



PROJECT ASSIGNMENT



April 2015 Dmytro Marushkevych/Troy Gatchell ROSETTA

Background

Rosetta is an agency focused on Customer Engagement. We connect rich data, engaging experiences and robust technologies to create meaningful customer relationships that drive business impact.

Being a part of Publicis.Sapient and Razorfish Global, Rosetta is helping major brands across virtually every industry to understand their customers and translate the intimate customer knowledge into award winning and highly effective Customer Engagement campaigns.

Rosetta has partnered with Cal Poly Computer Sciences department to help students get the hands-on experience with the database design and use in the business environment.

Data

Note: this is a hypothetical assignment and is not representative of any actual Rosetta's client. No real client data is provided as part of this assignment.

BrandX (not a real company) is an electronics manufacturer, specializing in mobile devices, such as smartphone and tablets. To drive customer engagement and retention, BrandX has created a database of their customers and their historical device purchases.

BrandX has hired a marketing agency to develop, run and optimize their email marketing program.

You are a Senior Associate at the agency's Tech group and you are tasked with developing a database to hold the account, device registration and campaign engagement data received from BrandX and their Email Service Provider (company that manages the physical deployment of the emails).

As a part of the exercise, you will receive 4 separate data sets. None of the datasets contains any Personally Identifiable Information (PII). Customer names and email addresses are kept on the Email Service Provider (ESP) side.

1. **Account Registration Data.**

This dataset contains information about the BrandX customer account registrations.

Major fields included:

Field	Description
Customer ID	Unique customer identifier
Email ID	Unique identifier of the customer email address (stored by ESP)
Registration Source ID	Identifier of the account registration source, e.g. web site or Call Center.
Registration Source Name	Name of the registration source
ZIP	
State	
Gender	
Income Level	Customer income level
Permission	Optin/optout indicator for marketing communications
Language	
Registration Date	
Email Domain	Domain of the customer email, e.g. gmail.com
Customer Tier	Tier customer has been assigned to by BrandX

2. **Device Registration Data.**

This dataset contains the information about BrandX devices purchased over time.

Major fields included:

Field	Description
Customer ID	Unique customer identifier
Registration Source ID	Identifier of the account registration source, e.g. web site or Call Center.
Registration Source Name	Name of the registration source
Device Model	
Serial Number	
Purchase Date	
Purchase Store Name	
Purchase Store State	
Purchase Store City	
Ecomm Flag	Indicator if purchased on ecomm site
Registration Date	
Number of Registrations	Number of registrations customer has
Registration ID	Unique registration ID

3. **Email Event Data.**

This dataset contains the information about customer-level email marketing campaign history and engagement: emails received, opened and clicks within the email.

Major fields included:

Field	Description
Customer ID	Unique customer identifier <DEPRECATED>
EmailID	Unique Id of the email address of the recipient of the email message
Audience	Customer segment
Campaign Name	Name of the email campaign
Version	Creative version
Subject Line	Email subject line
Deployment Date	When email has been sent
Deployment ID	Batch id (emails sent in batches)
Event Type ID	IDs for email types: send, open, click, bounce, unsubscribe
Event Type Name	Name of the email type (e.g. send, open, click)
Event Date	When event happened
Link Name	Link alias, e.g. Hero Module or Contact Customer Service
URL	Link URL

4. **Device Data.**

This dataset contains the mapping of device model number to the device name and wireless carrier.

Major fields included:

Field	Description
Device Model	See # 2
Device Name	E.g. 'Constellation 5'
Device Type	Phone vs Tablet
Carrier	AT&T, Verizon, etc.

Assignment

A. Design the database schema

1. For the purposes of this assignment, use MS SQL server or similar.
2. Keep in mind that actual dataset contains millions of records. Database schema should be developed in a way that avoids unnecessary data duplication and maximizes the performance (time to run) of the queries described in section C.
3. The database schema should be done within best practices of warehouse design consisting of star and/or snowflake schema design principals.

B. Create the ETL process for the daily file load (Optional)

1. 3 extract files: Account Registrations, Device Registrations and Email Events (Device categorization data is loaded ad-hoc)
2. Format:pipe delimited
3. Assume files are posted daily on FTP
4. Assume files are not encrypted
5. Need to load the daily files into developed schema.

C. Develop a set of reports

1. **Email Campaign Performance**

This report helps the business users to understand the performance of company's email marketing campaigns and drive future campaign optimization.

Fields to be included:

Field	Description
Campaign Name	Name of the email campaign
Audience	Customer segment
Version	Creative version
Subject Line	Email subject line
Deployment Date	When email has been sent
Unique Emails Delivered	Sum of unique emails sent - Sum of bounced emails
Unique Emails Opened	Sum of unique email opened
Unique Clickers	Sum of unique customers who opened this email
Open Rate	Unique emails opened/Unique emails delivered
Click to Open Rate	Unique emails clickers/Unique emails delivered
Click Rate	Unique emails clickers/Unique emails delivered
Unsub Rate	Unique emails unsubscribed/Unique emails opened

2. **Account Registration Report**

This report helps the business users to understand the trends in account registrations over time, specifically geography of the registrants and their optin status.

Fields to be included:

Field	Description
State	Registration state
Month	
Permission	See the "Data" section
Count of Customer IDs	

3. **Device Registration Report**

This report helps the business users to understand the trends in device registrations, by carrier, over time.

Fields to be included:

Field	Description
Carrier	Device carrier
Month	
Device	
Count of Customer IDs	