Rosetta: Final Design Report
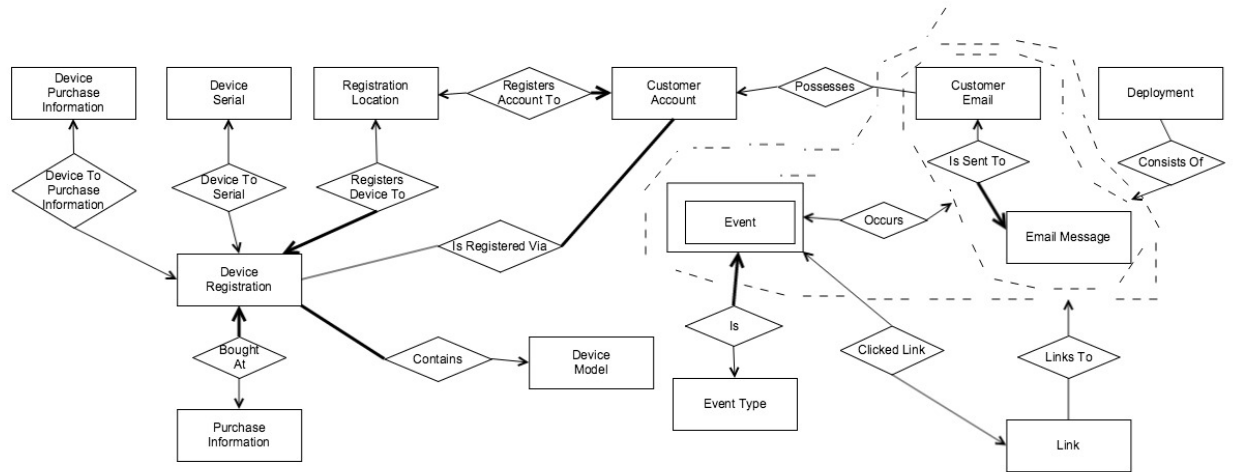# Team SELECT * FROM JoPo;

By: Esha Joshi, Andrew Wang, Delaney Giacalone, Nick Verbos
CPE 366-01: Database Modeling, Design, and Implementation

June 9, 2015

Contents

1 E-R Model, Operational Data Store

2 Relational Model, Operational Data Store

2.1 Entity Sets

CustomerEmail Entity Set

Name	Type	Description
email_id	Int	Unique Identifier of the customer email address
email_domain	String	Domain of customer email
audience	String	Customer segment

CustomerAccount Entity Set

Name	Type	Description
customer_id	int	Unique customer Identifier
permission	boolean	Optin/optout indicator for marketing communications
language	string	Language of customer
reg_date	Date	Date of customer registration
customer_tier	string	Tier customer has been assigned to by BrandX
gender	char	Gender of customer
zip	int	Zip code of customer
state	string	State of customer
income_level	string	Customer Income
num_registrations	int	Number of registrations customer has

RegistrationLocation Entity Set

RegistrationLocation Entity Set: Name	Type	Description
registration_source_id	int	Identifier of the account registration source
source_name	int	Name of the registration source

DeviceRegistration Entity Set

Name	Type	Description
registration_id	int	Unique registration
registration_date	Date	The date of registration

PurchaseInformation Entity Set

Name	Type	Description
purchase_store_name	String	Purchase store name
purchase_store_state	String	Purchase store state
purchase_store_city	String	Purchase store city
ecomm_flag	boolean	ecommFlag

DeviceSerial Entity Set

Name	Type	Description
serial_number	String	Serial number

DevicePurchaseDate Entity Set

Name	Type	Description
purchase_date	Date	Purchase date

Event Entity Set

Name	Type	Description
event_date	Date	When event happened

EventTypes Entity Set

Name	Type	Description
event_type_id	int	ID for email event types
event_type_name	String	Email event type

EmailSent Entity Set

Name	Type	Description
campaign_name	String	Name of email campaign
version	String	Creative version
subject_line	String	Email subject line
audience	String	Customer segment

Deployment Entity Set

Name	Type	Description
deployment_id	int	Batch ID
deployment_date	Date	When email has been sent

Link Entity Set

Name	Type	Description
link_name	String	Link Alias, e.g. Hero Module or Contact Customer Service
url	String	Link URL

DeviceModel Entity Set

Name	Type	Description
device_name	String	Name of device, such as Constellation 5
carrier	String	AT&T, Version, etc.
device_model	String	Model of Device
device_type	String	Phone, Tablet, etc.

