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

ACKNOWLEDGEMENTS

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

LIST OF TABLES

Tables Page

LIST OF FIGURES

Figures Page

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 ares. 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

Features	Traditional Data Base (OLTP)	Data Warehouse (OLAP)
Characteristic	It is based on Operational Processing.	It is based on Informational Processing.
Data	It mainly stores the current data that always guaranteed to be up-to-date.	It usually stores the Historical data whose accuracy is maintained over time.
Read/write	The most frequent type of access type is read/write.	It mostly use the read access for the stored data.
User Function	The common users are clerk, DBA, database professional.	The common users are knowledge worker (e.g., manager, executive, and analyst).
Data base design	The designing of database is ER based and application-oriented.	The designing is typically done using star or snowflake or hybrid schema and its subject-oriented
Summarization	The data is primitive and highly detailed.	The data is summarized and in consolidated form.
View	The view of the data is flat relational.	The view of the data is multidimensional
Function	It is used for day-to-day operations	It is used for trend analysis.

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]

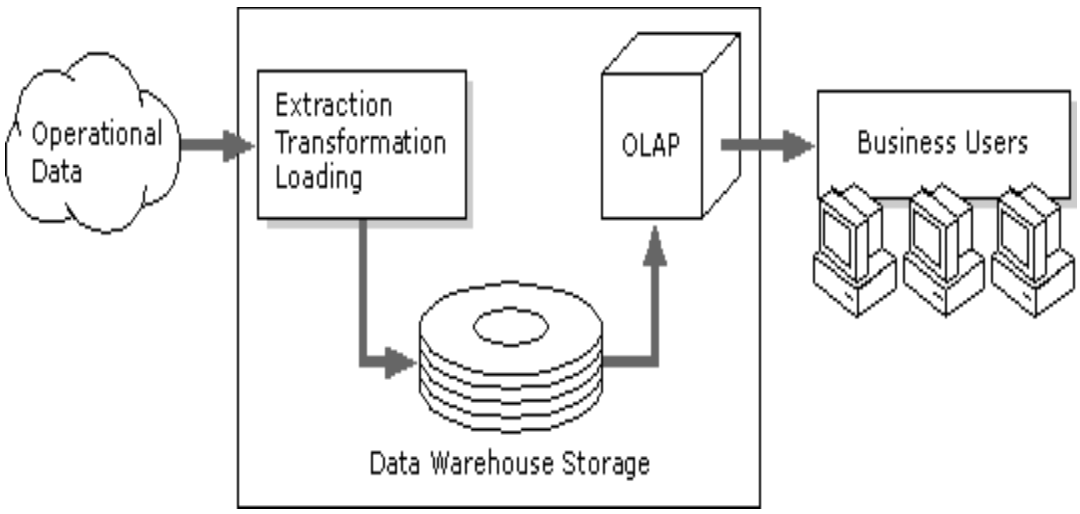


Figure 1 : Data warehouse and OLAP Cube[2]

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.

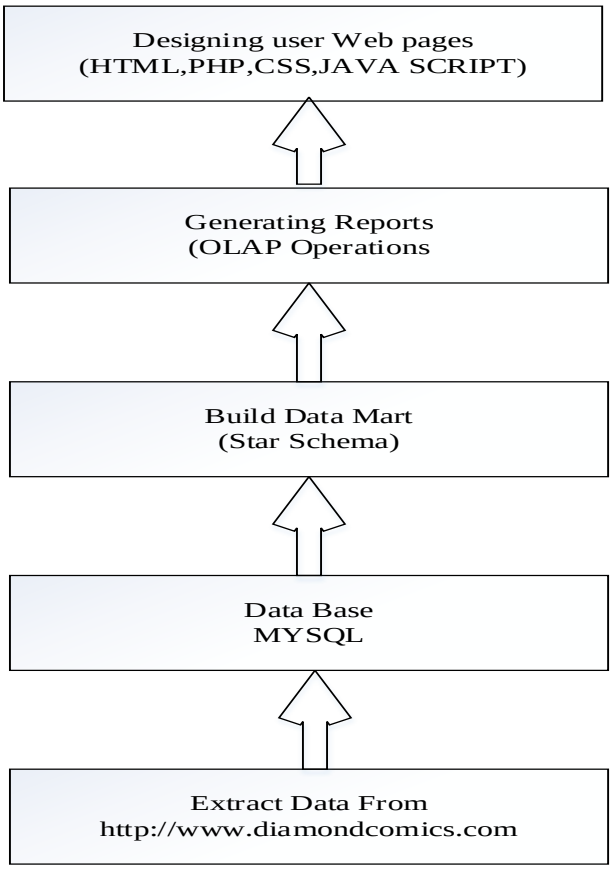


Figure 2: Flow Chart for Development of the Courseware

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.

