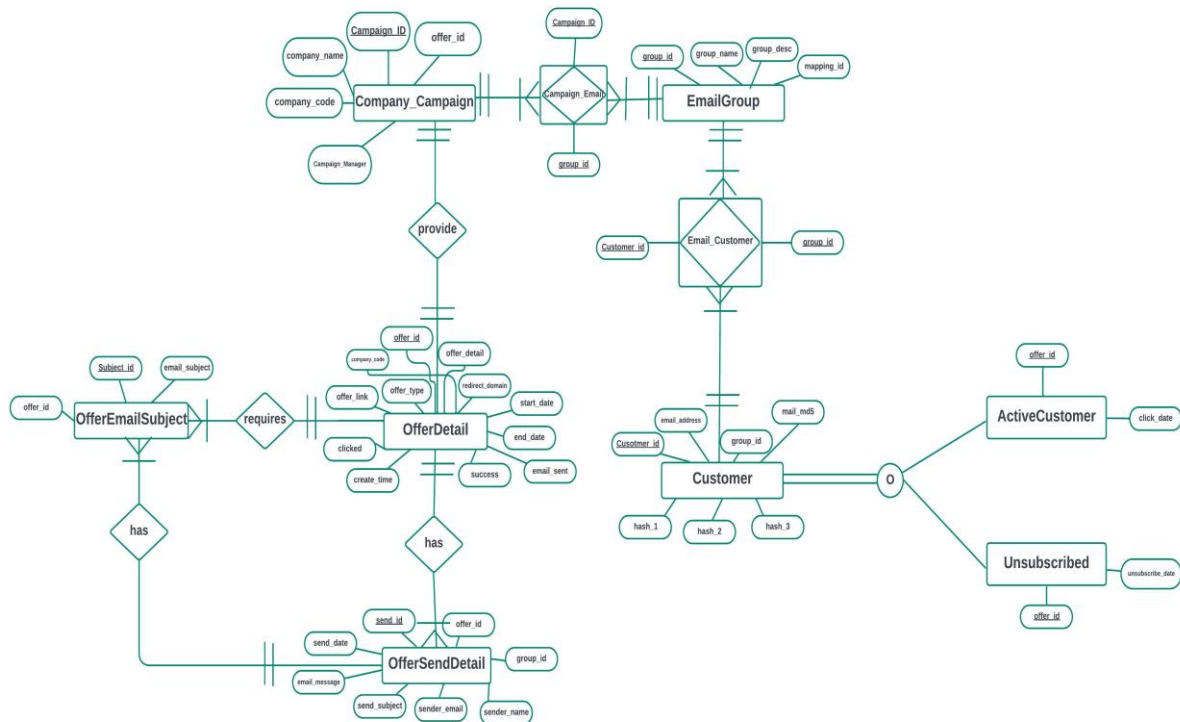
After successfully creating the database, our primary objective shifts towards conducting in-depth data analysis. Our aim is to derive valuable insights from the organized dataset, and we intend to achieve this by utilizing analytical techniques. We are particularly focused on determining the performance of different campaigns, leveraging key metrics such as click rate and the number of unsubscribers. Additionally, we are keen to identify which email subjects resonate most with our audience and garner the highest traction. This data-driven approach will not only allow us to gauge the effectiveness of our marketing efforts but also inform future campaign strategies and optimizations to further enhance the ASP OL Media project's success.

Approach

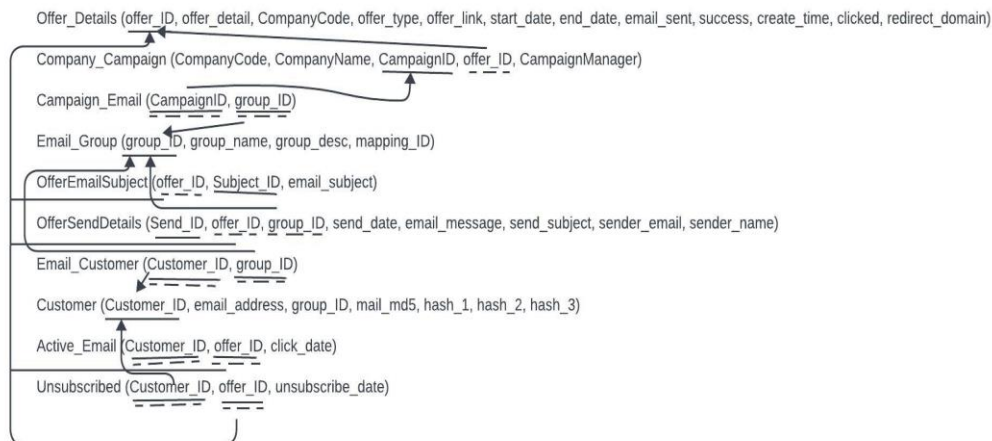
We began our work with csv files from the company, which contained both table creation statements and insert statements necessary for data retrieval. We extracted the data from these SQL tables, carefully interpreted the information they conveyed, and embarked on a data cleansing process. This was particularly crucial due to some tables featuring recurring column names and duplicate values. Moreover, we noticed that not all tables in the existing database had primary keys or foreign keys defined, so we undertook the task of creating and establishing these key relationships, resulting in a more organized database structure.

