Data Modelling

Content

- Second assignment
- Nature of data modelling
- Tools/Techniques used in data modelling
- Place in ISD
- Evaluation of data modelling
- Reading list

Assignment 2 – OO modelling

- Due Wednesday 5 October 2005 - week 11
- Undertake in pairs from the same tute group
- Individual submissions accepted (decide this week)

Assignment Pitfalls

- Not starting early
- Not confirming your understanding with the tutor
- Not starting early
- Not integrating the separate elements of the models
- Not starting early

Data modelling describe:

- Structure
- Meaning
- Relationship

Data modelling help us to grasp:

- Static Data in the organisation
- Fundamental building block of the system
- Two perspectives (Process and Data)
Techniques used in Data modelling

- Entity relationship diagrams
- Normalisation
- Data Dictionary
- What difference?
- Use both?

Entity

- Entity – things of interest to the business

Employee

- Identification of an entity is subjective
- Entities can be:
  - Real eg product
  - Abstract eg Quota
  - Event remembered eg sale
  - Role played eg employee

Relationship

- Relationship Between entities
  - Employee
  - Department

Cardinality (eg. One to many, one to one etc.)
Degree of relationship (Unary, Binary, Ternary)

Examples of Cardinalities

PATIENT
HISTORY
PATIENT
PROJECT
EMPLOYEE
PERSON

Has
Is assigned to
Is married to

Relationship Cardinality Summary

- Mandatory 1 cardinality
- Many cardinality (1,2 ... m)
- Optional (0 or 1) cardinality
- Optional (0 or many) cardinality

Unary Relationship

- Also called a recursive relationship
Binary Relationship

- A binary relationship is a relationship between instances of two entity types.

<table>
<thead>
<tr>
<th>PROJECT</th>
<th>LEADS</th>
<th>CUSTOMER</th>
<th>SELL ORDER</th>
<th>ITEM</th>
<th>SUPPLIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leads</td>
<td></td>
<td>CUSTOMER</td>
<td></td>
<td>ITEM</td>
<td>SUPPLIER</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ONE TO ONE</td>
<td></td>
<td>ONE TO MANY</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ONE TO MANY</td>
<td></td>
<td>MANY TO MANY</td>
</tr>
</tbody>
</table>

Ternary Relationship

- A ternary relationship is a relationship between instances of three entity types.

<table>
<thead>
<tr>
<th>CUSTOMER</th>
<th>VENDOR</th>
<th>PART</th>
<th>SUPPLIES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leads</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Attributes

- What we want to know about the entity or a relationship
- Types:
  - Derived,
  - multi-valued,
  - Composite,
  - Simple

Example of attributes

Why normalisation?

- Remove redundancy and incompleteness
- Bottom up process
- Rely on Maths – well researched

Determining columns

- One fact per column
- Hidden data
- Derivable data
- Determining the key
**Steps in Basic Normalisation**

Basic Normalisation is most often accomplished in three stages (these are the three basic normal forms):

- **First Normal Form**
  - Unnormalised table
  - Remove repeating groups

- **Second Normal Form**
  - First Normal Form
  - Remove partial dependencies

- **Third Normal Form**
  - Second Normal Form
  - Remove transitive dependencies

**First normal form**

Step 1: Remove the repeating group

- Why is repeating groups a problem?
- Determine the key for the new group.

**Second and Third normal forms**

- Eliminate redundancy
- Determinates – one or more columns which determines other column values

**Second and third normal form procedure**

- Identify any determinates (other than the key)
- Establish a separate table for each determinate and the columns it determines
- Name the new tables
- Remove the determined columns from the original table

**Third normal form**

A table is in third normal form if the only determinates of nonkey columns are candidate keys

**Advanced normalisation**

- A set of tables can be in 3NF and still not be fully normalised
- Further stages of normalisation are BCNF, 4NF, 5 NF and Domain key NF
Higher Normal forms

- Occur infrequently
- Most tables in 3 NF is already in BCNF, 4NF and 5 NF
- Data in 3NF but not in 5NF has
  - Redundancy
  - Insert/update/delete anomalies
  - Difficulty in storing facts independently

BC NF

Example

Branch-customer relationship (customer no, branch no, visiting frequency, date relationship established, salesperson no)

Problem

- Salesperson \( \rightarrow \) branch no
- Overlapping and candidate keys
  - Customer no and branch no
  - Customer no and salesperson
  - Branch -customer relationship (customer no, salesperson no, branch no, visiting frequency, date relationship established)
  - Customer salesperson relationship (customer no, salesperson no, visiting frequency, date relationship established)
  - Salesperson (Salesperson No, branch no)

B-C NF

Every determinant must be a candidate key
(must have overlapping keys)

Fourth and Fifth NF

- Apply to all-key tables, degree > 2

Problems are the result of multi-valued dependencies
Fifth NF

Keep splitting tables until:

- Any further splitting would lead to tables unable to be joined to produce the original table
- The only splits left are trivial

Dealing authority problem

Many to many relationships resolved

Combined table

Combined table which cannot be split

Other normalisation issues

- Normalisation and redundancy
  - Overlapping classifications
  - Derivable data
- Selecting primary keys
Thinking in Data modelling

- Hard Vs Soft ??
- Perspective
  - Objective vs Subjective
  - Nature of the organisation

Evaluation of Data modelling

<table>
<thead>
<tr>
<th>Conceptual</th>
<th>Problem oriented</th>
<th>Product oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structured analysis Entity relationship modeling Logical construction of systems Modern structured analysis Object oriented analysis</td>
<td>Structured design Object oriented design</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Formal</th>
<th>Levels of abstraction Stepwise refinement Proof of correctness Data abstraction JSP Object oriented programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSL/PSA</td>
<td>JSD VDM</td>
</tr>
</tbody>
</table>

Advantages of Data modelling

- Data model is not computer oriented (agree??)
- Model understandable by technologist and users
- Does not show bias
- UoD can vary (whole organisation or department)
- Readily transformable into other models
- Different data analysis techniques
- Data modelling is rule-based

Disadvantages

- Does not encourage/support user participation
- Your view on the organisation –people or data
- The idea that the model is THE model
- Subjective view
- One-side into data
- Others??

Advantages of Normalisation

- Rid the data of redundancy and other problems
- Very well researched
- Math basis for normalisation

Disadvantages of normalisation

- Does not encourage/support user participation
- Your view on the organisation –people or data
- One-side into data
- Can be done mechanistically without thought
- Others??
Process modelling view of ISD

- Development group
- Object system
- Change process
- Object system
- Object system
- Environment

Reading for next week