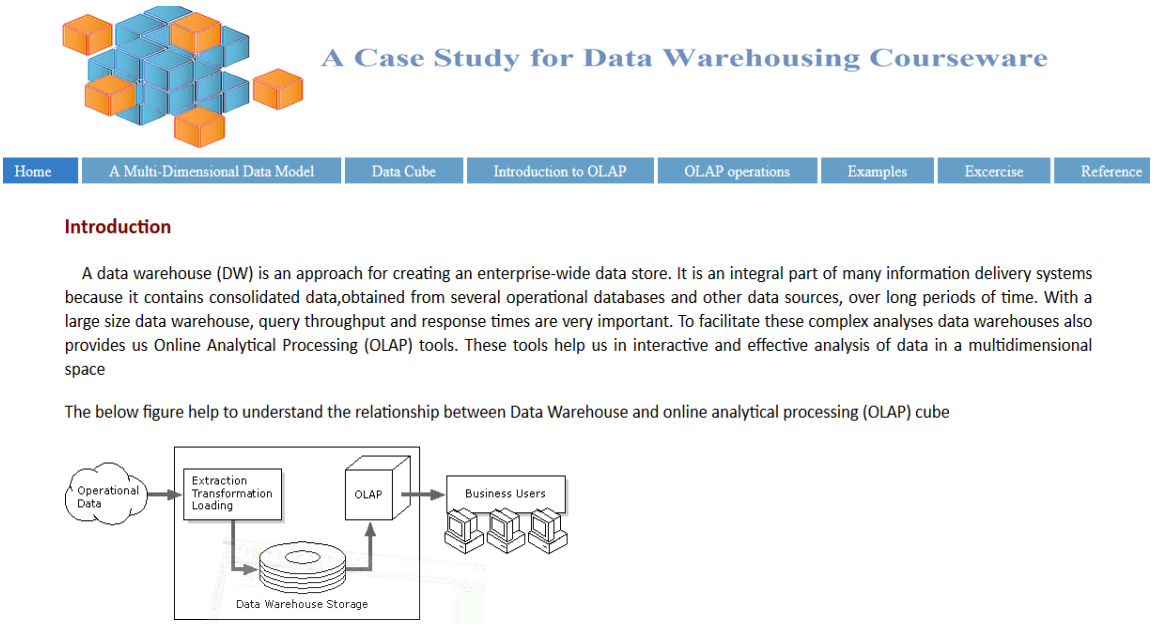


Figure 3 : A Case Study for Data Warehousing Courseware

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.

