A CASE STUDY FOR DATA WAREHOUSING COURSEWARE

A Project

Presented to the faculty of the Department of Computer Science

California State University, Sacramento

Submitted in partial satisfaction of
the requirements for the degree of

MASTER OF SCIENCE

in

Computer Science

by

Shwetha Biligere Prabhuswamy

SPRING
2015
A CASE STUDY FOR DATA WAREHOUSING COURSEWARE

A Project

by

Shwetha Biligere Prabhuswamy

Approved by:

__________________________________, Committee Chair
Dr. Meiliu Lu

__________________________________, Second Reader
Dr. Bill Mitchell

__________________________________
Date
Student: Shwetha Biligere Prabhuswamy

I certify that this student has met the requirements for format contained in the University format manual, and that this project is suitable for shelving in the Library and credit is to be awarded for the project.

__________________________, Graduate Coordinator

Dr. Jinsong Ouyang

Date

Department of Computer Science
Abstract

of

A CASE STUDY FOR DATA WAREHOUSING COURSEWARE

by

Shwetha Biligere Prabhuswamy

Data Warehouse along with Online Analytical Processing (OLAP) are essential elements in making any decisions, which has increasingly become one of the focus of the database industry. Data Warehouse provides an effective way for the analysis of mass data and helps in the decision making process. The objective of this project is to develop a web-based interactive courseware to help data warehouse designers to enhance understanding of the key concepts of OLAP using a case study approach.

The courseware will help users to understand the concepts of OLAP with practical examples. This courseware provides an opportunity for students to generate various summary reports from example data with the help of dropdown list on the web pages. In addition, the students can also work on exercises based on the examples provided in the courseware. This project is developed using MYSQL, PHP, HTML, CSS, and Java Scripts technologies.

_______________________, Committee Chair
Dr. Meiliu Lu

_______________________
Date
I would like to take this opportunity to thank all the people who have helped me walk through this process.

I would like to express my deep and sincere gratitude to my project advisor, Dr. Meiliu Lu for her support and guidance throughout the project. I thank her for giving me an opportunity to work under her on my master’s project. She has been very patient, encouraging and guided me through the entire process. Her detailed feedback was helpful during my project design and development. My sincere thanks to Prof. Bill Mitchell for being my second reader. I also take this opportunity to thank Dr. Jinsong Ouyang for his review of the project.

I would like to thank my husband Mr Chetan Patil, without his support and understanding, this project would have been impossible. I extend my thanks to all those who have helped me directly or indirectly in the completion of this project. Last but not the least; thanks are to the Almighty for all the blessings.
Table of Contents

Acknowledgments................................................................................................................v

List of Tables ................................................................................................................... viii

List of Figures .................................................................................................................... ix

Chapter

1. INTRODUCTION ...........................................................................................................1

1.1 Background................................................................................................................2

1.2 Need and Scope of the courseware ........................................................................4

2. PROJECT DEVELOPMENT STAGES ..........................................................................6

2.1 Research and Analysis..............................................................................................6

2.2 Design and Development.........................................................................................6

3. ROAD MAP FOR DATAWAREHOUSING COURSEWARE .....................................9

3.1 Multidimensional Data Model ...............................................................................10

3.2 The Relational Implementation of the Model .........................................................12

3.3 Data Cube.................................................................................................................13

3.4 On-line Analytical Processing (OLAP) .................................................................15

3.5 OLAP Categories ....................................................................................................17

3.6 OLAP Operations ....................................................................................................23

3.7 Examples of OLAP Operations .............................................................................27

3.8 Exercises ..................................................................................................................41

4. CONCLUSION ..............................................................................................................43
Appendix............................................................................................................................44
References..........................................................................................................................99
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1: Difference between OLAP and OLTP</td>
<td>3</td>
</tr>
<tr>
<td>Table 2: Two-dimensional representation of data mart</td>
<td>14</td>
</tr>
<tr>
<td>Table 3: Three-dimensional representation of data model</td>
<td>15</td>
</tr>
<tr>
<td>Table 4: Data in Single table before creating OLAP cube</td>
<td>30</td>
</tr>
<tr>
<td>Table 5: Sample data in Books Name Table</td>
<td>31</td>
</tr>
<tr>
<td>Table 6: Data in Book Categories</td>
<td>31</td>
</tr>
<tr>
<td>Table 7: Data in Distributor Table</td>
<td>32</td>
</tr>
<tr>
<td>Table 8: Sample Data in Book Publisher</td>
<td>32</td>
</tr>
<tr>
<td>Table 9: Sample table in Time Table</td>
<td>32</td>
</tr>
<tr>
<td>Table 10: Sample data in Comic Book Fact Table</td>
<td>33</td>
</tr>
<tr>
<td>Table 11: Results of Roll up OLAP Operation</td>
<td>35</td>
</tr>
<tr>
<td>Table 12: Results of OLAP Slice operations</td>
<td>39</td>
</tr>
<tr>
<td>Table 13: Results of OLAP Dice operation</td>
<td>41</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figures ........................................ Page

Figure 1 : Data warehouse and OLAP Cube ................................................................. 4
Figure 2: Flow Chart for Development of the Courseware ........................................ 7
Figure 3 : A Case Study for Data Warehousing Courseware ...................................... 9
Figure 4: Shows the relationships among the basic elements .................................. 10
Figure 5: Levels and Hierarchy relation in a cube ...................................................... 12
Figure 6: Sample representation of Star Schema ...................................................... 13
Figure 7: OLAP Architecture ..................................................................................... 16
Figure 8: ROLAP Server Architecture .................................................................... 18
Figure 9: MOLAP Server Architecture .................................................................... 19
Figure 10: Hybrid OLAP Architecture ..................................................................... 20
Figure 11: DOLAP Architecture .............................................................................. 23
Figure 12: Rollup and Dice operations ................................................................. 25
Figure 13 : Slice and Dice Operations ................................................................... 26
Figure 14 : OLAP Pivot Operations ........................................................................ 27
Figure 15 : Star Schema ........................................................................................... 28
Figure 16 : Web page to generate Roll up operation .............................................. 34
Figure 17 : Web page to generate Drill down results ............................................. 36
Figure 18 : Results of Drill down OLAP operations ............................................... 38
Figure 19: Exercise Page ......................................................................................... 42
Chapter 1

INTRODUCTION

A Data Warehouse is designed for query and analysis rather than for transaction processing. According to William Inmon Data Warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process [1].

- **Subject-Oriented**: A data warehouse can be used to analyze particular subject areas. For example, sales data of retail shop would be subject area.
- **Integrated**: A data warehouse integrates data from multiple data sources.
- **Time-Variant**: Data Warehouse typically stores historical data. For example, one can retrieve data from 3 months, 6 months, 12 months, or even older data from a data warehouse. This contrasts with a transactions system, where daily inserted and/or updated data is stored.
- **Non-volatile**: Once data is in the data warehouse, it will not have frequent updates.

Every organization, small or big, requires exploitation of a large scale of chronological data. An analytical prediction model for this data can help management functions such as decision-making and planning. This data helps analysts to take informed decisions in an organization. Consider a business executive of a Super Mart wants to analyze the trends of products sold region wise over a period and make decisions to offer coupons on his products. Since transactional database changes on daily basis, they cannot rely. In such situations, we need a separate database system that maintains historical data. A data
warehouse is a database containing data that usually represents the business history of an organization. It is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, and analyst) to make better and faster decisions. It provides architecture and tools for business executives to systematically organize, understand and use their data to make strategic decisions.

1.1 Background

A data warehouse (DW) is an integral part of many information delivery systems because it contains consolidated data from several operational databases and other data sources where uploaded. With a large data warehouse, query throughput and response times are very important. To facilitate these complex analyses data warehouses also provides Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in a multidimensional space.

