CRM Data Warehouse

Announcements

• Assignment 2 is on the subject website
• Students must form assignment groups
  ASAP: refer to the assignment for details

Customer-Centric Data Warehouse, design issues

• Data modelling issues
• Capturing changes
• Consolidation and integration issues
• Storage of unstructured information
• Data storage requirements
• Privacy issues
CRM data warehouse

- Customer dimension is critical for effective CRM
- It is most challenging dimension for any data warehouse
- Can have millions of rows (credit card companies, government agencies exceed 100,000 millions of rows)
- Hundreds of attributes
- One leading marketer maintains 3000 customer attributes
- (Kimball 2002, The data warehouse toolkit, chapter 6)

Data modelling for CRM data warehouse

- What works for CRM?
- We will examine Todman’s and Kimbal’s approaches

Customer data warehouse design (Todman, 2001)

- Two types of customer data
  - Customer’s behaviour
  - Customer’s circumstances
- Customer table contains data about customer circumstances
- Fact tables contain customer’s behavioural data, for example:
  - Sales
  - Shipments
  - Insurance claims
  - Flight bookings
Dimensional Model Characteristics

- Dimensional model is Behaviour centric
  - It’s purpose is to enable comprehensive analysis of behavioural data
  - Dimensional models are ideal for slicing and dicing the fact table data

Example Star Schema

- Sales
  - Customer
  - Time
  - product
  - store

Example Star Schema

- Orders
  - Customer
  - Time
  - product
  - salesperson
Example Star Schema

Customer-centric models

- If we want to examine the effect of change in customer circumstances we need to focus on modelling customer dimension
- CRM data warehouse is customer-centric

(Todman, 2001)

Customer-centric approach

- The emphasis is shifted from behaviour
- More value attached to the customer’s personal circumstances
- This enables us to classify customers into segments
Modelling customer data

- Dimensional modelling or E-R modelling
- What works for CRM?
- Todman’s approach:
  - Use dimensional models in modelling customer behaviour
  - But use more conventional approaches when modelling customers’ circumstances

General Conceptual Model (GCL) (Todman 2001)

- Todman proposes General Conceptual Model for customer-centric warehouse
- Three general entities
  - Customer
  - Customer changing circumstances
  - Customer behaviour
  - Customer derived segment

General conceptual model for a customer-centric data warehouse

(Adapted from Todman 2001)
CGL model for a customer with changing circumstances

Customer contains customer static information
- Information which does not change or changes do not need to be kept eg: title, name, date of birth, sex

Customer Changing Circumstances contains information we want to keep
- eg: address, marital status, hobbies, profession, income, employer, children, spouse

Customer Attributes

- Customer's Static Information
  - example
    - Title
    - Name
    - Telephone number
    - Date of Birth
    - Sex
Customer Attributes, example

- Customer’s Changing Information
  - Address
  - Marital Status
  - Children’s details
  - Spouse details
  - Income
  - Hobbies and interests
  - Trade of profession
  - Employer’s details

Customer changing circumstances

- Typically in data warehousing we analyse behaviour – customer interaction with business
  - e.g. Customer orders, purchases, ...
- In CRM we need to know, for example:
  - what customer circumstances cause customers to churn,
  - what circumstances cause customers to change their product preferences etc...
  - Customer moves to a different area
  - Customer gets married
  - Starts to play tennis

Customer changing circumstances

- In the CRM data warehouse we need to analyse the effect of changes in customer circumstances
- We have to identify the changes we need to store (address, marital status, profession)
- Some previous values do not change (date of birth)
- Some previous values can be lost (e.g. previous title)
- What changes we store depend on the type of business
Example: customer with changing circumstances

- Customer
- Address
- Marital Status

Each customer can have many changes of addresses and marital status.

Note:
The model does not show that there are many attributes that can change. It shows that each attribute can change many times.

The CGL model extended to include behavior

- Customer Behavior
- Customer
- Customer Changing Circumstances

Customer behavior: eg sales, flight reservations.
Example GML for customer behavior

Sales → Customer

orders → shipments

Customer derived segment

Customer Behavior → Customer

Customer → Customer Changing Circumstances

Customer Derived Segment

Segmentation

- Todman categorises segmentation into:
  - Circumstances
    - eg group customers according to sex, age, group, occupation
  - Behavioural segmentation
    - eg Purchased type of products (car insurance, home insurance)
  - Derived segmentation
Customer derived segment

- Segmentation derived from customer’s circumstances or behaviour
- Examples of derived segments:
  - Propensity to churn
  - Estimated life time value
  - Upsell and crosssell potential
  - Credit risk
  - This type of segments can be derived from data using data mining, OLAP and other analytical methods

Example: the Wine Club case study (Todman, 2001)

- The organisation: mail order wine club
- Sells wines and accessories (glassware, tableware, books)
- Organises trips to special events
- Existing systems: customer admin., stock control, order processing, shipments, trip bookings

Example: the Wine Club case study (read, Todman, 2001, pp 27-30)

Business requirements
The club is loosing its market share. Customers are leaving the club.
It is difficult to get answers to strategic questions from existing systems. The directors of the Wine Club want to know:
Which product lines are increasing in popularity and which are decreasing?
Do customers tend to purchase a particular class of product?
Which customers are likely to churn?
For special promotions. Which customers are likely to buy the product?
Wine Club customer changing circumstances

Address
Spouse
Hobby/Interest
Marital Status
Customer
Profession
Child
Income
Employer