2.2 Relationship Sets

- Contains(DeviceRegistration, DeviceModel)

  - many-to-one

- DeviceRegistration participates in DeviceModel

- BoughtAt(DeviceRegistration, PurchaseInfo)

  - many-to-one

- DeviceRegistration participates in PurchaseInfo

  - PurchaseInfo is weak

- OptionallyHas(DeviceRegistration, DeviceInfo)

  - one-to-one

- RegistersDeviceTo(DeviceRegistration, RegistrationLocation)

  - one-to-one
  - DeviceRegistration participates in RegistrationLocation

- IsRegisteredVia(CustomerAccount, DeviceRegistration)

  - many-to-many
  - DeviceRegistration participates in CustomerAccount

- RegistersAccountTo(CustomerAccount, RegistrationLocation)

  - one-to-one
  - CustomerAccount participates in RegistrationLocation

- Possesses(CustomerAccount, CustomerEmail)

  - one-to-many

- IsSentTo(EmailSent, CustomerEmail)

- many-to-many
- EmailSent participates in CustomerEmail

- Happens(Event, IsSentTo(CustomerEmail, EmailSent))

  - many-to-many

- Link(Links, Happens(Event, IsSentTo(CustomerEmail, EmailSent)))

  - many-to-many

- Is(Event, EventType)

  - one-to-one
  - Event participates in EventType

- ConsistsOf(EmailSent, Deployments)

  - one-to-many
  - Deployments participates in EmailSent

- LinksTo(EmailSent, Link)

  - one-to-many

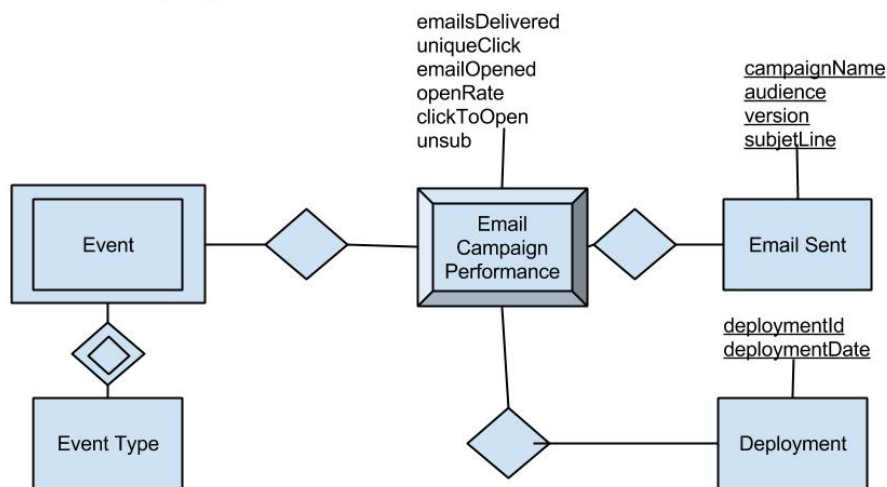- ClickedLink(Event, Link)

  - one-to-many

3 Brief Description of the ETL Process

- Created an class for each relational table.

- Created a dictionary (Python hashtable) for each relational table.

- Parsed each data set provided by Rosetta, built a table entry tuple, and added it to the dictionary.

- Primary key of each table is represented by an auto-incrementing integer each time a new tuple is added to the table.

- Each table tuple in the dictionary was converted into INSERT INTO statements after each data set .csv file was parsed.

- These insert statements are in DB-inserts.sql which calls ¡TableName¿-insert.sql.