In general, Data can be modeled by two ways On Line Transaction Processing (OLTP) and Online Analytical Processing (OLAP). Online Transaction Processing, or OLTP, is a class of information systems that facilitates and manages transaction-oriented applications, typically for data entry and retrieval transaction processing. On the other hand, OLAP deals with large amount of historical data, answers multidimensional queries and provides an approach for users to view data in different dimensions. Differences between OLAP and OLTP are listed in Table 1.
Table 1: Difference between OLAP and OLTP

<table>
<thead>
<tr>
<th>Features</th>
<th>Traditional Data Base (OLTP)</th>
<th>Data Warehouse (OLAP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
<td>It is based on Operational Processing.</td>
<td>It is based on Informational Processing.</td>
</tr>
<tr>
<td>Data</td>
<td>It mainly stores the current data that always guaranteed to be up-to-date.</td>
<td>It usually stores the Historical data whose accuracy is maintained over time.</td>
</tr>
<tr>
<td>Read/write</td>
<td>The most frequent type of access type is read/write.</td>
<td>It mostly use the read access for the stored data.</td>
</tr>
<tr>
<td>User Function</td>
<td>The common users are clerk, DBA, database professional.</td>
<td>The common users are knowledge worker (e.g., manager, executive, and analyst).</td>
</tr>
<tr>
<td>Data base design</td>
<td>The designing of database is ER based and application-oriented.</td>
<td>The designing is typically done using star or snowflake or hybrid schema and its subject-oriented</td>
</tr>
<tr>
<td>Summarization</td>
<td>The data is primitive and highly detailed.</td>
<td>The data is summarized and in consolidated form.</td>
</tr>
<tr>
<td>View</td>
<td>The view of the data is flat relational.</td>
<td>The view of the data is multidimensional</td>
</tr>
<tr>
<td>Function</td>
<td>It is used for day-to-day operations</td>
<td>It is used for trend analysis.</td>
</tr>
</tbody>
</table>
OLAP technology enables data warehouses to be used effectively for online analysis, providing rapid responses to iterative complex analytical queries. OLAP's multidimensional data model and data aggregation techniques organize and summarize large amounts of data so it can be evaluated quickly using online analysis and graphical tools. Data mining functions such as association, clustering, classification, prediction can be integrated with OLAP operations to enhance the interactive mining of knowledge at multiple level of abstraction. These are the data warehouse has now become an important platform for data analysis and online analytical processing.

Figure 1 demonstrates relation between DW and OLAP [2]

1.2 Need and Scope of the courseware

The main objective of this courseware is to help students to understand how the OLAP methods and tools can be used to perform multidimensional analysis of data with the help
of an example. Courseware illuminates basic concepts and design principles of data warehousing. The tool supports the course material using illustrative examples, interactive demonstrations and visual diagrams for the topic explanation. This generates interests and insight among students during learning process. The students can assess their understanding of OLAP concepts via exercises provided at the end of courseware. The case study uses the comic books sales data of Diamond distributors as example data. In the courseware, we demonstrate steps to build a data warehouse for the sales data. This tool not only illustrates the data warehousing design process but also demonstrates OLAP operations on designed data marts. Our case study includes examples for every OLAP operations and demonstrated using open sample queries.
Chapter 2

PROJECT DEVELOPMENT STAGES

The project development took place in two phases:

2.1 RESEARCH AND ANALYSIS

Before developing the courseware considerable time was spent to understand concepts of DW and OLAP. I have enhanced my knowledge on Data Warehouse through books and research papers [3, 4 and 5]. Along with this, I had to read a plenty of articles online and do a lot of research on these technologies, which are complex, understand their working. All this effort made to move ahead with confidence in designing and implementing the courseware. To implement OLAP operations Comic books sales data was obtained from Icv2 [6]. This data set includes sales data of Top 300 comic books sold by Diamond Distributors

2.2 DESIGN AND DEVELOPMENT

The project development mainly consists of the following six stages as shown in Figure 2.

1) Top 300 Comic books sales data of Diamond Distributors is downloaded from Icv2[6]. The data for each year downloaded separately. I downloaded the data in .csv format in Excel files so that loading the data in the database would be easy. The data underwent some preprocessing before loading into the database using Microsoft Excel functions.
Figure 2: Flow Chart for Development of the Courseware

2) With the help MySQL workbench database and tables are created, the data from the Excel files are inserted into each individual table with the help of MYSQL server using below command.

LOAD DATA LOCAL INFILE './gaia/class/student/biligers/publisher.csv' INTO TABLE Publisher.

3) Data mart is segment of Data warehouse that provides data for reporting and analysis. To demonstrate the examples in the courseware Data Mart created for Diamond distributors, which keep track of Top 300 comic books sold every year. It provides
information about comic book titles, publishers and total sales on monthly basis. Star
schema concept is used to design a data mart. In star schema, single dimensional table
represents each dimension. The project uses five Dimension tables and one fact table.
The fact table represents the monthly sales, while the dimension tables provide detail
information about publishers, comic books name, time, book categories and
distributors. OLAP cube uses Star Schema to represent data. Once the structure of the
cube is defined, the cube allows us to perform various calculations and computations
on the data.

4) In the courseware examples SQL queries constructed to generate various reports to
demonstrate OLAP operations.

5) Finally, the user interface (UI) webpages are developed using HTML and CSS,
PHP, CSS, JAVA SCRIPT and by using PHP data connection is established between
database and web pages. The code for obtaining the configuration is:
<?php
$link = mysql_connect('localhost', 'mysql_user', 'mysql_password');
if (!$link) {
    die('Could not connect: ' . mysql_error());
}
echo 'Connected successfully';
mysql_close($link);?>
Chapter 3

ROAD MAP FOR DATAWAREHOUSING COURSEWARE

This chapter describes the Data Warehouse courseware in detail. Courseware is organized into six main tabs that covers all the information about OLAP : “Introduction”, “A multidimensional data model”, “Data cube”, “Introduction to OLAP”, “OLAP operations”, “Examples” and “Exercises”. The content of the courseware includes visual diagrams, examples and supporting information. To make students familiar with OLAP operations user interactive web pages are designed where students can generate various reports with the help of dropdown lists. Data from generated reports are displayed in tables on web pages, so that the users are able to view the data in a neat tabular form.

Figure 3 below shows the overview of the courseware.

![A Case Study for Data Warehousing Courseware](image)

**Introduction**

A data warehouse (DW) is an approach for creating an enterprise-wide data store. It is an integral part of many information delivery systems because it contains consolidated data, obtained from several operational databases and other data sources, over long periods of time. With a large size data warehouse, query throughput and response times are very important. To facilitate these complex analyses data warehouses also provides us Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in a multidimensional space.

The below figure help to understand the relationship between Data Warehouse and online analytical processing (OLAP) cube

**Figure 3 : A Case Study for Data Warehousing Courseware**
3.1 MULTIDIMENSIONAL DATA MODEL

This section covers a multidimensional data model and how it is implemented in tables. Data warehouses and OLAP tools are based on a multidimensional data model. This model views data in several forms, one of which is a data cube. The multidimensional data model is composed of some basic elements like logical cubes, measures, dimensions, hierarchies, levels, and attributes are demonstrated in Figure 4.

![Figure 4: Shows the relationships among the basic elements [7]](image)

Dimensions

Dimensions are the perspectives or entities with respect to which an organization wants to keep record. For example if user wants to keep track of comic books that are published in a year 2000 with respect to dimensions, where dimension are books categories, year,
Each dimension may have a table associated with it called a dimension table.

**Hierarchy**

A hierarchy is a way to organize data at different levels of aggregation. In viewing data, analysts use dimension hierarchies to recognize trends at one level, drill down to lower levels to identify reasons for these trends, and roll up to higher levels to see what affect these trends have on a larger sector of the business.

**Level**

It is a column within a dimension table that can be used for aggregating data. For example, a product dimension can have levels of product type (beverage), product category (alcoholic beverage), product class (beer), product name (miller lite, budlite, corona, etc.).

**Cube**

A cube is a logical organization of multidimensional data. A cube is derived from a fact table. Dimensions categorize a cube’s data and a cube contains measures that share the same dimensionality. Cubes are not usually exposed to end-users since they are more interested in the measures.

**Measures**

Measures are numeric representations of a set of facts that have occurred. Examples of measures include dollars of sales, number of credit hours, store profit percentage, dollars of operating expenses etc.
Figure 5 shows the relation between levels, hierarchy and dimensions.

![Figure 5: Levels and Hierarchy relation in a cube](image)

### 3.2 The Relational Implementation of the Model

The logical design of the multidimensional data model is typically a star schema, or a snowflake schema. This model helps in organizing the data into dimension tables, fact tables and materialized views.

**Star Schema**

Star schema is the simplest form of a dimensional model, in which data is organized into facts and dimensions. A fact is an event that is counted or measured. A dimension contains reference information about the fact such as date, product, or customer. Each fact table with its associated dimensions surrounds a star schema shown in the Figure 6.
Figure 6: Sample representation of Star Schema.

