Module 1: Introduction to Data Warehousing and OLAP

Raw Data vs. Business Information
- Capturing Raw Data
  - Gathering data recorded in everyday operations
- Deriving Business Information
  - Deriving meaningful information from raw data
- Turning Data into Information
  - Implementing a decision support system

OLTP Source Systems
- OLTP System Characteristics
  - Processes real-time transactions of a business
  - Contains data structures optimized for entries and edits
  - Provides limited decision support capabilities
- OLTP Examples
  - Order tracking
  - Customer service
  - Point-of-sales
  - Service-based sales
  - Banking functions

Data Warehouse Characteristics
- Provides Data for Business Analysis Processes
- Integrates Data from Heterogeneous Source Systems
- Combines Validated Source Data
- Organizes Data into Non-Volatile, Subject-Specific Groups
- Stores Data in Structures that Are Optimized for Extraction and Querying

Data Warehouse System Components

Defining OLAP Solutions
- OLAP Databases
- Common OLAP Applications
- Relational Data Marts and OLAP Cubes
- OLAP in SQL Server 2000
OLAP Databases

- Optimized Schema for Fast User Queries
- Robust Calculation Engine for Numeric Analysis
- Conceptual, Intuitive Data Model
- Multidimensional View of Data
  - Drill down and drill up
  - Pivot views of data

Common OLAP Applications

- Executive Information Systems
  - Performance measures
  - Exception reporting
- Financial Applications
  - Reporting
  - Planning
  - Analysis
- Sales/Marketing Applications
  - Booking/billing
  - Product analysis
  - Customer analysis
- Operations Applications
  - Manufacturing
  - Customer service
  - Product cost

Relational Data Marts and OLAP Cubes

<table>
<thead>
<tr>
<th>Relational Data Mart</th>
<th>OLAP Cube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Storage</td>
<td>Relational Data Structure</td>
</tr>
<tr>
<td>Data Content</td>
<td>Detailed and Summarized Data</td>
</tr>
<tr>
<td>Data Sources</td>
<td>Relational and Non-relational Sources</td>
</tr>
<tr>
<td>Data Retrieval</td>
<td>Fast Performance for Data Extract Queries</td>
</tr>
</tbody>
</table>

OLAP in SQL Server 2000

- Microsoft Is One of Several OLAP Vendors
- Analysis Services Is Bundled with Microsoft SQL Server 2000
- Analysis Services Include
  - OLAP engine
  - Data mining technology

Understanding Data Warehouse Design

- The Star Schema
- Fact Table Components
- Dimension Table Characteristics
- The Snowflake Schema

The Star Schema
Fact Table Components

<table>
<thead>
<tr>
<th>Dimension Tables</th>
<th>sales_fact Table</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>customer_dim</strong></td>
<td><strong>product_dim</strong></td>
</tr>
<tr>
<td>customer_key</td>
<td>product_key</td>
</tr>
<tr>
<td>quantity_sales</td>
<td>amount_sales</td>
</tr>
</tbody>
</table>

The grain of the sales_fact table is defined by the lowest level of detail stored in each dimension.

Dimension Table Characteristics

- Describes Business Entities
- Contains Attributes That Provide Context to Numeric Data
- Presents Data Organized into Hierarchies

The Snowflake Schema

- Defines Hierarchies by Using Multiple Dimension Tables
- Is More Normalized than a Single Table Dimension
- Is Supported within Analysis Services

Understanding OLAP Models

- OLAP Database Components
- OLAP Dimensions vs. Relational Dimensions
- Dimension Fundamentals
- Dimension Family Relationships
- Cube Measures
- Relational Data Sources

OLAP Database Components

- Numeric Measures
- Dimensions
- Cubes

OLAP Dimensions vs. Relational Dimensions

<table>
<thead>
<tr>
<th>OLAP</th>
<th>Relational</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGION</td>
<td>REGION</td>
</tr>
<tr>
<td>West</td>
<td>West</td>
</tr>
<tr>
<td>CA</td>
<td>East</td>
</tr>
<tr>
<td>OR</td>
<td>MA</td>
</tr>
<tr>
<td>East</td>
<td>NY</td>
</tr>
<tr>
<td>MA</td>
<td>NY</td>
</tr>
<tr>
<td>NY</td>
<td>East</td>
</tr>
</tbody>
</table>
Dimension Fundamentals

**Time**
- Year
  - 1999
  - 2000
  - 2001
- Quarter
  - Q1
  - Q2
  - Q3
  - Q4
- Month
  - Jan
  - Feb
  - Mar

Dimension Family Relationships

- USA is the parent of North West and South West
- North West and South West are children of USA
- North West and California are descendants of USA
- North West and USA are ancestors of Washington
- North West and South West are siblings
- Oregon and California are cousins
- All are dimension members

Cube Measures

- Are the Numeric Values of Principle Interest
- Correspond to Fact Table Facts
- Intersect All Dimensions at All Levels
- Are Aggregated at All Levels of Detail
- Form a Dimension

Relational Data Sources

- Star and Snowflake Schemas
  - Are required to build a cube with Analysis Services
- Fact Table
  - Contains measures
  - Contains keys that join to dimension tables
- Dimension Tables
  - Must exist in same database as fact table
  - Contain primary keys that identify each member

Applying OLAP Cubes

- Defining a Cube
- Querying a Cube
- Defining a Cube Slice
- Working with Dimensions and Hierarchies
- Visualizing Cube Dimensions
- Connecting to an OLAP Cube

Defining a Cube
Defining a Cube Slice

Working with Dimensions and Hierarchies

Visualizing Cube Dimensions

Connecting to an OLAP Cube

Review

- Introducing Data Warehousing
- Defining OLAP Solutions
- Understanding Data Warehouse Design
- Understanding OLAP Models
- Applying OLAP Cubes