Data Warehousing

Multidimensional Analysis

Joachim Hammer
Data Warehousing Architecture

Information Sources
- Semistructured Sources
- Operational DB’s

Data Warehouse
- extract
- transform
- load
- refresh
- etc.

Data Warehouse Server

OLAP Servers
- MOLAP
- ROLAP

Data Marts

Clients
- Analysis
- Query/Reporting
- Data Mining
Multidimensional Modeling

• Support ad-hoc querying for business analyst
• Think in terms of spreadsheets
  – View sales data by geography, time, or product
• Away from traditional ER models
  – “One fact in one location”
  – Entities are inter-related through a series of joins
Multidimensional Data Model

• Database is a set of facts in multidimensional space

• Measures
  – Numerical data being tracked
  – Stored in central fact table
  – E.g., sales, inventory, expenditures

• Dimensions
  – Business parameters
  – E.g., time, geography, account data
Example

- “Sales by product line manager over the past six months”

<table>
<thead>
<tr>
<th>Key columns joining fact table to dimension tables</th>
<th>Numerical Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prod Code</td>
<td>Time Code</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...
Multidimensional Data Model Cont’d

• Dimensions have attributes
• Organized into hierarchies
  – E.g., Time dimension: days → weeks → quarters
  – E.g., Product dimension: product → product line → brand
• Operators to navigate the hierarchies
  – “Roll-up”
  – Drill-down is the opposite of roll-up
  – Slice (defines a subcube)
  – Various visualization ops (e.g., pivot)
• Physical architecture of dimensional model is described by “star” schema
Star Schema

- Single fact table and a single table for each dimension
- Dimension tables are denormalized
  - E.g., dimension attributes may be stored multiple times

Example:
100 separate products
5 brands
lots of redundancies
Advantages of Star Schema

- Reduces the number of physical joins
- Simplify the view of the data model
- Allows rel. easy maintenance
Example

**Dimensions:**
- Time, Product, Geography

**Attributes:**
- Product (upc, price, …)
- Geography …
  - …

**Hierarchies:**
- Product → Brand → …
- Day → Week → Quarter
- City → Region → Country

56 units of bread sold in LA on M
Example Cont’d

Time
- Time Code
- Quarter Code
- Quarter Name
- Week Code
- Day Code
- Day name

Account
- Account Code
- Key Account Code
- Account Name
- Account Type
- Account Market

Sales
- Geography Code
- Time Code
- Account Code
- Dollar Amount
- Units

Geography
- Geography Code
- Region Code
- Region Mgr
- City Code
- City Name

Product
- Product Code
- Product Name
- Brand Mgr
- Brand Code
- Prod. Line Code
- Prod. Line Name
- Prod. Name
- ...