Snow Flake Schema

The snowflake schema consists of one fact table that is connected to many dimension tables, which can be connected to other dimension tables through a many-to-one relationship. Tables in a snowflake schema are usually normalized to the third normal form. Each dimension table represents exactly one level in a hierarchy.

3.3 DATA CUBE

In this section we will discuss about how the multidimensional cube represent on data mart. A data cube is a type of multidimensional matrix that lets users explore and analyze a collection of data from different perspectives. The cube is used to represent data along some measure of interest. Although called a cube, it can be 2-dimensional, 3-dimensional, or higher-dimensional. Each dimension represents some attribute in the database and the cells in the data cube represent the measure of interest. For example,
they could contain a count for the number of times that attribute combination occurs in the database, or the minimum, maximum, sum or average value of some attribute. Queries are performed on the cube to retrieve decision support information. The example in Table 2 demonstrates two dimensional representation of data mart.

Consider comic books distributor company sales fact table in North California region.

Table 2: Two-dimensional representation of data mart

<table>
<thead>
<tr>
<th>Title</th>
<th>year</th>
<th>Est sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Avengers</td>
<td>2004</td>
<td>73,362</td>
</tr>
<tr>
<td>The Avengers</td>
<td>2005</td>
<td>No Data</td>
</tr>
<tr>
<td>The Avengers</td>
<td>2006</td>
<td>150,000</td>
</tr>
<tr>
<td>The Avengers</td>
<td>2007</td>
<td>159,141</td>
</tr>
<tr>
<td>Amazing Spider-Man</td>
<td>2004</td>
<td>123,540</td>
</tr>
<tr>
<td>Amazing Spider-Man</td>
<td>2005</td>
<td>112,564</td>
</tr>
<tr>
<td>Amazing Spider-Man</td>
<td>2006</td>
<td>No Data</td>
</tr>
<tr>
<td>Amazing Spider-Man</td>
<td>2007</td>
<td>Pending</td>
</tr>
</tbody>
</table>

Multi-dimensional databases are a compact and easy way of visualizing and manipulating data elements that have many inter-relationships. If the cube is expanded to include another dimension, for example, sales fact compared with South California region then the cube is viewed in Table 3.
3.4 On-line Analytical Processing (OLAP)

OLAP systems are part of decision support systems and will assist analysts and managers, those who are responsible for the smooth running of an organization by giving them quick access to data. OLAP tools provide users with a fast response even if the query request is made on a large volume of data. Basically, OLAP tools provide the ability to transform huge volumes of data that exist in the organization into useful information to support decision-making process.

**OLAP Architecture**

OLAP systems have a structured architecture based on three essential components as shown in Figure 7.
**Data warehouse tier**

This layer deals with preparing data for OLAP analysis. It collects the data from various data source (relational databases, files, csv, etc), processes, transforms, loads the data into fact, and dimensions tables on different levels. Data Warehouse tier includes following steps

- Extracting data from multiple operational databases and external sources
- Cleaning, transforming, and integrating the data
- Loading data into the data warehouse
• Periodically refreshing the data warehouse to reflect updates at the source and to purge data from the data warehouse

**OLAP server tier**

It manages multidimensional data structure and at the same time it links between the Data warehouse and OLAP customer.

**Client tier**

It provides data mining applications and involve in report generation.

### 3.5 OLAP Categories

OLAP tools are categorized according to the architecture used to store and process multidimensional data. There are four main categories of OLAP tools. The four main categories of OLAP tools are listed below.

**Relational OLAP (ROLAP)**

ROLAP is the fastest-growing type of OLAP tools. ROLAP can handle large volumes of data, all data resides in the relational database management system where relational tables are optimized for low-level dimensional requests, and aggregate indexes are created for higher-level OLAP requests [8]. Figure 8 shows the ROLAP server architecture.
Advantages of ROALP

- ROLAP supports RDBMS products with a metadata layer, thus avoiding the requirement to create a static multi-dimensional data structure.
- This facilitates the creation of multiple multi-dimensional views of the two-dimensional relation.
- To improve performance, some ROLAP products have enhanced SQL engines to support the complexity of multi-dimensional analysis, while others recommend, or require, the use of highly deormalized database designs such as the star schema.
- Data management remains within the RDBMS, not within the cube.

Disadvantages of ROLAP

- Performance problems associated with the processing of complex queries that require multiple passes through the relational data.
- Development of middleware to facilitate the development of multi-dimensional applications.
Multi-dimensional OLAP (MOLAP)

MOLAP tools use specialized data structures and multi-dimensional database management systems (MDDBMS) to organize, navigate, and analyze data. MOLAP data structures use array technology and efficient storage techniques that minimize the disk space requirements through sparse data management [8]. Figure 9 shows MOLAP server architecture.

Advantages of MOLAP

- Excellent performance: MOLAP cubes are built for fast data retrieval, and is optimal for slicing and dicing operations
- Can perform complex calculations: All calculations have been pre-generated when the cube is created. Hence, complex calculations are not only doable, but they return quickly.

Disadvantages of MOLAP

- Limited for data it can handle because all calculations are performed when the cube is built, it is not possible to include a large amount of data in the cube itself.
This is not to say that the data in the cube cannot be derived from a large amount of data. Indeed, this is possible. In this case, only summary-level information will be included in the cube itself.

- Requires additional investment: Cube technology are often proprietary and do not exist in the organization. Therefore, to adopt MOLAP technology, chances of additional investments in human and capital resources are needed.

**Hybrid OLAP (HOLAP)**

Hybrid Online Analytical Processing (HOLAP) is a combination of MOLAP and ROLAP. HOLAP stores the detail data in the relational database but stores the aggregations in multidimensional format. With HOLAP, we will have medium query performance not as slow as ROLAP, but not as fast as MOLAP. However, you were only querying aggregated data or using a cached query, query performance would be similar to MOLAP but when you need to get that detail data, performance is closer to ROLAP [8]. Figure 10 shows the HOLAP server architecture.

![Figure 10: Hybrid OLAP Architecture](image)
Advantages of HOLAP

- HOLAP is best used when large amounts of aggregations are queried often with little detail data, offering high performance and lower storage requirements.
- Cubes are smaller than MOLAP since the detail data is kept in the relational database.
- Processing time is less than MOLAP since only aggregations are stored in multidimensional format.
- Low latency since processing takes place when changes occur and detail data is kept in the relational database.

Disadvantages of HOLAP

- The architecture results in significant data redundancy and may cause problems for networks that support many users.
- Ability of each user to build a custom data cube may cause a lack of data consistency among users.
- Only a limited amount of data can be efficiently maintained.

Desktop OLAP (DOLAP)

Desktop OLAP or DOLAP is based on the idea that a user can download a section of the data from the database or source, and work with that dataset locally, or on their desktop. DOLAP is easier to deploy and has a cheaper cost but comes with a very limited functionality in comparison with other OLAP applications. Figure 9 shows the DOLAP server architecture.
Advantages of DOALP

- DOLAP tools store the OLAP data in client-based files and support multi-dimensional processing using a client multi-dimensional engine. DOLAP requires that relatively small extracts of data are held on client machines.
- The administration of a DOLAP database is typically performed by a central server or processing routine that prepares data cubes or sets of data for each user.

Disadvantages of DOLAP

- Provision of appropriate security controls to support all parts of the DOLAP environment.
- Reduction in the effort involved in deploying and maintaining the DOLAP tools.
3.6 OLAP Operations

OLAP provides user with the flexibility to view data from different perspectives. Hence OLAP operations are discussed on multidimensional data.

**ROLL UP**

- A roll-up involves summarizing the data along a dimension.
• The roll-up operation is performed by climbing up a concept hierarchy for the dimension location.

• When roll-up operation is performed one or more dimensions from the data cube are removed.

**Drill-down**

Drill down is the reverse of roll-up. Navigates from less detailed data to more detailed data it can achieved by any of the following way.

• Stepping down a concept hierarchy for a dimension.

• Introduces additional dimensions.

The Figure 12 shows demo of Roll up and Drill down operations.
Figure 12: Rollup and Dice operations

Slice

- Performs a selection on one dimension of the given cube, resulting in a sub-cube.
- Reduces the dimensionality of the cubes.
- Sets one or more dimensions to specific values and keeps a subset of dimensions for selected values.

