Data Warehouse and OLAP II

Week 6
Team Homework Assignment #8

- Using a data warehousing tool and a data set, play four OLAP operations (Roll-up (drill-up), Drill-down (roll down), Slice and dice, Pivot (rotate)) and show the results.
- Due date
  - beginning of the lecture on Friday March 11th.
Typical OLAP Operations

- Roll-up (drill-up)
- Drill-down (roll down)
- Slice and dice
- Pivot (rotate)
- Drill-across
- Drill-through
Roll-up

• Perform aggregation on a data cube by
  – Climbing up a concept hierarchy for a dimension
  – Dimension reduction
Roll-up
Drill-down

- Drill-down is the reverse of roll-up
- Navigates from less detailed data to more detailed data by
  - Stepping down a concept hierarchy for a dimension
  - Introducing additional dimensions
Drill-down
Slice and Dice

• The slice operation performs a selection on one dimension of the given cube, resulting in a sub-cube
• The dice operation defines a sub-cube by performing a selection on two or more dimensions
Slice
Dice

Dice for
(location = “Toronto” or “Vancouver”)
and (time = “Q1” or “Q2”)
and (item = “home entertainment” or “computer”)

Location (cities)

Toronto
Vancouver

Time (quarters)

Q1
Q2

Item (types)

computer
home entertainment

Home entertainment

Time (quarters)

Q1
Q2
Q3
Q4

Item (types)

computer
security
phone

Home entertainment

City

Chicago
New York

1,400

1,500

2,000
Pivot (Rotate)

- Visualization operation that rotate the data axes in view in order to provide an alternative presentation of the data
Pivot
Drill-across

- An additional drilling operation
- Executes queries involving (i.e., across) more than one fact table
Drill-through

- An additional drilling operation
- Uses relational SQL facilities to drill through the bottom level of a data cube down to its back-end relational tables
Figure 3.10. Examples of Typical OLAP operations on multidimensional data cube, commonly used for data warehousing.
Motivation for Building Data Warehouse

- Building and using a data warehouse is a complex, difficult, and long-term task.
- The construction of a large and complex information system can be viewed as the construction of large and complex building.
Data Warehouse Project Process (1)

- Top-down, bottom-up approaches or a combination of both
  - Top-down: Starts with overall design and planning (mature)
  - Bottom-up: Starts with experiments and prototypes (rapid)
Data Warehouse Project Process (2)

• Typical data warehouse design process
  – Choose a business process to model, e.g., orders, invoices, etc.
  – Choose the grain (atomic level of data) of the business process
  – Choose the dimensions that will apply to each fact table record
  – Choose the measure that will populate each fact table record
Three Data Warehouse Models

- **Enterprise warehouse**
  - Collects all of the information about subjects spanning the entire organization

- **Data mart**
  - A subset of corporate-wide data that is of value to a specific groups of users. Its scope is confined to specific, selected groups, such as marketing data mart
    - Independent vs. dependent (directly from warehouse) data mart

- **Virtual warehouse**
  - A set of views over operational databases
  - Only some of the possible summary views may be materialized
Data Warehouse Development: A Recommended Approach

**Figure 3.13** A recommended approach for data warehouse development.
Figure 3.12 A three-tier data warehousing architecture.
OLAP Server Architectures

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)
ROLAP

• Advantages
  – Can handle large amounts of data
  – Can leverage functionalities inherent in the relational database

• Disadvantages
  – Performance can be slow
  – Limited by SQL functionalities
MOLAP

• Advantages
  – Excellent performance
  – Can perform complex calculations

• Disadvantages
  – Limited in the amount of data it can handle
  – Requires additional investment
HOLAP

• HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP.
Data Warehouse Vendors

- IBM
- Microsoft
- Oracle
- Business Objects
Data Warehouse Vendors (cont’d)

• Microstrategy
  – http://www.microstrategy.com/
• Cognos
  – http://www.cognos.com/
• Informatica
  – http://www.informatica.com/
• Actuate
Open Source Data Warehousing Tools

- MySQL-based data warehouse
- Open data warehouse
Data Warehouse Usage (1)

• **Information processing**
  – supports querying, basic statistical analysis, reporting using cross-tabs, tables, charts and graphs

• **Analytical processing**
  – multidimensional analysis of data warehouse data
  – supports basic OLAP operations, slice-dice, drilling, pivoting
Data Warehouse Usage (2)

• Data mining
  – knowledge discovery from hidden patterns
  – Supports associations, constructing analytical models, performing classification and prediction, and presenting the mining results using visualization tools
From OLAP to OLAM

- **On-Line Analytical Mining**
  - High quality of data in data warehouses
  - Available information processing infrastructure surrounding data warehouses
  - OLAP-based exploratory data analysis
  - On-line selection of data mining functions
Figure 3.18: An Integrated OLAM and OLAP Architecture.