4 E-R Model, Data Warehouse

Email Campaign Performance

Device Registration Report

Count of
Customer IDs

Device Registration	◇	Device Registration Report	◇	Device Model

month

◇

Customer Account

carrier
device
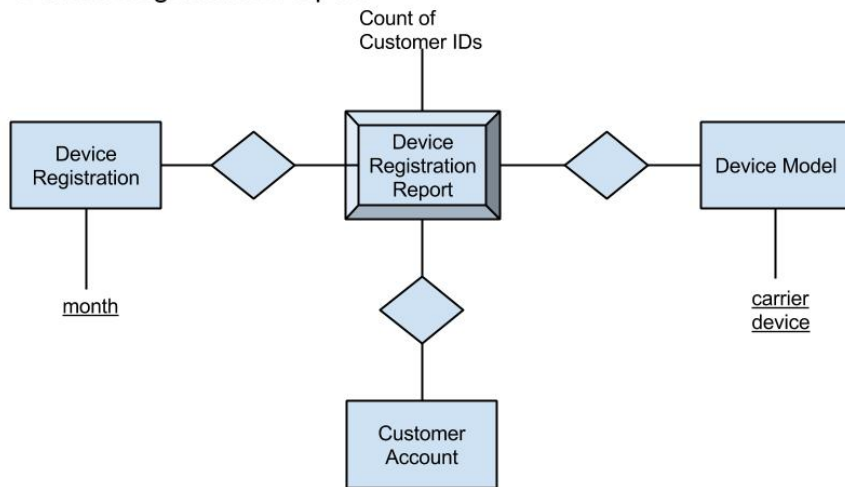
Account Registration Report

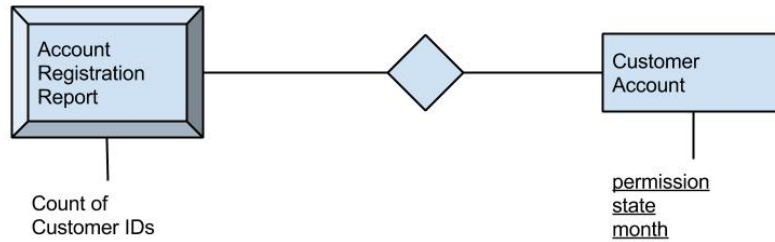# 5 Relational Model, Data Warehouse

Email Campaign Performance Entity Set

Name	Type	Description
campaign_name	String	Campaign name
audience	String	Audience
version	String	Version
subject_line	String	Subject line
deployment_date	Date	Date of deployment of campaign
UniqueEmailsDelivered	int	Sum of unique emails sent Sum of bounced emails
UniqueEmailsOpened	int	Sum of unique email opened
UniqueClickers	int	Sum of unique customers who opened this email
OpenRate	int	Unique emails opened/Unique emails delivered
ClickRate	int	Unique emails clickers/Unique emails delivered
ClickToOpenRate	int	Unique emails clickers/Unique emails delivered
UnsubRate	int	Unique emails unsubscribed/Unique emails opened

AccountRegistration Entity Set

Name	Type	Description
state	String	State
month	String	Month
device	String	Device
customer_ID_count	int	Count of Customer IDs

DeviceRegistration Entity Set

Name	Type	Description
carrier	String	Carrier
month	String	Month
device	String	Device
customer_ID_count	int	Count of Customer IDs

6   Data Dictionary

6.1   Entity Sets

CustomerEmail Entity Set

Data Name	Source
email_id	Account Registration Data Set: Email ID
email_domain	Account Registration Data Set: Email Domain
audience	Email Event Data Set: Audience

CustomerAccount Entity Set

Name	Source
customer_id	Account Registration Data Set: Customer ID
permission	Account Registration Data Set: Permission
language	Account Registration Data Set: Language
reg_date	Account Registration Data Set: Registration Date
customer_tier	Account Registration Data Set: Customer Tier
gender	Account Registration Data Set: Gender
zip	Account Registration Data Set: ZIP
state	Account Registration Data Set: State
income_level	Account Registration Data Set: Income Level
num_registrations	Device Registration Data Set: Number of Registrations

RegistrationLocation Entity Set

Name	Source
registration_source_id	Device Registration Data Set: Registration Source ID
source_name	Device Registration Data Set: Registration Source Name

DeviceRegistration Entity Set

Name	Source
registration_id	Device Registration Data Set: Registration ID
registration_date	Device Registration Data Set: Registration Date

PurchaseInformation Entity Set

Name	Source
purchase_store_name	Device Registration Data Set: Purchase Store Name
purchase_store_state	Device Registration Data Set: Purchase Store State
purchase_store_city	Device Registration Data Set: Purchase Store City
ecomm_flag	Device Registration Data Set: Ecomm Flag

DeviceSerial Entity Set

Name	Source
serial_number	Device Registration Data Set: Serial Number

DevicePurchaseDate Entity Set

Name	Source
purchase_date	Device Registration Data Set: Purchase Date

Event Entity Set

Name	Source
event_date	Email Event Data Set: Event Date

EventTypes Entity Set

Name	Source
event_type_id	Email Event Data Set: Event Type ID
event_type_name	Email Event Data Set: Event Type Name