Dice

- Define a sub-cube by performing a selection of one or more dimensions.
• Refers to range select condition on one dimension, or to select condition on more than one dimension.

• Reduces the number of member values of one or more dimensions.

The Figure 13 shows demo of slice and Dice operations

![Slice and Dice Operations](image)

**Figure 13 : Slice and Dice Operations**

**Pivot (or rotate)**

• Rotates the data axis to view the data from different perspectives.

• Groups data with different dimensions.
The Figure 14 shows demo of OLAP Pivot operations.

![OLAP Pivot Operations Diagram](image)

**Figure 14 : OLAP Pivot Operations**

**Some more OLAP operations are listed below**

**Drill-across**

- An additional drilling operation.
- Executes queries involving 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.

**3.7 Examples of OLAP Operations**

In our courseware, we are using Comic books sales data of a Diamond Distributors to illustrate OLAP operations. Here we are considering data of Top 300 comic books sold in every year. To demonstrate this example I have created five dimension tables and one fact table. Fact table referred to as a cube and the columns within the table are referred to
as measures. Cube has edges, which are referred to as dimensions. Figure 15 shows Sample representation of Strat Schema.

**Dimension tables**

- Books Name
- Publisher
- Books Categories
- Distributors
- Time

**Fact table**

Comic Books Fact table

![Star Schema Diagram](image)

Figure 15 : Star Schema
Star Schema is the simplest style of data mart schema. It consists of one dimension table along with any numbers of dimension table. In above star schema fact table contains sales data of Top300 Comic books sold in every month. It also includes a set of columns that form a concatenated or composite key. Each column of the concatenated key is a foreign key drawn from a dimensional table primary key. Each row in a fact table must contain a primary key value from each dimension table. This rule is called referential integrity and is an important requirement in decision-support databases. In star schema, referential integrity is maintained to ensure valid query results.

**Star Schema Advantages**

The Main advantages behind using star schema are

- **Query Performance**
  Queries run faster against a star schema database than an OLTP system because of the clear join paths it retrieves only necessary rows from database.

- **Load Performance**
  The star schema structure reduces the time required to load large batches of data into a database. By defining facts and dimensions and separating them into different tables, the impact of a load operation is reduced. Dimension tables can be populated once and occasionally refreshed. New facts can be added regularly and selectively by appending records to a fact table.

- **Easily understood**
  Structure of schema is very simple to understand. Navigating through data is efficient because dimensions are joined through fact tables. These joins are
significant because they represent fundamental relationships of data model. We can navigate to a single dimension table in order to select attribute values to construct an efficient query.

Before creating data mart a sample data in the single table is shown in Table 4.

**Table 4: Data in Single table before creating OLAP cube**

<table>
<thead>
<tr>
<th>Book_ID</th>
<th>Book_Name</th>
<th>Issue</th>
<th>Price</th>
<th>Publisher</th>
<th>Sales</th>
<th>Year</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Uncanny X-Men</td>
<td>378</td>
<td>1.99</td>
<td>Marvel</td>
<td>113703</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>2</td>
<td>X-Men</td>
<td>98</td>
<td>1.99</td>
<td>Marvel</td>
<td>109676</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>3</td>
<td>Wolverine</td>
<td>148</td>
<td>1.99</td>
<td>Marvel</td>
<td>87852</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>4</td>
<td>JLA</td>
<td>39</td>
<td>1.99</td>
<td>DC</td>
<td>78273</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>5</td>
<td>Avengers</td>
<td>26</td>
<td>1.99</td>
<td>Marvel</td>
<td>76535</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>6</td>
<td>Spawn</td>
<td>94</td>
<td>1.95</td>
<td>Image</td>
<td>73517</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>7</td>
<td>Soul Saga</td>
<td>1</td>
<td>2.5</td>
<td>Image</td>
<td>66273</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>8</td>
<td>Earth X</td>
<td>11 (Res)</td>
<td>2.99</td>
<td>Marvel</td>
<td>64788</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>9</td>
<td>Tomb Raider</td>
<td>3</td>
<td>2.5</td>
<td>Image</td>
<td>63902</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>10</td>
<td>X-Men</td>
<td>26</td>
<td>2.99</td>
<td>Marvel</td>
<td>62057</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>11</td>
<td>Fantastic Four</td>
<td>27</td>
<td>1.99</td>
<td>Marvel</td>
<td>58575</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>12</td>
<td>Cable</td>
<td>77</td>
<td>1.99</td>
<td>Marvel</td>
<td>57214</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>13</td>
<td>Amazing</td>
<td>15</td>
<td>1.99</td>
<td>Marvel</td>
<td>56413</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>14</td>
<td>Spider-Man</td>
<td>21</td>
<td>1.99</td>
<td>Marvel</td>
<td>55965</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>15</td>
<td>Thor</td>
<td>15</td>
<td>1.99</td>
<td>Marvel</td>
<td>54318</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>16</td>
<td>Iron Man</td>
<td>26</td>
<td>1.99</td>
<td>Marvel</td>
<td>52590</td>
<td>2006</td>
<td>Jan</td>
</tr>
<tr>
<td>17</td>
<td>Batman Gotham Knights</td>
<td>1</td>
<td>2.5</td>
<td>DC</td>
<td>51264</td>
<td>2006</td>
<td>Jan</td>
</tr>
</tbody>
</table>

Data from single table is processed and created Star schema to perform OLAP operations

**Sample Data in the data mart**

The Table 5 shows sample data in Books Name table.
• **Book Name**

**Table 5: Sample data in Books Name Table**

<table>
<thead>
<tr>
<th>Book_id</th>
<th>Title Name</th>
<th>Issuance</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1111</td>
<td>Uncanny X-Men</td>
<td>378</td>
<td>1.99</td>
</tr>
<tr>
<td>1112</td>
<td>X-Men</td>
<td>98</td>
<td>1.99</td>
</tr>
<tr>
<td>1113</td>
<td>Wolverine</td>
<td>148</td>
<td>1.99</td>
</tr>
<tr>
<td>1114</td>
<td>JLA</td>
<td>39</td>
<td>1.99</td>
</tr>
<tr>
<td>1115</td>
<td>Avengers</td>
<td>26</td>
<td>1.99</td>
</tr>
<tr>
<td>1116</td>
<td>Spawn</td>
<td>94</td>
<td>1.95</td>
</tr>
<tr>
<td>1117</td>
<td>Soul Saga</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>1118</td>
<td>Earth X</td>
<td>11 (Res)</td>
<td>2.99</td>
</tr>
<tr>
<td>1119</td>
<td>Tomb Raider</td>
<td>3</td>
<td>2.5</td>
</tr>
<tr>
<td>1120</td>
<td>X-Men Unlimited</td>
<td>26</td>
<td>2.99</td>
</tr>
<tr>
<td>1121</td>
<td>Fantastic Four</td>
<td>27</td>
<td>1.99</td>
</tr>
<tr>
<td>1122</td>
<td>Cable</td>
<td>77</td>
<td>1.99</td>
</tr>
<tr>
<td>1123</td>
<td>Amazing Spider-Man</td>
<td>15</td>
<td>1.99</td>
</tr>
<tr>
<td>1124</td>
<td>Thor</td>
<td>21</td>
<td>1.99</td>
</tr>
<tr>
<td>1125</td>
<td>Peter Parker Spide</td>
<td>15</td>
<td>1.99</td>
</tr>
</tbody>
</table>

• **Book Categories**

Sample data in Book_categories table is show in Table 6.

**Table 6: Data in Book Categories**

<table>
<thead>
<tr>
<th>Book_category_id</th>
<th>Book_category_Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Comic Books</td>
</tr>
<tr>
<td>102</td>
<td>Graphics Novel</td>
</tr>
</tbody>
</table>

• **Distributor**

Sample data in Distributor table shown in Table 7.
Table 7: Data in Distributor Table

<table>
<thead>
<tr>
<th>Distributor_id</th>
<th>Distributor_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>100001</td>
<td>Diamond distributor</td>
</tr>
</tbody>
</table>

- Book Publisher

Sample data in Book Publisher table shown in Table 8.

