Data Warehouse for CRM

Introduction to data warehousing

Topics
- Introduction to data warehousing (today)
- Data warehouse modelling
- What works for CRM?
- Modelling data for Customer Data Warehouse.

Presentation structure

- Data warehouse definitions
- Data warehouse architectures
- Data warehousing process
- Data warehouse and analytical CRM
Technologies for Analytical CRM

- Data warehouse
  - Analysis and delivery systems:
    - OLAP, query and reporting
    - Data Mining

Architecture for Analytical CRM

Data warehouse: definitions

- Definition 1
  Data Warehouse is an integrated and consistent store of subject-oriented data that is obtained from a variety of sources and formatted into a meaningful context to support decision-making in an organization
Need for Data Warehousing

- Information gap: data → information
- Most systems are developed to support operational (transaction) processing, not decision making.

Operational vs. Informational Systems

- Operational systems support transaction processing, example: order processing, reservation system.
  - Transaction processing captures, stores, and manipulates data to support daily operations of business.
- Informational systems support decision making, example: sales trend analysis, customer profile, customer segmentation.
  - Analytical information processing is the analysis of data to support decision making.
- Data warehouse is a database designed specifically to facilitate analytical information processing.

Definitions

- Definition 2
  "... a subject-oriented, integrated, time-variant, and non-volatile collection of data* used in support of management’s decisions"  
  Inmon and Hackathorn (1994)
  * (High quality data)
Data warehouse - subject oriented

- The data warehouse is organised around key subjects (high level entities) of the enterprise e.g. customers, students, products.
- This may be contrasted with the process orientation of many OLTP systems.

Data warehouse - integrated

- DW integrates data collected from disparate sources:
  - Operational systems
  - Internal documents
  - External data
  - The data must relate specifically to the area of knowledge to be attained.
    - For example, data for CRM DW can be collected from:
      - Customer touchpoints
      - Operational systems: sales, order processing, marketing
      - Application forms, reports
      - Credit agencies, marketing agencies
  - Data in DW is integrated - consistent names, formats, encoding structures etc.

Data warehouse - time variant

- Data in the data warehouse is periodic (time dimension)
- The data consists of a series of “snapshots” which are time-stamped and record values at a moment in time.
- This supports trend analysis of the data.
- Operational data is transient.
Example of periodic data in a data warehouse
(McFadden et al 1999)

Data warehouse – non-volatile

- Data in OLTP system databases is continuously updated (inserts, deletes and changes) by end-users
- Data in a data warehouse is periodically up-loaded from operational systems, but cannot be updated by end-users

Motivations for data warehousing

- Demands on OLTP data bases for query processing is too great
- Data warehousing is designed for efficient retrieval
- Data in legacy systems is frequently inconsistent, of poor quality and stored in different formats
- Data warehouses are motivated by the need to view the entire enterprise from a single point
An enterprise data warehouse (motivated by the need to view the entire enterprise from a single point)

**Purposes of Data Warehouses**

- Data warehouses, utilized correctly, can provide business insight
  - Product performance
  - Campaign performance
  - Profitability
  - Cost structures
  - Trend in sales
  - Customer behavior etc..

**Warehouse architecture**

- Major options:
  - Enterprise data warehouse
  - Dependent data mart
  - Independent data mart
Enterprise data warehouse

• Single central data source
• Large in scope and often size
• Maximises the benefits of integration
• User views
• Difficult to meet the requirements of different user groups
• Projects often fail
Data Marts

- Sometimes seen as departmental or even workgroup level operation of the warehouse
- Geared to anticipated queries and operations [reports, projections, specific analyses]
- Data marts customise decision support for different groups

Types of Data Marts

- Dependent - Populated from the EDW
- Independent - Data taken directly from the operational databases.
Dependent data marts

- Three-level architecture
  - Operational data
  - Enterprise data warehouse (EDW) - single source of data for decision making
  - Data marts - limited scope; data selected from EDW