EmailSent Entity Set

Name	Source
campaign_name	Email Event Data Set: Campaign Name
version	Email Event Data Set: Version
subject_line	Email Event Data Set: Subject Line
audience	Email Event Data Set: Audience

Deployment Entity Set

Name	Source
deployment_id	Email Event Data Set: Deployment ID
deployment_date	Email Event Data Set: Deployment Date

Link Entity Set

link_name	Email Event Data Set: Link Name
url	Email Event Data Set: URL

DeviceModel Entity Set

Name	Source
device_name	Device Data Set: Device Name
carrier	Device Data Set: Carrier
device_model	Device Data Set: Device Model
device_type	Device Data Set: Device Type

6.2 Relationship Sets

Contains Relationship Set

Name	Source
DeviceRegistration	Data Store: DeviceRegistration Table
DeviceModel	Data Store: DeviceModel Table

BoughtAt Relationship Set

Name	Source
DeviceRegistration	Data Store: DeviceRegistration Table
PurchaseInfo	Data Store: PurchaseInfo Table

OptionallyHas Relationship Set

Name	Source
DeviceRegistration	Data Store: DeviceRegistration Table
DeviceInfo	Data Store: DeviceInfo Table

RegistersDeviceTo Relationship Set

Name	Source
DeviceRegistration	Data Store: DeviceRegistration Table
RegistrationLocation	Data Store: RegistrationLocation Table

IsRegisteredVia Relationship Set

Name	Source
CustomerAccount	Data Store: CustomerAccount Table
DeviceRegistration	Data Store: DeviceRegistration Table

RegistersAccountTo Relationship Set

Name	Source
CustomerAccount	Data Store: CustomerAccount Table
RegistrationLocation	Data Store: RegistrationLocation Table

Possesses Relationship Set

Name	Source
CustomerAccount	Data Store: CustomerAccount Table
CustomerEmail	Data Store: CustomerEmail Table

IsSentTo Relationship Set

Name	Source
EmailSent	Data Store: EmailSent Table
CustomerEmail	Data Store: CustomerEmail Table

Happens Relationship Set

Name	Source
Event	Data Store: Event Table
SentTo(CustomerEmail, EmailSent)	Data Store: SentTo(CustomerEmail, EmailSent) Relationship Set

Link Relationship Set

Name	Source
Links	Data Store: Link Table
Happens(Event, IsSentTo(CustomerEmail, EmailSent))	Data Store: Happens(Event, IsSentTo(CustomerEmail, EmailSent)) Relationship Set

ConsistsOf Relationship Set

Name	Source
EmailSent	Data Store: EmailSent Table
Deployments	Data Store: Deployment Table

LinksTo Relationship Set

Name	Source
EmailSent	Data Store: EmailSent
Link	Data Store: Link Table

ClickedLink Relationship Set

Name	Source
Event	Data Store: Event Table
Link	Data Store: Link Table

6.3   Data Warehouse

Email Campaign Performance

Name	Source
Campaign Name	Email Event Data Set: Campaign Name
Audience	Email Event Data Set: Audience
Subject Line	Email Event Data Set: Subject Line
Deployment Date	Email Event Data Set: Deployment Date
Unique Emails Delivered	Aggregation of isSentTo with EmailSentId and Deployment Date where Type is Bounced (event_type is [32, 48])
Unique Emails Opened	Aggregation of isSentTo with EmailSentId and Deployment Date where Type is Opened (event_type is 2)
Unique Clickers	Aggregation of isSentTo with EmailSentId and Deployment Date where Type is Clicked (event_type is 0)
Open Rate	Comes from Unique Emails Opened / Unique Emails Delivered
Click to Open Rate	Comes from Click Rate / Open Rate
Click Rate	Comes from Unique Emails Clicked / Unique Emails Delivered
Unsub Rate	Comes from Unique Emails Unsubscribed / Unique Emails Delivered

Account Registration Report

Name	Source
State	Account Registration Data Set: State
Month	Account Registration Data Set: Registration Date
Permission	Account Registration Data Set: Permission
Count of Customer IDs	Aggregation of Count of the grouping of state, month, year, permission

Device Registration Report

Name	Source
Carrier	Device Data Set: Carrier
Month	Device Registration Data Set: Registration Date
Device	Device Registration Data Set: Device Model
Count of Customer IDs	Aggregation of the Count of the grouping of carrier, Device, month, year