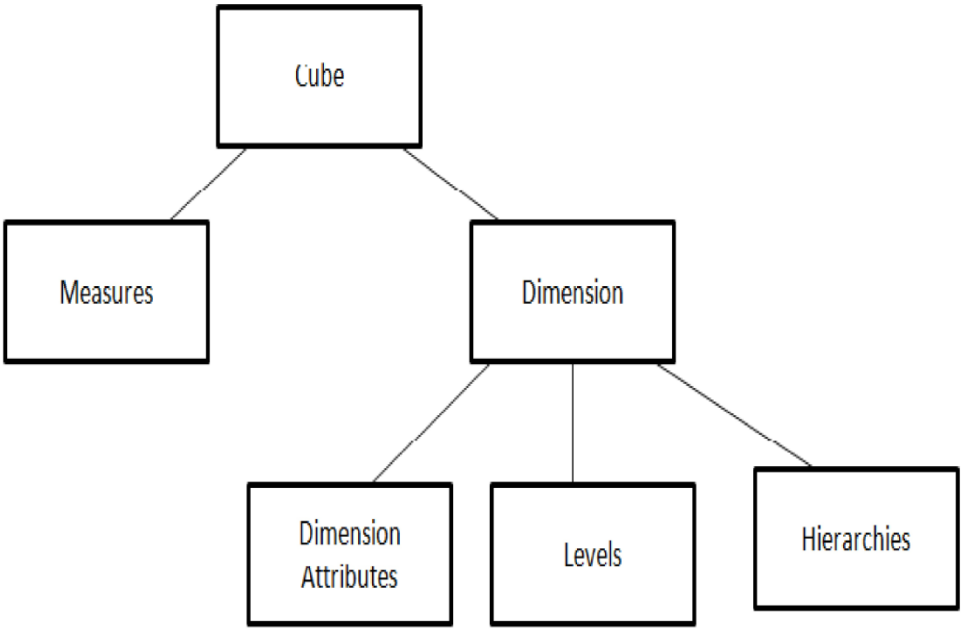


Figure 4: Shows the relationships among the basic elements [7]

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,

publisher etc. 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, 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.

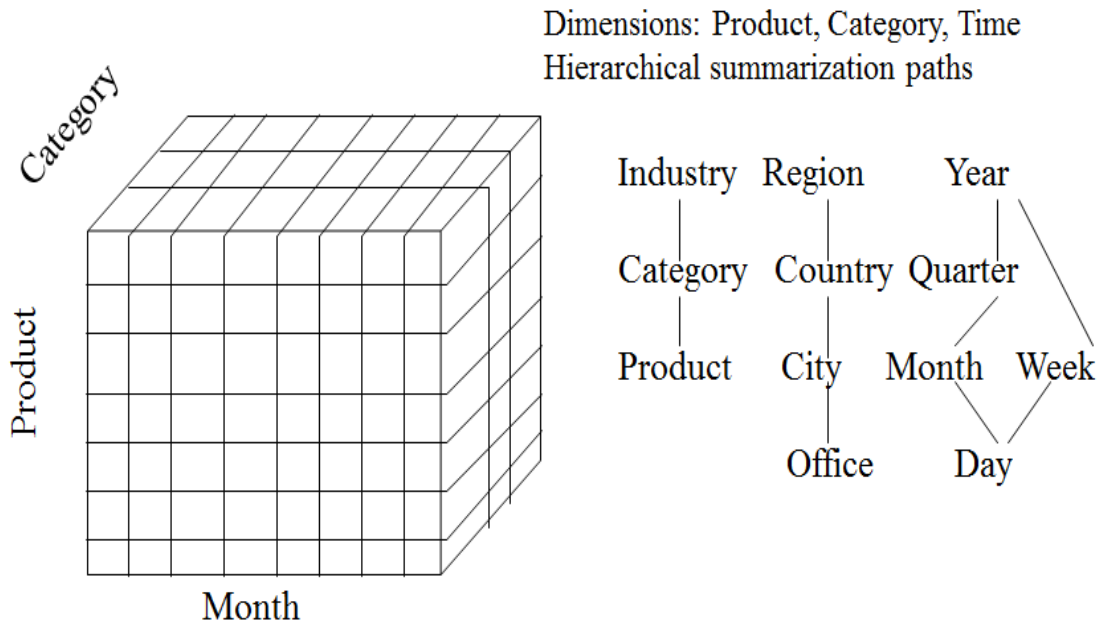


Figure 5: Levels and Hierarchy relation in a cube

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.