Dependant data mart

- Subsets of data are taken from enterprise data warehouse and organised to fit requirements of a business unit or application
Independent data mart

Data sources  Independent data marts  Clients

Independent data marts

- Many small data warehouses
- Integration is a problem
- Duplication and inconsistency

The Data Warehousing process

- Data warehousing is an evolutionary process involving:
  - sourcing
  - ETL (extract, transform, load)
  - storing
  - delivering data to support decision makers
Data warehousing

- Source
- ETL
- Metadata

Data warehouse development

- Requirements identification
- Logical design, data modelling
- Warehouse architecture, technology and tools
- Data extract, transform and load (ETL)
- Physical database design
- Delivery systems
- Operational policies

Requirements identification

- Users needs
- Data availability

• Once a system exists it becomes the major source of requirements
  - Feedback mechanics are critical

Evolutionary approach
Logical design, Data modelling

- Designing the database of the data warehouse is a key issue
- There are two main approaches
  - entity relationship modelling and normalisation
  - dimensional modelling

Data modelling will be covered in next lecture.

Data extract transform and load (ETL)

- Data has to be extracted from data sources and then cleaned and integrated before it enters the data warehouse
- Data staging is a set of processes that clean, transform, combine and prepare data for use in DW

Data quality

- Data quality is a critical issue in data warehousing
- Data sourcing and cleansing is the largest task in most data warehousing projects
Examples of heterogeneous data

- Data in operational systems are typically of poor quality.
- Distributed on variety of incompatible hardware and software platforms.
- Transient.

Data quality is a key factor in the success of data warehousing and CRM systems.

Data quality problems are widespread.

- Poor quality data may lead to misguided decisions.
- Data sourcing and cleansing is the largest task in most data warehousing projects.

On its way to the warehouse, data from the operational systems is cleaned and transformed.
Extract, Transform, Load (ETL) Processes (data reconciliation)

- Extract - capture data from operational DBs and other data sources
- Data scrubbing (cleansing) and transformation
- Load
- Most effort, time and cost go into ETL
- ETL tools

Data transformation

- Transform
  - Convert the data format from the source to the target system
- Data structures need to be made alike
- Data may need to be aggregated
- Fields can be split or merged
- Data may need to be changed from one format to another [e.g. °F to °C]

Data extract, transform and load

- Load and Index
  - When the warehouse is first created.
  - Static data capture.
- Update Mode
  - Ongoing update of the warehouse.
  - Incremental data capture.
Tools for data quality

- Various tools are available to support data sourcing and cleansing
- These typically support transformation from various file formats and some data validation
- But they can’t do lots of things…
  - Know where somebody lives, what their name is etc.

Data Quality

- two levels
  - data (content)
  - metadata (structure)
Meta-data

- Meta-data is data about data
- In traditional information systems, metadata recorded details of data elements, their formats and validation constraints
- In a data warehouse, richer metadata is required

Meta-data in a Data Warehouse

- Meta-data in a data warehouse should include
  - Definitions of data elements
  - Validation constraints
  - Details of source systems and transformations required
  - Comments about data quality
  - Date at which data was sourced
  - Details of who is responsible for the data

Data and metadata Integration

- Data integration:
  - Combines data from multiple sources into a coherent store
- Schema integration
  - Integrate metadata from different sources
  - Entity identification problem: How do you know that the customer ID field from two DBs are coded the same?
Data warehouse and Analytical CRM

- Analytical CRM consolidates customer data across the enterprise and provides **unified view of the customer**
- Data warehouse provides data infrastructure for analytical CRM

References

- Zikmund R., McLeod R., Gilbert F. “Customer Relationship Management, Integrating Marketing Strategy and Information Technology”
- The Groth text, chapter on data warehouse
- “Modern Data Base Management”, by Hoffer, et al, Chapter on Data Warehouse