Table 8: Sample Data in Book Publisher

<table>
<thead>
<tr>
<th>Publisher Name</th>
<th>Publisher_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 Gauge</td>
<td>2222</td>
</tr>
<tr>
<td>3 Finger Prints</td>
<td>2223</td>
</tr>
<tr>
<td>360ep</td>
<td>2224</td>
</tr>
<tr>
<td>3BP</td>
<td>2225</td>
</tr>
<tr>
<td>88 MPH</td>
<td>2226</td>
</tr>
<tr>
<td>A Silent</td>
<td>2227</td>
</tr>
<tr>
<td>AAA Milwaukee</td>
<td>2228</td>
</tr>
</tbody>
</table>

- Time

Sample data in Time table shown in Table 9.

Table 9: Sample table in Time Table

<table>
<thead>
<tr>
<th>ID</th>
<th>Year</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>2008</td>
<td>Sept</td>
</tr>
<tr>
<td>106</td>
<td>2008</td>
<td>Oct</td>
</tr>
<tr>
<td>107</td>
<td>2008</td>
<td>Nov</td>
</tr>
<tr>
<td>108</td>
<td>2008</td>
<td>Dec</td>
</tr>
<tr>
<td>109</td>
<td>2009</td>
<td>Jan</td>
</tr>
<tr>
<td>110</td>
<td>2009</td>
<td>Feb</td>
</tr>
<tr>
<td>111</td>
<td>2009</td>
<td>Mar</td>
</tr>
<tr>
<td>112</td>
<td>2009</td>
<td>Apr</td>
</tr>
<tr>
<td>113</td>
<td>2009</td>
<td>May</td>
</tr>
<tr>
<td>114</td>
<td>2009</td>
<td>June</td>
</tr>
<tr>
<td>115</td>
<td>2009</td>
<td>July</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• Comic Book Fact Table

Sample data in Comic_Book_Fact table is shown in Table 10.

Table 10 : Sample data in Comic Book Fact Table

<table>
<thead>
<tr>
<th>ID</th>
<th>Book_ID</th>
<th>Publisher_id</th>
<th>Time_id</th>
<th>Distributor_id</th>
<th>Book_Category_id</th>
<th>Est_sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>11111111</td>
<td>1111</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>113,703</td>
</tr>
<tr>
<td>11111112</td>
<td>1112</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>109,676</td>
</tr>
<tr>
<td>11111113</td>
<td>1113</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>87,852</td>
</tr>
<tr>
<td>11111114</td>
<td>1114</td>
<td>2353</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>78,273</td>
</tr>
<tr>
<td>11111115</td>
<td>1115</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>76,535</td>
</tr>
<tr>
<td>11111116</td>
<td>1116</td>
<td>2439</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>73,517</td>
</tr>
<tr>
<td>11111117</td>
<td>1117</td>
<td>2439</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>66,273</td>
</tr>
<tr>
<td>11111118</td>
<td>1118</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>64,789</td>
</tr>
<tr>
<td>11111119</td>
<td>1119</td>
<td>2439</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>63,902</td>
</tr>
<tr>
<td>11111120</td>
<td>1120</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>62,097</td>
</tr>
<tr>
<td>11111121</td>
<td>1121</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>58,575</td>
</tr>
<tr>
<td>11111122</td>
<td>1122</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>57,214</td>
</tr>
<tr>
<td>11111123</td>
<td>1123</td>
<td>2465</td>
<td>1</td>
<td>100001</td>
<td>101</td>
<td>56,413</td>
</tr>
</tbody>
</table>

Roll up

The Roll Up analytical operation performed by navigating up a dimensional hierarchy to a more summarized level. Here by using OLAP Roll up operation on Star schema [Figure 11] we can find contributions of each publisher in Top 300 comic books sold in a year by Diamond distributor. To demonstrate an interactive example I have created drop down list to generate Roll up operation results on OLAP cube. Figure 16 shows web page of Roll up operation.
Rollup operation

In our example we are considering yearly sales data of Top300 comic books of Diamond Distributors. By computing OLAP Roll up operation on Star schema we are calculating the contribution of each publishers in Top 300 comic books sold every year.

Diamond distributors sales data report can be viewed using Roll up operation.

* required field

Distributors: Select Distributors

Books Categories: Select Categories

Year: Select Year

We can generate Roll up operation results by using below query.

Select bp.Publisher_Name , count(bp.Publisher_Name) from Comic_Books_fact cbf, Book_publisher bp, Time t, Distributor d, Book_categories bc where cbf.Distributor_id = d.Distributor_id and t.Time_id = cbf.Time_id and cbf.Book_catgry_id = bc.Book_catgry_id and t.year = 2006 and cbf.publisher_id = bp.Publisher_id group by bp.Publisher_Name;
In the above query, select operation will select respective year, publisher name dimensions and count (Publisher_Name) function will calculate total number of times publisher name repeated in a sales data. Group by function is used for grouping the result set by publisher name. The Table 11 shows query output for roll up using star schema.

Table 11: Results of Roll up OLAP Operation

<table>
<thead>
<tr>
<th>Publisher Name</th>
<th>Publisher items in Top 300 comic books</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marvel</td>
<td>1060</td>
</tr>
<tr>
<td>Image</td>
<td>331</td>
</tr>
<tr>
<td>Idea and Design</td>
<td>153</td>
</tr>
<tr>
<td>Dark Horse</td>
<td>143</td>
</tr>
<tr>
<td>Archie</td>
<td>98</td>
</tr>
<tr>
<td>Devil's Due</td>
<td>70</td>
</tr>
<tr>
<td>Dynamic</td>
<td>64</td>
</tr>
<tr>
<td>Avatar</td>
<td>60</td>
</tr>
<tr>
<td>Gemstone</td>
<td>48</td>
</tr>
<tr>
<td>Bongo</td>
<td>32</td>
</tr>
<tr>
<td>Boom</td>
<td>31</td>
</tr>
<tr>
<td>Antarctic</td>
<td>28</td>
</tr>
<tr>
<td>Oni</td>
<td>27</td>
</tr>
<tr>
<td>Virgin</td>
<td>25</td>
</tr>
<tr>
<td>Slave Labor</td>
<td>22</td>
</tr>
<tr>
<td>Viz</td>
<td>20</td>
</tr>
<tr>
<td>Aspen</td>
<td>20</td>
</tr>
</tbody>
</table>
Drill down

Drill down operation is reverse of Roll up operation. By computing OLAP Drill down operation on Star Schema [Figure 11], we are calculating the contribution of each publishers in Top 300 comic books sold every year we also can drill down to one more hierarchy for the dimension time. On drilling down, the time dimension is descended from the level of year to the level of month. I have created drop down list to generate Drill down operation results on OLAP cube. Figure 17 shows web page for drilldown operation.

![A Case Study for Data Warehousing Courseware](image)

**Figure 17 : Web page to generate Drill down results**

We can generate drilldown up operation results by using below query.

Select bp.Publisher_Name ,t.month,t.year,
count(bp.Publisher_Name)
from Comic_Books_fact cbf, Book_publisher bp,
Time t,Distributor d, Book_categories bc
where cbf.Distributor_id = d.Distributor_id
and t.Time_id = cbf.Time_id
and cbf.Book_catgry_id = bc.Book_catgry_id
and t.year= 2006 and t.month = 'Jan'
and cbf.publisher_id = bp.Publisher_id
and bp.Publisher_Name like 'Marvel%' group by cbf.publisher_id;

In the above query, select operation will select respective year, month, publisher name
dimensions and count (Publisher_Name ) function will calculate total number of times
publisher name repeated in a sales data. With the help of where clause we can drill down to
individual publisher contribution in Top300 comic books sold in a respective month and year
.Group by function is used for grouping the result set by publisher name.
Figure 18 shows query output for drill down operation using star schema.
Figure 18 : Results of Drill down OLAP operations

Slice Operation

OLAP Slice operation performs selection on one dimension of a given cube, resulting in a sub cube. Hence, by applying OLAP slice operation we can just slice the comic books name dimensions data in a given year from the data cube. Slice operation can be performed using query below.

```sql
select bn.Title_name ,bn.Issuance,
bn.price,t.year
from Comic_Books_fact cbf ,
Book_Name bn ,Time t
where cbf.Book_id =bn.Book_id
and t.year =2006 and
```

Click see SQL query used to generate Drill down operation. SQL Script
\begin{verbatim}
t.month = 'jan' and
cbf.Time_id = t.time_id;
\end{verbatim}

From the above query, the select operation will select the respective time and Book Name dimensions. From clause explains that the time, Books Name dimensions should be selected from the star schema fact and dimension tables. The where clause performs referencing of foreign and primary keys between the above-mentioned tables.