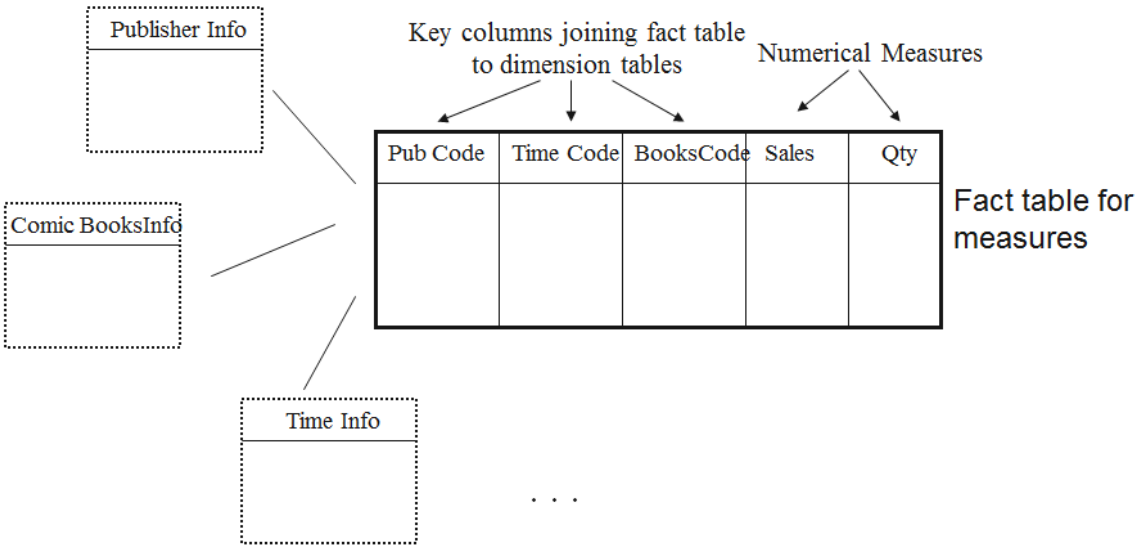


Figure 6: Sample representation of Star Schema.

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

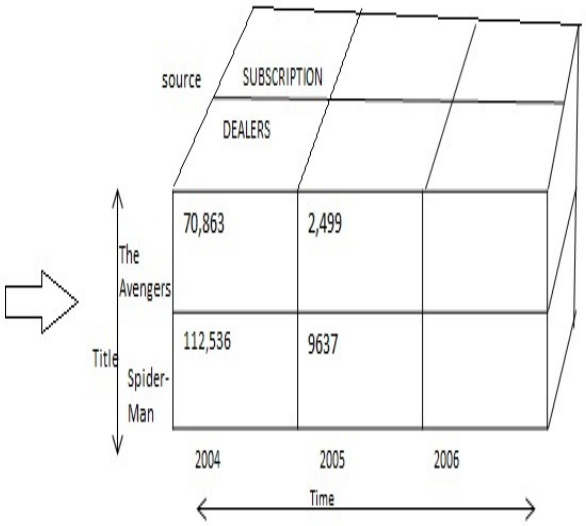
Title	year	Est sales
The Avengers	2004	73,362
The Avengers	2005	No Data
The Avengers	2006	150,000
The Avengers	2007	159,141
Amazing Spider-Man	2004	123,540
Amazing Spider-Man	2005	112,564
Amazing Spider-Man	2006	No Data
Amazing Spider-Man	2007	Pending

Time \ Title	The Avengers	Amazing Spider-Man
2004	73,362	123,540
2005	No Data	112,564
2006	150,000	No Data
2007	159,141	Pending

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.

Table 3: Three-dimensional representation of data model

Sources	Title	Time	Est sales
DEALERS	Amazing Spider-Man	2004	112,536
SUBSCRIPTION	Amazing Spider-Man	2004	102,377
DEALERS	Amazing Spider-Man	2005	9637
SUBSCRIPTION	Amazing Spider-Man	2005	10,187
DEALERS	The Avengers	2004	70,863
SUBSCRIPTION	The Avengers	2004	2,499
DEALERS	The Avengers	2005	45,150
SUBSCRIPTION	The Avengers	2005	5,612



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.