Wine Club customer behavior

Wine Sales
Customer
Accessory Sales
Trips Taken

The example model extended to include behavior (incomplete)

Wine Sales
Tips
Customer
Address
Accessories
Marital Status

(Adapted from Todman, 2001)
Derived segment examples for the Wine Club

- Lifetime Values
  - Customer
    - Recently Churned and Winback
    - Special Promotions

Todman's modelling approach

- General Conceptual Model
- Conceptual model (detailed information requirements)
- Logical model (relational schema)

Example Relational logical schema (the Wine Club case study)

- Relation Customer
  - Customer_code
  - Customer_name
  - Hobby_code
  - Date_joined
  - PK(customer_code)
  - FK(Hobby_code)

- Relation Customer address exist
  - Customer_code
  - Customer_address_exist_start
  - Customer_address_exist_end
  - Customer_address
  - PK(Customer_code, Customer_address_exist_start)
  - FK(Customer_code)

- Relation Customer exist
  - Customer_code
  - Customer_exist_start
  - Customer_exist_end
  - PK(Customer_code)
Capturing changes

• Capturing behaviour data: fact table linked to time dimension, data captured on a periodic, basis (hourly, daily, etc.)

• Capturing circumstances (dimensional changes) is difficult
  - historic values of addresses, product prices

Problems and issues involving time

• Identifying and capturing the temporal requirements
• Capture of dimensional updates
• The timelines of capture
• Synchronisation of changes
• Dynamic nature of segments

CRM data warehouse modelling (Kimbal)

- Customer dimension is a critical element for effective CRM
- Modelling customer dimension is very difficult
- Customer table can have millions of rows
- Customer dimension can have hundreds of attributes
Large changing customer dimensions

- Multimillion-row customer dimensions
- Too long to constrain or browse among the relationships in such a big table
- Place frequently changing attributes into a separate dimension called minidimension

Demographic minidimension with a customer dimension

<table>
<thead>
<tr>
<th>Customer Dimension</th>
<th>Customer Key (PK)</th>
<th>Customer ID (Natural key)</th>
<th>Customer Name</th>
<th>Customer Address</th>
<th>Customer Date of Birth</th>
<th>Customer Date of 1st Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>Gender</td>
<td>Annual income</td>
<td>Number of children</td>
<td>Customer marital status</td>
<td></td>
</tr>
</tbody>
</table>

Fact Table

- Customer Key (FK)
- Demographics Key (FK)
- More Foreign Keys

Customer Demographics Dimension

- Customer Demographics Key (PK)
- Age Band
- Gender
- Income Band
- Number of Children Band
- Marital Status

Separate demographic and behavioral minidimensions

Customer Dimension

- Customer Key (PK)
- Relative constant attributes

Customer Demographics Dimension

- Customer Demographics Key (PK)
- Demographics attributes

Customer Purchase-Credit Dimension

- Customer Purchase-Credit Key (PK)
- Credit/payment behavioral attributes
Snowflaking with a dimension outrigger of cluster of low-cardinality attributes (adapted from Kimba 2002)

<table>
<thead>
<tr>
<th>Fact Table</th>
<th>Customer Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Key (PK)</td>
<td>Customer Key (PK)</td>
</tr>
<tr>
<td>More foreign keys</td>
<td>Customer ID (natural key)</td>
</tr>
<tr>
<td>Facts…</td>
<td>Customer Salutation</td>
</tr>
<tr>
<td>… and more</td>
<td>Customer First Name</td>
</tr>
<tr>
<td></td>
<td>Customer Surname</td>
</tr>
<tr>
<td></td>
<td>Customer City</td>
</tr>
<tr>
<td></td>
<td>Customer County</td>
</tr>
<tr>
<td></td>
<td>County Demographics Key (FK)</td>
</tr>
<tr>
<td></td>
<td>Customer State</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>County Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>County Demographics Key (PK)</td>
</tr>
<tr>
<td>Total Population</td>
</tr>
<tr>
<td>% Pop. Under 5 Years</td>
</tr>
<tr>
<td>% Pop. Under 18 Years</td>
</tr>
<tr>
<td>% Pop. 65 Years and Older</td>
</tr>
<tr>
<td>Female Pop.</td>
</tr>
<tr>
<td>% Female Pop.</td>
</tr>
<tr>
<td>Male Pop.</td>
</tr>
<tr>
<td>% Male Pop.</td>
</tr>
<tr>
<td>N. of High School Graduates</td>
</tr>
<tr>
<td>N. Of College Graduates</td>
</tr>
<tr>
<td>N. Of Housing Units</td>
</tr>
<tr>
<td>Homeownership Rate</td>
</tr>
</tbody>
</table>

More Challenges to developing Customer Data Warehouse

- Performance and storage requirements are critical
- High volumes of data:
  - Typically extensive history is required
  - Data from web-based channels (several terabytes)
  - Granularity: analysis of behavioral data requires access to the individual customer transaction level
  - As a result, order of magnitude more data required to be stored than in other types of DW

More Challenges to developing customer data warehouse

- Integration of multiple sources of customer data
- Changes in the business environment
- Accommodate new and changing data sources
- Rapidly change to accommodate new data and new data structures
- Storing unstructured information such as feedback and complaints
- Privacy issues
Data sources

Customer Data Gathering

Customer Data Warehouse

Information delivery systems

Decision makers, users

References

• Todman C. Designing a Data Warehouse Supporting Customer Relationship Management, Prentice Hall, 2001