The Table 12 displays the sample results of the above query.

\textbf{Table 12 : Results of OLAP Slice operations}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Comic Books Name & Issuance & Price & Year & Month \\
\hline
All Star Superman & 2 & $2.99 & 2006 & Jan \\
New Avengers & 15 & $2.50 & 2006 & Jan \\
Amazing Spider-Man & 523 & $2.50 & 2006 & Jan \\
Green Lantern & 7 & $2.99 & 2006 & Jan \\
Ultimate Extinction & 1 & $2.99 & 2006 & Jan \\
Uncanny X-Men & 463 & $2.50 & 2006 & Jan \\
Friendly Neighborhood Spider-Man & 4 & $2.99 & 2006 & Jan \\
Marvel Knights Spider-Man & 22 & $2.99 & 2006 & Jan \\
X-Men & 181 & $2.50 & 2006 & Jan \\
JLA & 124 & $2.50 & 2006 & Jan \\
Ultimate X-Men & 66 & $2.50 & 2006 & Jan \\
X-Men Deadly Genesis & 3 & $3.50 & 2006 & Jan \\
\hline
\end{tabular}
\end{table}
**Dice operation**

Dice selects two or more dimensions from a given cube and provides a new sub-cube. By computing OLAP Dice operation on Star Schema [Figure 15], we can calculate all the books published by a given publisher in given year. This involves three dimensions

- Books Name
- Time
- Publisher

Dice operation can be performed by using below query

```sql
Select bn.Title_name ,
bn.Issuance,bn.price,
t.year , bp.Publisher_Name
from Comic_Books_fact cbf ,
Book_Name bn , Time t,
Book_publisher bp
where cbf.Book_id =bn.Book_id
and t.year =2006 and t.month = 'jan'
and cbf.Time_id =t.time_id
and bp.Publisher_id =cbf.publisher_id
and bp.Publisher_Name like 'Dark Horse%';
```

From the above query, the select operation will select the respective time and Book Name and publisher dimensions. From clause explains that the Time, Books Name, Publisher dimensions should be selected from the star schema fact and dimension tables. The where
clause performs referencing of foreign and primary keys between the above-mentioned tables.

The Table 13 displays sample results of the above query.

**Table 13: Results of OLAP Dice operation**

<table>
<thead>
<tr>
<th>Books Title</th>
<th>Issuence</th>
<th>Price</th>
<th>Publisher</th>
<th>Month</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star Wars Knights of the</td>
<td>1</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>Star Wars Purge One Shc0</td>
<td>2</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>Star Wars Republic</td>
<td>81</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>Star Wars Republic</td>
<td>82</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>BPRD Black Flame</td>
<td>6</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>Revelations</td>
<td>6</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>Aeon Flux</td>
<td>4</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>Usagi Yojimbo</td>
<td>90</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
<tr>
<td>13th Son Worse Thing W 3</td>
<td>109</td>
<td>$2.99</td>
<td>Dark Hors</td>
<td>Jan</td>
<td>2006</td>
</tr>
</tbody>
</table>

### 3.8 Exercises

This section of the courseware helps the students to evaluate their understandings based on the examples demonstrated to them. Provided an example data set in exercises section for students use. Students can download and add more data to it. The users can create their own database and perform OLAP operations on the data set. Questions for OLAP operations are provided for practical learning experience for the users. Figure 19 shows snapshot of exercise page.
Exercises

This section contains exercises on OLAP operations.
Please download the sales data of Comic books in the Excel sheet Comic. BOOKS. SALES. DATA

Answer the following questions using OLAP Operations.

1. Find total sales of Marvel publisher over each month in 2004 using Rollup operation.

2. Total sales in 2005 is 76101989 Million perform drill down operation to find each individual publisher contribution in total sales.

3. By using Slice operation list all Publishers who have published Comic books in 2006.

Click here to check results for Exercise

Figure 19: Exercise Page
Chapter 4

CONCLUSION

Developing “A Case Study for Data warehousing Courseware” has also been a great learning experience. I was able to learn about OLAP and Data Warehousing methodologies in detail. I also learnt that the frameworks such as PHP, Java script, CSS, HTML, which are very powerful in developing custom websites and user interfaces.

The main objective of this project was to develop a web based interactive courseware that helps the users to understand the OLAP concepts with practical examples. As a conclusion to the project report. I feel that courseware would now illustrate the key concepts of OLAP cube using example demonstrations. It also allow users to understand the concepts of data mining, and implement it on raw data that is provided to the users for performing the exercise section. Overall, I feel I was able to accomplish the goals that I had set for myself at the beginning of the project.
APPENDIX

Code of Courseware Web Development

//index.php

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">

<head>

<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<title>A Case Study for Data Warehousing Courseware</title>

<meta name="keywords" content="Data Warehousing Courseware,CSC177,CSUS" />

<meta name="description" content="This is a courseware for CSC177" />

<style>

#navcontainer ul

{

margin: 0;

padding: 10px;

list-style-type: none;

text-align: center;

}

#navcontainer ul li { display: inline; }

#navcontainer ul li a

{

text-decoration: none;

}
padding: .2em 1.8em;
color: #fff;
background-color: #659EC7;
}
#navcontainer ul li a:hover,a:active
{
color: #fff;
background-color: #659EC7;
}
.currentLink {
color:red;
}
#navcontainer li a.current{
color: #fff;
background-color: #357EC7;
}
#header {
background-color:#996633;
color:white;
text-align:center;
padding:5px;
}
<body>

<div align="center">
    <img src="OLAPimage.png" align="center" width='1100px' height='150px'>
</div>

<div id="navcontainer" align= "center">
    <ul>
        <li><a class ="current" href="index.php">Home</a></li>
        <li><a  href="reportdemo1.php">A Multi-Dimensional Data Model</a></li>
    </ul>
</div>
<table width="1100px" border="0" align="center">
  <tr>
    <td colspan="2">
      <font style="font-family:'Calibri';
      font-size:22px;
      color:maroon;
      font-weight:bold;">Introduction</font>
      <p>Introduction</p>
    </td>
  </tr>
  <tr>
    <td colspan="2">
      &nbsp;&nbsp;&nbsp;&nbsp;
    </td>
  </tr>
</table>
A data warehouse (DW) is an approach for creating an enterprise-wide data store. It is an integral part of many information delivery systems because it contains consolidated data, obtained from several operational databases and other data sources, over long periods of time. With a large size data warehouse, query throughput and response times are very important. To facilitate these complex analyses data warehouses also provide us Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in a multidimensional space.

The below figure helps to understand the relationship between Data Warehouse and online analytical processing (OLAP) cube.

Features of Data warehousing:

- Features of Data warehousing:
A physical repository where relational data is specially organized to provide enterprise wide cleansed data in a standardized format.

The Data Warehouse is that database which is kept separate from the organization's operational database.

The Data warehouse supports On-Line Analytical Processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases.

There is no frequent updating done in data warehouse.

Data warehouse helps the executives to organize, understand and use their data to make strategic decision.

To facilitate complex analyses and visualization, the data in a warehouse is typically modeled multi dimensionally.

To facilitate complex analyses and visualization, the data in a warehouse is typically modeled multi dimensionally.
Difference between Data warehouse and other Operational Database Systems"</font></strong></a></p><br><font size="18"></font><p><br><br></p></td></tr></table><br><br>//reportdemo1.php<!DOCTYPE html PUBLIC '-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"> <html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta http-equiv="Content-Type" content="text/html; charset=utf-8" /> <title>A Case Study for Data Warehousing Courseware</title> <meta name="keywords" content="Data Warehousing Courseware,CSC177,CSUS" />
<meta name="description" content="This is a courseware for CSC177" />
<style>
#navcontainer ul
{
  margin: 0;
  padding: 10px;
list-style-type: none;
text-align: center;
}
#navcontainer ul li { display: inline; }
#navcontainer ul li a
{
text-decoration: none;
padding: .2em 1.8em;
color: #fff;
background-color: #659EC7;
}
#navcontainer ul li a:hover,a:active
{
color: #fff;
background-color: #659EC7;
}
.currentLink {

color:red;
}
#navcontainer li a.current{
color: #fff;
background-color: #357EC7;
<div align="center">

