Chapter 7
Multidimensional Data Modeling (MDDM)
## Learning Objectives and Learning Outcomes

<table>
<thead>
<tr>
<th>Learning Objectives</th>
<th>Learning Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. To assess the capabilities of OLTP and OLAP systems</strong></td>
<td>(a) Appreciate the differences between OLTP and OLAP systems</td>
</tr>
<tr>
<td><strong>2. To understand Dimensional Modeling (DM)</strong></td>
<td>(b) Understanding of DM, basics of data warehousing and related terminology; also an overview of what facts and dimensions are</td>
</tr>
<tr>
<td><strong>3. To learn data warehousing in further detail through a case-study</strong></td>
<td>(c) Understand how to convert an OLTP schema into a dimensional schema model through various techniques of data warehousing</td>
</tr>
</tbody>
</table>
Session Plan

Lecture time : 90 minutes approx.

Q/A : 15 minutes
Agenda

• Database Concepts – Recap
• Introduction to On-line Analytical Processing (OLAP)
• Multidimensional Data Modeling (MDDM)
  – To Answer Why? Where? When? and How?
Recap

• Databases and Tables

• Normalization and Keys
  – ACID Properties
  – Primary, Foreign and Surrogate Keys
  – Cardinality

• Transactions
  – On-Line Transaction Processing
  – On-Line Analytical Processing
Recap (contd.)

- **Difference between OLTP and OLAP**

<table>
<thead>
<tr>
<th></th>
<th>OLTP</th>
<th>OLAP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong></td>
<td>On Line Transaction Processing</td>
<td>On Line Analytical Processing</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>Dynamic (day to day transaction / operational data)</td>
<td>Static (historical data)</td>
</tr>
<tr>
<td><strong>Data Atomicity</strong></td>
<td>Data is stored at microscopic level</td>
<td>Data is aggregated or summarized and stored at the higher level</td>
</tr>
<tr>
<td><strong>Normalization</strong></td>
<td>Normalized Databases to facilitate insertion, deletion and updation</td>
<td>De-normalized Databases to facilitate queries and analysis</td>
</tr>
<tr>
<td><strong>History</strong></td>
<td>Old data is purged or archived</td>
<td>Historical data stored to enable trend analysis and future predictions</td>
</tr>
<tr>
<td><strong>Queries</strong></td>
<td>Simple queries and updates. Queries use small amounts of data</td>
<td>Complex queries. Queries use large amounts of data</td>
</tr>
<tr>
<td></td>
<td>(one record or a few records)</td>
<td>Example: Total annual sales for north region</td>
</tr>
<tr>
<td></td>
<td>Example: update account balance enroll for a course</td>
<td>Total monthly sales for north region</td>
</tr>
</tbody>
</table>

“Fundamentals of Business Analytics”
RN Prasad and Seema Acharya
Copyright © 2011 Wiley India Pvt. Ltd. All rights reserved.
Comparison of OLTP and DSS

OLTP Capability Examples

- Search & locate student(s)
- Print student scores
- Filter students above 90% marks
- Update student Grade
- Group by Batch and compute average score
- Find top 10 high performance students

DSS Capability Examples

- Which courses have productivity impact on-the-job?
- Which colleges need to be rewarded for supplying students with consistent high on-the-job performance?
- What is the customer satisfaction improvement due to extended training?
- How project level profitability is influenced by certification?
- How much training is needed on future technologies for non-linear growth in BI?
Still a little fuzzy about what OLAP can do?
Introduction to On-line Analytical Processing (OLAP)

Scenario:
Internal systems department at Infosys maintains all relevant data in a database. Conceptual schema is as shown below.
OLAP Contd.

• CEO of the company wants the following information from the IS department.

  • Number of employees added in the role of the company during the last quarter/6 months/1 year

Q1. How many table(s) is/are required?

![Employee Table Diagram]
OLAP Contd.

Q2. How many employees are currently in the projects?

Q3. How many are on bench?
Q4. Which customer from a region has given maximum business during the previous quarter on a domain under specific technology and who are the PMs of the project and assets owned by them.
SOLUTION?

MDDM
Answer a Quick Question

So if there were very few updates and more of data-retrieval queries being made on your database, what do you think would be a better schema to adopt?

OLTP or OLAP
Introduction to Dimensional Modeling (DM)

- DM is a logical design technique used in Data Warehouses (DW). It is quite directly related to OLAP systems.
- DM is a design technique for databases intended to support end-user queries in a DW.
- It is oriented around understandability, as opposed to database administration.
However, before we actually jump into MDDM…

let’s first understand the language of Dimensional Modeling
MDDM Terminology

• Grain
• Fact
• Dimension
• Cube
• Star
• Snowflake
Of hierarchies and levels...
What Is a Grain?

• Identifying the grain also means deciding the level of detail that will be made available in the dimensional model

• *Granularity* is defined as the detailed level of information stored in a table

• The more the detail, the lower is the level of granularity

• The lesser the detail, higher is the level of granularity
What can be measured, can be controlled…

…and do you know how such measurements are stored in a data warehouse?
Facts and Fact Tables

- Consists of at least two or more foreign keys
- Generally has huge numbers of records
- Useful facts tend to be numeric and additive
Fact Types

- Additive
- Semi Additive
- Non Additive
- Factless Fact

“Fundamentals of Business Analytics”
RN Prasad and Seema Acharya
Copyright © 2011 Wiley India Pvt. Ltd. All rights reserved.
And what about descriptive data?
What Are Dimensions/Dimension Tables?

• The dimensional tables contain attributes (descriptive) which are typically static values containing textual data or discrete numbers which behave as text values.

• Main functionalities:
  • Query filtering/constraining
  • Query result set labeling
Types of Dimensions

- Slowly Changing Dimension
- Rapidly Changing Dimension
- Degenerate Dimension
- Junk (garbage) Dimension
- Role-playing Dimension
Understanding Dimension – Cube

• An extension to the two-dimensional Table.
• For example in the previous scenario CEO wants a report on revenue generated by different services across regions during each quarter.
Understanding Dimension – Cube (contd.)

Dimension Hierarchy

Grain
Fact

N. America
Europe
Asia Pacific

Q1  Q2  Q3  Q4

Testing
Consulting
Production Support

“Fundamentals of Business Analytics”
RN Prasad and Seema Acharya
Copyright © 2011 Wiley India Pvt. Ltd. All rights reserved.
Can a data warehouse schema take different forms with respect to normalization?
Star Schema

The basic star schema contains four components. These are:

Fact table, Dimension tables, Attributes and Dimension hierarchies
Snow Flake Schema

- Normalization and expansion of the dimension tables in a star schema result in the implementation of a snowflake design.
- A dimension table is said to be snow flaked when the low-cardinality attributes in the dimension have been removed to separate normalized tables and these normalized tables are then joined back into the original dimension table.
Snow Flaking Example

- Consider the Normalized form of Region dimension

Decreases performance because more tables will need to be joined to satisfy queries
Armed with these weapons that we call ‘Concepts’, let’s step into the battlefield!
Case Study
Conversion of a ER Model to a Dimensional Model
Summary

- Basics of Database
- OLTP
- MDDM
- Cube
- Star Schema
- Snowflake schema
Food for Thought!

1. Who according to you would be the user of OLAP?
2. Who would need the Multidimensional perspective of data?