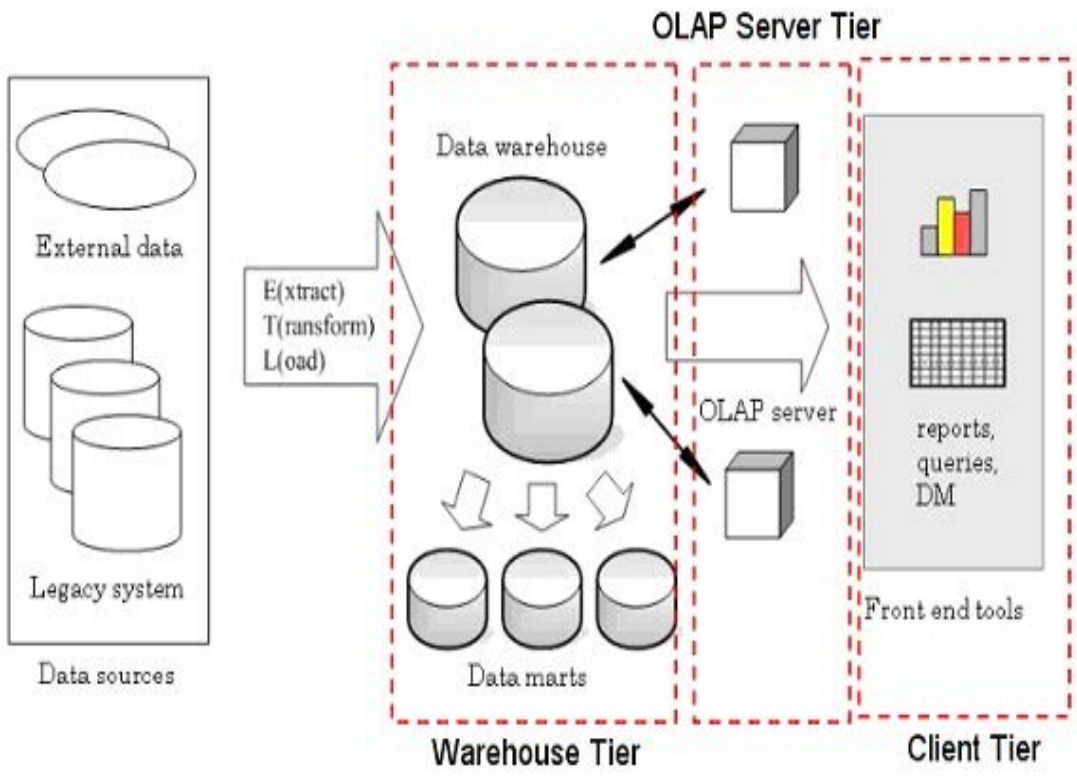


Figure 7: OLAP Architecture [4]

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 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 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.

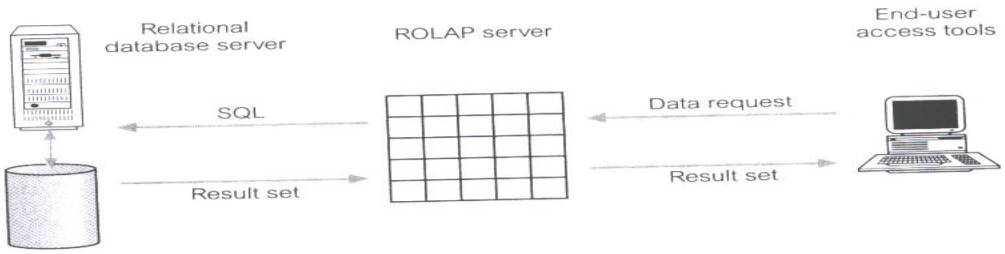


Figure 8: ROLAP Server Architecture

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 deformalized 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.

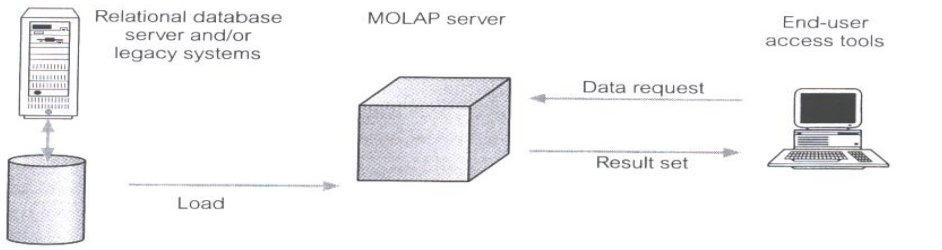


Figure 9: MOLAP Server Architecture

Advantages of MOALP

- 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.