<img src="OLAPimage.png" align="center" width='1100px' height='150px'>

</div>

<div id="navcontainer"; align="center">
<ul>
<li><a href="index.php">Home</a></li>
<li><a class="current" href="reportdemo1.php">A Multi-Dimensional Data Model</a></li>
<li><a href="reportdemo2.php">Data Cube</a></li>
<li><a href="report3.php">Introduction to OLAP</a></li>
<li><a href="report4.php">OLAP operations</a></li>
<li><a href="example.php">Examples</a></li>
<li><a href="contactus.html">Quizzes</a></li>
<li><a href="contactus.html">Reference</a></li>
</ul>
</div>

<table width="1100px" border="0" align="center" >
<tr>
<td colspan="2" >

<font style="font-family:'Calibri';
font-size:22px;
color:maroon;" >

</font>

</td>
</tr>
</table>
Multi dimensional Data Model

Data warehouses and OLAP tools are based on a multidimensional data model. This model views data in the form of a data cube. The multidimensional data model is composed of some basic elements like logical cubes, measures, dimensions, hierarchies, levels, and attributes.
&nbsp;&nbsp;&nbsp;
A cube is a logical organization of multidimensional data. A cube is derived from a fact table. Dimensions categorize a cube's data and a cube contains measures that share the same dimensionality. Cubes are not usually exposed to end-users since they are more interested in the measure(s) contained within the cubes.
Measures are numeric representations of a set of facts that have occurred. Examples of measures include dollars of sales, number of credit hours, store profit percentage, dollars of operating expenses etc.

Dimensions are the perspectives or entities with respect to which an organization wants to keep record. For example if user wants to keep track of comic books that are published in a year 2000 with respect to dimensions that user needs to keep track are books_categories, year, publisher etc. Each dimension may have a table associated with it called a dimension table.
A hierarchy is a way to organize data at different levels of aggregation. In viewing data, analysts use dimension hierarchies to recognize trends at one level, drill down to lower levels to identify reasons for these trends, and roll up to higher levels to see what affect these trends have on a larger sector of the business.
color:maroon;
font-weight:bold;"}>Level </font><br />
</td>
</tr>
<tr>
<td>
It is a column within a dimension table that could be used for aggregating data. For example, product dimension could have levels of product type (beverage), product category (alcoholic beverage), product class (beer), product name (miller lite, budlite, corona, etc)</p>
</td>
</tr>
<tr>
<td>
<p><img src="http://athena.ecs.csus.edu/~biligers/msproject/Level.png" alt="" align="left" width="750" height="250" /></p>
</td>
</tr>
<tr>
<td>
</td>
</tr>
<tr>
<td>
</td>
</tr>
<tr>
<td>
</td>
</tr>
<tr>
<td>
</td>
</tr>
</tr>
<tr>
<td>
<font style="font-family:'Calibri';"}


The relational implementation of the multidimensional data model is typically a star schema, or a snowflake schema. This model helps in organizing the data into dimension tables, fact tables and materialized views.
The snowflake schema consists of one fact table that is connected to many
dimension tables, which can be connected to other dimension tables through a many-to-
one relationship. Tables in a snowflake schema are usually normalized to the third normal form. Each dimension table represents exactly one level in a hierarchy.

Star Schema

Star schema is the simplest form of a dimensional model, in which data is
organized into facts and dimensions. A fact is an event that is counted or measured. A
dimension contains reference information about the fact such as date, product, or customer. A star schema is diagramed by surrounding each fact with its associated dimensions. The resulting diagram resembles a star schema.

```html
//reportdemo2.php

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
</head>
<html xmlns="http://www.w3.org/1999/xhtml">
```
#navcontainer ul li a:hover, a:active
{

color: #fff;
background-color: #659EC7;
}

.currentLink {

color:red;
}

#navcontainer li a.current{

color: #fff;
background-color: #357EC7;
}

#header {

background-color:#996633;
color:white;
text-align:center;
padding:5px;
}

p.ex {

margin-top: 10px;
margin-bottom: 10px;
margin-right: 1500px;
<div align="center">
<img src="OLAPimage.png" align="center" width='1100px' height='150px'>
</div>

<div id="navcontainer" align= "center">
<ul>
<li><a  href="index.php">Home</a></li>
<li><a  href="reportdemo1.php">A Multi-Dimensional Data Model</a></li>
<li><a class ="current" href="reportdemo2.php">Data Cube</a></li>
<li><a href="report3.php">Introduction to OLAP</a></li>
<li><a href="report4.php">OLAP operations</a></li>
<li><a href="example.php">Examples</a></li>
</ul>
</div>
A data cube is a type of multidimensional matrix that lets users explore and analyze a collection of data from many different perspectives. The cube is used to represent data along some measure of interest. Although called a <b>cube</b>, it can be 2-dimensional, 3-dimensional, or higher-dimensional. Each dimension represents some attribute in the database and the cells in the data cube represent the measure of interest. For example, they could contain a count for the number of times that attribute combination occurs in the database, or the minimum, maximum, sum or average value of some attribute. Queries are performed on the cube to retrieve decision support information.
<table>
<thead>
<tr>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider comic books distributor company sales fact table in North California region.</td>
</tr>
<tr>
<td>Example for two Dimensional representation of Data Cube</td>
</tr>
</tbody>
</table>
Multi-dimensional databases are a compact and easy-to-understand way of visualizing and manipulating data elements that have many inter-relationships. The cube can be expanded to include another dimension, for example, sales fact compared with SouthCalifornia region.

Example for three Dimensional representation of Data Cube
<tr>
<td>
<img align="left" src="http://athena.ecs.csus.edu/~biligers/msproject/cube2.jpg" alt="" >
</td></tr >
</table>

//reportdemo3.php

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>A Case Study for Data Warehousing Courseware</title>
<meta name="keywords" content="Data Warehousing Courseware,CSC177,CSUS" />
<meta name="description" content="This is a courseware for CSC177" />
<style>
#navcontainer ul
{
 margin: 0;
 padding: 10px;
 list-style-type: none;
}
text-align: center;

}

#navcontainer ul li { display: inline; }

#navcontainer ul li a
{

text-decoration: none;
padding: .2em 1.8em;
color: #fff;
background-color: #659EC7;
}

#navcontainer ul li a:hover,a:active
{
color: #fff;
background-color: #659EC7;
}

.currentLink {

color:red;
}

#navcontainer li a.current{

color: #fff;
background-color: #357EC7;
}
#header {
background-color:#996633;
color:white;
text-align:center;
padding:5px;
}
p.ex {
margin-top: 10px;
margin-bottom: 10px;
margin-right: 1500px;
margin-left: 2500px;
}
</style>

<STYLE TYPE="text/css">
<!--
TD{font-family: calibri; font-size: 14pt; text-align: justify;
   text-justify: inter-word;}
--->
</STYLE>

</head>

<body>

<div align="center">
<font style="font-family:'calibri';"
font-size:22px;
color:maroon;
font-weight:bold;">
On-line Analytical Processing (OLAP)
</font>
OLAP systems are part of decision support systems and will assist analysts and managers, those who are responsible for the smooth running of an organization by giving them quick access to data. OLAP tools provide users with a fast response even if the query request is made on a large volume of data. Basically, OLAP tools provide the ability to transform huge volumes of data that exist in the organization into useful information to support decision-making process.
OLAP tools are categorized according to the architecture used to store and process multi-dimensional data. There are four main categories of OLAP tools.

The four main categories of OLAP tools are listed below.

Relational OLAP (ROLAP)
ROLAP stores all data, including aggregations, in the source relational database. This type of storage is good for enterprises that need larger data warehousing. It uses an SQL reporting tool to query data directly from the data warehouse.

<p></p></td>
</tr>
<tr><td>
<font style="font-family:'calibri'; font-size:18px; color:maroon; font-weight:bold;">
Multidimensional OLAP (MOLAP)<br>
</font>
</td></tr>
<tr><td>
MOLAP is the more traditional OLAP type. In MOLAP, both the source data and the aggregation calculations are stored in a multidimensional format. This type is the fastest option for data retrieval, but it also requires the most storage space</p>
Hybrid OLAP (HOLAP)

HOLAP technique is combination of ROLAP and MOLAP in a single system. It has both the higher scalability of ROLAP and faster computation of MOLAP. HOLAP
server allows to store the large data volumes of detail data and the aggregations are stored separated in MOLAP store.

