Margaret H. Dunham
Southern Methodist University

DATA WAREHOUSING INTRODUCTION
DATA WAREHOUSE OUTLINE

• OLAP

• Data Warehousing

• Dimensional Modeling
Specific data values stored are called the facts (usually numeric)

- minute, second, etc.
- Time - millennium, century, decade, year, month, day, hour

Dimension - collection of logically related attributes viewed as

different values.

View each traditional schema attribute as a dimension with
data

Dimensional modeling different way to view and interrogate

DIMENSIONAL MODELING
DIMENSION EXAMPLE

- Dimensions
  - Time - Year, Month, Day, Hour
  - Location - City, County, Country, Planet
  - Product - ProductID, Type, Company

- Facts: Quantity, Unit Price
<table>
<thead>
<tr>
<th>Unit Price</th>
<th>Quantity</th>
<th>Date</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>10</td>
<td>012000</td>
<td>Chicago</td>
</tr>
<tr>
<td>20</td>
<td>15</td>
<td>022000</td>
<td>Bradenton</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
<td>021500</td>
<td>Rochester</td>
</tr>
<tr>
<td>70</td>
<td>5</td>
<td>030100</td>
<td>Seattle</td>
</tr>
<tr>
<td>75</td>
<td>20</td>
<td>012000</td>
<td>Chicago</td>
</tr>
<tr>
<td>80</td>
<td>5</td>
<td>021000</td>
<td>Fort Worth</td>
</tr>
<tr>
<td>95</td>
<td>5</td>
<td>031500</td>
<td>Dallas</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>031500</td>
<td>Dallas</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>020100</td>
<td>Houston</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>022900</td>
<td>Dallas</td>
</tr>
</tbody>
</table>

**Relational View**

/data warehousing introduction (p. 5)
• Aggregation queries used
  - drill down
  - More specific level query at lower level in dimension granularity
  - drill up
  - More general level query at higher level in dimension granularity
• Partial or total order
• Levels of dimension granularity viewed as hierarchy or DAG.

AGGREGATION HIERARCHIES
ACCGRETATION
View as relations, problem volume of data and indexing:

(Location.City = "Dallas")

and

WHERE (Facts.LocationID = Location.LocationID)

FROM Facts, Location

SELECT Quantity, Price

– Access to fact table from dimension table via join.

– Divide dimension tables.

– Outside of the facts, each dimension is shown separately in

  – Center of the star has facts shown in fact tables.

  – Star schema shows facts and dimensions.

  – Specialized schemas data.

MULTIDIMENSIONAL SCHEMAS
<table>
<thead>
<tr>
<th>Product ID</th>
<th>Location ID</th>
<th>Salesman ID</th>
<th>Quantity</th>
<th>Unit Price</th>
<th>Month</th>
<th>Day</th>
<th>Year</th>
<th>Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>State</td>
<td>Zip Code</td>
<td>Location ID</td>
<td>Salesman ID</td>
<td>Day</td>
<td>Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>State</td>
<td>Zip Code</td>
<td>Location ID</td>
<td>Salesman ID</td>
<td>Day</td>
<td>Year</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**STAR SCHEMA**
DSS use "what if" questions.

- Sales, billing, employee, manufacturing, and warehousing
- Operational Data supports day to day running of the company
- Forecasting
- Informational Data supports functions such as planning and decision making and solving problems.
- Decision Support Systems - tools to assist managers in making decisions
- Data warehouse - data which supports DSS

DATA WAREHOUSING
<table>
<thead>
<tr>
<th>Star/Snowflake</th>
<th>Relational</th>
<th>Data Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutes</td>
<td>Few Seconds</td>
<td>Response</td>
</tr>
<tr>
<td>Less</td>
<td>Often</td>
<td>Access</td>
</tr>
<tr>
<td>Summarized</td>
<td>Detailed</td>
<td>Level</td>
</tr>
<tr>
<td>Terabits</td>
<td>Gigabytes</td>
<td>Size</td>
</tr>
<tr>
<td>Integrated</td>
<td>Operational Values</td>
<td>Data Application</td>
</tr>
<tr>
<td>Business</td>
<td>Application</td>
<td>Orientation</td>
</tr>
<tr>
<td>Static</td>
<td>Dynamic</td>
<td>Modification</td>
</tr>
<tr>
<td>Historical</td>
<td>Snapshot</td>
<td>Temporal</td>
</tr>
<tr>
<td>Aha</td>
<td>Predictive Queries</td>
<td>Use</td>
</tr>
<tr>
<td>OLAP</td>
<td>Application</td>
<td>OLTP</td>
</tr>
</tbody>
</table>

OPERATIONAL DATA VS. DATA WAREHOUSE
Handling missing and erroneous data

New derived data

Summarizing data

Historical view.

Multiple snapshots may need to be merged to create the actual operational data is probably a snapshot of the data.

Converting heterogeneous sources into one common schema.

Unwanted data must be removed.

WAREHOUSE
DATA TRANSFORMATION FOR DATA
Operational data.

Virtual warehouse - warehouse implemented as a view from

Subset of complete data warehouse - data mart.

DATA WAREHOUSE VARIATIONS
Multidimensional view is fundamental.

OLAP tools may be used in DSS systems.

Aggregation functions available in SQL.

Can be thought of as an extension of some of the basic

Require analysis of data.

Support ad hoc querying.

Query results then traditional OLTP or database systems.

OLAP

Online Analytic Processing (OLAP) systems are targeted to

©Margaret H. Dunham
OLAP IMPLEMENTATION

- Updated frequently - RDB
- Not updated frequently - MDD

\textbf{Hybrid OLAP (HOLAP)}•

- Less complex; Less efficient
  for the user.
- ROLAP server (middleware) creates the multidimensional view
- Data stored in a relational database.

\textbf{ROLAP (Relational OLAP)}•

- Indexes used to speed up processing.
- Data stored as an n-dimensional array (cube).
- Data stored as an n-dimensional array (cube).
- Specialized DBMS and software systems capable of supporting
  the multidimensional data directly.
- \textbf{Multidimensional Database (MDD)}.

\textbf{MOLAP (Multidimensional OLAP)}•
OLAP OPERATIONS

(a) Single Cell
(b) Multiple Cells
(c) Slice
(d) Dice

Drill Down
Roll Up
actually "see" results of an operation.

Visualization: These operations allow the OLAP users to

- Drill Down
- Roll Up (Dimension Reduction, Aggregation)
- Dice: Rotate cube to look at another dimension.
- Slice: Look at a subcube to get more specific information.
- Simple query - single cell the cube.

OLAP OPERATIONS
<table>
<thead>
<tr>
<th>Area</th>
<th>Query</th>
<th>Data</th>
<th>Results</th>
<th>Preprocessed</th>
<th>Dimensional Analysis</th>
<th>Vague Preprocessed</th>
<th>Vague Dimensional Analysis</th>
<th>Data Warehouse Integration</th>
<th>Data Warehouse Integration (p. 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>Query</td>
<td>Data</td>
<td>Results</td>
<td>Preprocessed</td>
<td>Dimensional Analysis</td>
<td>Vague Preprocessed</td>
<td>Vague Dimensional Analysis</td>
<td>Data Warehouse Integration</td>
<td>Data Warehouse Integration (p. 20)</td>
</tr>
</tbody>
</table>

**RELATIONSHIP BETWEEN TOPICS**