With the Entity-Relationship Diagram (ERD) and relational schema successfully created, we proceeded to import clean CSV files into the MySQL server. Following the data import, we diligently assigned primary and foreign keys to each identifier attribute within the tables, further enhancing data integrity. Additionally, we introduced two additional tables for managing associative entities, namely "campaign_email" and "email_customer".

ERD Diagram



Relational Schema



Description of ERD and Relational Schema

The Company Campaign table has Campaign ID as the Primary Key. The other attributes are (CompanyCode, CompanyName, CampaignID, offer_ID, CampaignManager). This table has information about each campaign associated to each company and also has the offer id attribute which

is a foreign key from the offers detail table. Each Campaign is related to one offer id and each offer id is related to one campaign id.

The Offer Details table has Offer ID as the Primary Key. The other attributes are (offer_ID, offer_detail, CompanyCode, offer_type, offer_link, start_date, end_date, email_sent, success, create_time, clicked, redirect_domain). Each Offer ID can have multiple Send IDs and each Send ID can have only one Offer ID.

SendID is the primary key of the OfferSendDetails table. This table also has Offer ID from the Offer Details table as the foreign key and Group ID also as the foreign key from the Email Group Table. The other attributes of these table are (Send_ID, offer_ID, group_ID, send_date, email_message, send_subject, sender_email, sender_name)

Offer Email Subject table has Subject_ID as the primary key. The other attributes are (offer_ID, Subject_ID, email_subject)

Campaign_Email table is the associative entity between the company campaign table and email group table. This table has the following attributes (CampaignID, group_ID)

Email Group table has Group ID as the primary key. The other attributes are (group_ID, group_name, group_desc, mapping_ID). Each email group is associated to many to one customer and each customer can be attributed to many to one group_id.

This makes an associative relationship between these two entities represented in the Email_Customer table. This table has the following attributes (Customer_ID, group_ID)

Customer table has the primary key as Customer_ID. The other attributes in this table are (Customer_ID, email_address, group_ID, mail_md5, hash_1, hash_2, hash_3)

Customers can be of two types, Active Customers or Unsubscribed Customers. There exists an overlap of relationship between active and unsubscribed customers explained by the “O” in the ERD diagram. Such a relation exists because one customer could be active for one offer but may choose to unsubscribe from another offer.

The attributes of Active Customer are (Customer_ID, offer_ID, click_date)

The attributes of Unsubscribed Customer are (Customer_ID, offer_ID, unsubscribe_date)

Normalization

Normalization plays a crucial role in database design and management by helping to improve data integrity, reduce data redundancy, and make the database more efficient. After working through 1 NF, 2 NF and 3 NF, we have the following normalized form.

offer_id → offer_detail, company_code, offer_type, offer_link, start_date, end_date,
email_sent, success, create_time, clicked, redirect_domain

company_code → company_name, Campaign_Manager

Campaign_ID → offer_id, CompanyCode

CampaignID, group_Id →

group_Id → group_name, group_desc, mapping_ID

Subject_ID → offer_id, email_subject

Send_ID → offer_id, group_Id, send_date, email_message, Subject_ID, sender_email,
sender_name

Customer_Id, group_Id →

Customer_Id → email_address, group_Id, mail_md5, hash_1, hash_2, hash_3)

Customer_Id, offer_id → click_date

Customer_Id, offer_id → unsubscribe_date

Prior to the normalization, company and campaign tables were clubbed and as part of removing the partial dependency, we have split the table into two separate ones. 'OfferSendDetail' table has send_subject unrelated to the email_subject in "OfferEmailSubject" though both are essentially the same. Hence, to remove the redundancy and improve data integrity, we have included Subject_ID in the "OfferSendDetail" table.

From the derived tables, we performed analysis to determine the most influential variables leading to a successful advertisement campaign. Based on these insights, we developed some key recommendations that could improve campaign performance.