Desktop OLAP (DOLAP)

DOLAP is based on the idea that a user can download a data cube and work with it locally. This type of application is easy to deploy and has lower costs, but it is very limited in its performance. With Active Pivot, Excel sheets can be saved to the user's desktop, for local access to OLAP data.
<a href="HOLAPandDOLAP.php"><strong>Click</strong> to know about more about HOLAP and DOLAP categories</a></p>
<br>
</td>
</tr>
</table>

//report4.php

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>A Case Study for Data Warehousing Courseware</title>
<meta name="keywords" content="Data Warehousing Courseware, CSC177, CSUS" />
<meta name="description" content="This is a courseware for CSC177" />
<style>
#navcontainer ul

margin: 0;
padding: -5px;
list-style-type: none;
text-align: center;
}

#navcontainer ul li { display: inline; }

#navcontainer ul li a
{
text-decoration: none;
padding: .2em 1.8em;
color: #fff;
background-color: #659EC7;
}

#navcontainer ul li a:hover,a:active
{
color: #fff;
background-color: #659EC7;
}

#navcontainer li a.current
{ .currentLink
{
color:red;
}
#navcontainer li a.current{
<body>
<div align="center">
<img src="OLAPimage.png" align="center" width='1100px' height='150px'>
</div>
<div id="navcontainer" align="center">
<ul>
<li><a href="index.php">Home</a></li>
<li><a href="reportdemo1.php">A Multi-Dimensional Data Model</a></li>
<li><a href="reportdemo2.php">Data Cube</a></li>
<li><a href="report3.php">Introduction to OLAP</a></li>
<li><a href="report4.php" class="current">OLAP operations</a></li>
<li><a href="example.php">Examples</a></li>
<li><a href="contactus.html">Quizzes</a></li>
<li><a href="contactus.html">Reference</a></li>
</ul>
</div>
<table width="1100px" border="0" align="center">
<tr>
</tr>
<tr>
</tr>
<td>
<font style="font-family:'calibri';"
OLAP provides user with the flexibility to view data from different perspectives hence we will discuss the OLAP operations in multidimensional data.

- **ROLL UP**

```html
<ul>
    <li>ROLL UP</li>
</ul>
```
<li>A roll-up involves summarizing the data along a dimension.</li>

The roll-up operation is performed by climbing up a concept hierarchy for the dimension location.

When roll-up operation is performed then one or more dimensions from the data cube are removed.

</ul>

<font style="font-family:'calibri';
font-size:18px;
color:maroon;
font-weight:bold;">Drill-down</font>

</td></tr>
<tr><td width="100%" style="vertical-align:top">
<ul>
<li>Drill down is the reverse of roll-up. Navigates from more detailed data it can achieved by any of the following way.</li>
<li>Stepping down a concept hierarchy for a dimension</li>
<li>Introduces additional dimensions

</ul>
Below figure shows the demo of Roll up and Drill down operations.

![Rollup.jpg](http://athena.ecs.csus.edu/~biligers/msproject/Rollup.jpg)

### Slice

- Performs a selection on one dimension of the given cube, resulting in a sub-cube.
<li>Reduces the dimensionality of the cubes.</li>
<li>Sets one or more dimensions to specific values and keeps a subset of dimensions for selected values</li>
</ul>
</td>
</tr>
<tr>
<td>
<font style="font-family:'calibri';
    font-size:18px;
    color:maroon;
    font-weight:bold;">Dice</font>
</td>
</tr>
<tr>
<td width="100%" style="vertical-align:top">
<ul>
<li>Define a sub-cube by performing a selection of one or more dimensions.</li>
<li>Refers to range select condition on one dimension, or to select condition on more than one dimension.</li>
<li>Reduces the number of member values of one or more dimensions</li>
</ul>
</td>
</tr>
Below figure shows the demo of slice and Dice operations

![Slice and Dice Operations](http://athena.ecs.csus.edu/~biligers/msproject/sliceanddice.jpg)

<table>
<thead>
<tr>
<th>Pivot (or rotate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotates the data axis to view the data from different perspectives.</td>
</tr>
<tr>
<td>Groups data with different dimensions</td>
</tr>
</tbody>
</table>

Below figure shows the demo of OLAP Pivot operations
Some more OLAP operations are listed below:

Drill-across

An additional drilling operation.
<li>Executes queries involving more than one fact table</li>
</ul>
</td>
</tr>
<tr><td><font style="font-family:'calibri';
font-size:18px;
color:maroon;
font-weight:bold;">Drill-through</font>
</td>
</tr>
<tr><td style="vertical-align:top">
<ul>
<li>An additional drilling operation.</li>
<li>Uses relational SQL facilities to drill through the bottom level of a data cube down to its back end relational tables</li>
</ul>
</td>
</tr>

//example.php
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>A Case Study for Data Warehousing Courseware</title>
<meta name="keywords" content="Data Warehousing Courseware, CSC177, CSUS" />
<meta name="description" content="This is a courseware for CSC177" />
<style>
#navcontainer ul
{
margin: 0;
padding: 10px;
list-style-type: none;
text-align: center;
}
#navcontainer ul li { display: inline; }
#navcontainer ul li a
{
text-decoration: none;
padding: .2em 1.8em;
color: #fff;
background-color: #659EC7;
}
#navcontainer ul li a:hover,a:active
{
color: #fff;
background-color: #659EC7;
}
.currentLink {
color:red;
}
#navcontainer li a.current{
color: #fff;
background-color: #357EC7;
}
#header {
background-color:#996633;
color:white;
text-align:center;
padding:5px;
}
p.ex {
<div align="center">
<img src="OLAPimage.png" align="center" width='1100px' height='150px'>
</div>

<div id="navcontainer" align= "center">
<ul>
<li><a  href="index.php">Home</a></li>
<li><a  href="reportdemo1.php">A Multi-Dimensional Data Model</a></li>
<li><a href="reportdemo2.php">Data Cube</a></li>
</ul>
</div>
In our courseware we are using Comic books sales data of a Diamond Distributors to illustrate OLAP operations. Here we are considering data of Top 300 comic books sold in every year.

Data mart used to organize data is listed below.
font-size: 18px;
color: maroon;
font-weight: bold;”> Dimension Tables</font></td></tr></table>

</td></tr>

</ul>

<li><a href="BooksNameTable.png"><b>BooksName</b></a></li>
<li><a href="TimeTable.png"><b>Time</b></a></li>
<li><a href="BookCategories.png"><b>BooksCategories</b></a></li>
<li><a href="DistributorsTable.png"><b>Distributors</b></a></li>
<li><a href="Publisher.png"><b>Publishers</b></a></li>

</ul>

</td></tr>

</table>

</td></tr>

</ul>

<li><a href="Facttable.png"><b>ComicBooksFactTable</b></a></li>

</ul>

</td></tr>

</table>

</td></tr>

</ul>

<li><a href="Facttable.png"><b>ComicBooksFactTable</b></a></li>

</ul>

</td>
The below Star schema shows the way Data has organized in the data Model:

Roll up operation
The Roll Up analytical operation is performed by navigating up a dimensional hierarchy to a more summarized level. Here by using OLAP Roll up operation we can find contributions of each publisher in Top 300 comic books sold in a year by Diamond distributor.

Rollup Operation example

Drill down operation
Drilling down is an analytical technique where a user navigates among levels of data ranging from most summarized to the most detailed one.

Here, by using OLAP drill down operation, we can drill down to each publisher contribution in Top 300 comic books of Diamond distributor.

<a href="Drilldownex.php">Drill down Operation example</a>
OLAP slice operation selects one particular dimension from a given cube and provides a new sub-cube. In our example we can obtain total sales information of a given books vs time.

Slice Operation example

Dice operation

Dice selects two or more dimensions from a given cube and provides a new sub-cube.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;tr&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;td vertical-align:Top; font-family:'Calibri';text-align: left width=&quot;10%&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;a href=&quot;diceex.php&quot;&gt;&lt;strong&gt;Dice Operation example&lt;/strong&gt;&lt;/a&gt;&lt;/p&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;font size=&quot;18&quot;&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/font&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/td&gt;</td>
<td></td>
</tr>
<tr>
<td>&lt;/tr&gt;</td>
<td></td>
</tr>
<tr>
<td>/table&gt;</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


[6] ICV2'S TOP 300 COMICS & TOP 300 GN'S INDEX,

http://web.stanford.edu/dept/itss/docs/oracle/10g/olap.101/b103/multimodel.html.

[8] MOLAP, ROLAP, And OLAP,