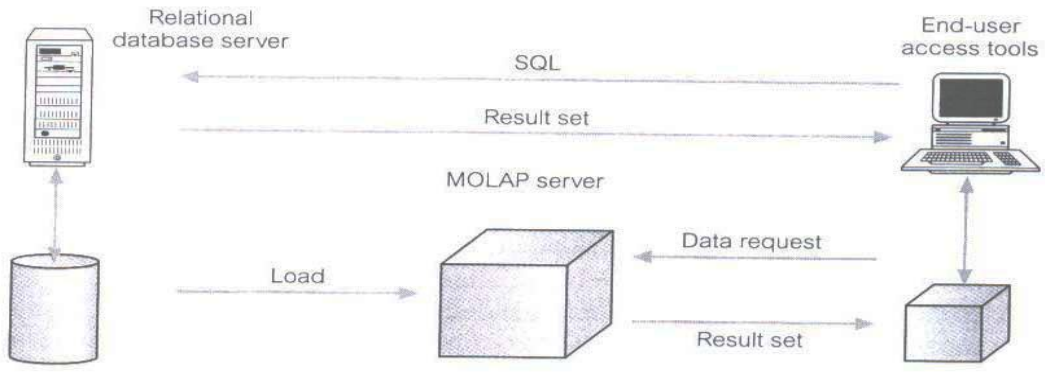


Figure 10: Hybrid OLAP Architecture

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.

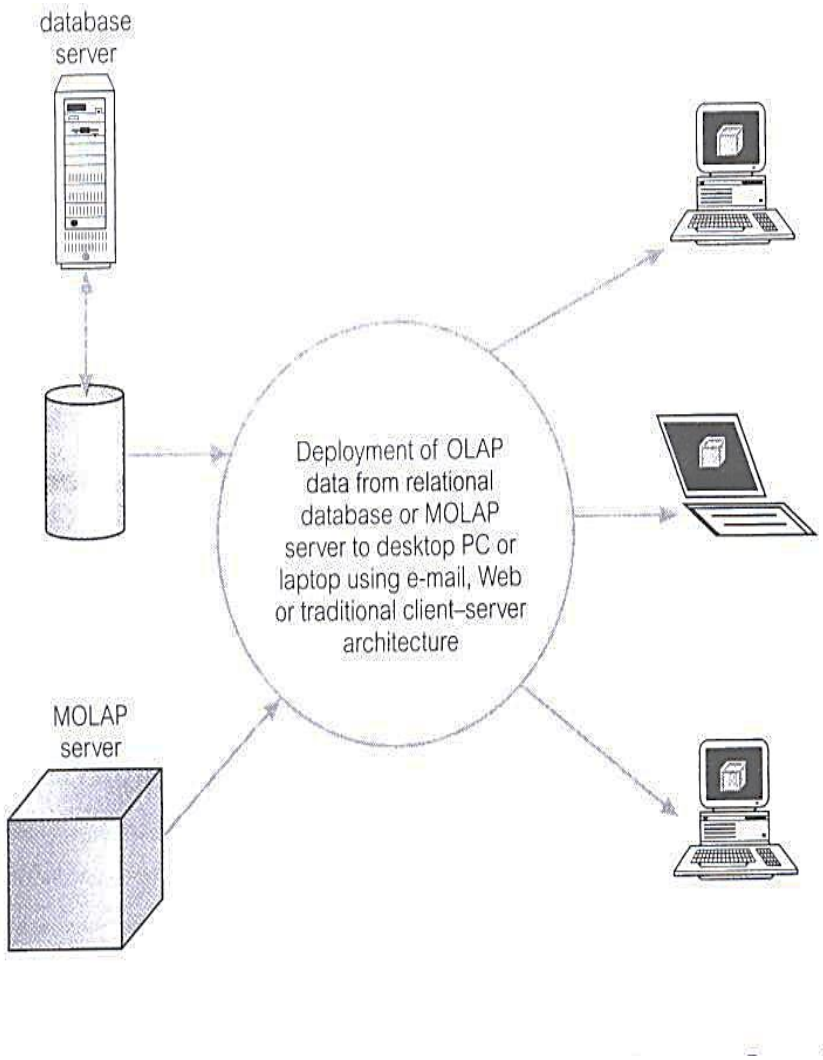


Figure 11: DOLAP Architecture [9]

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

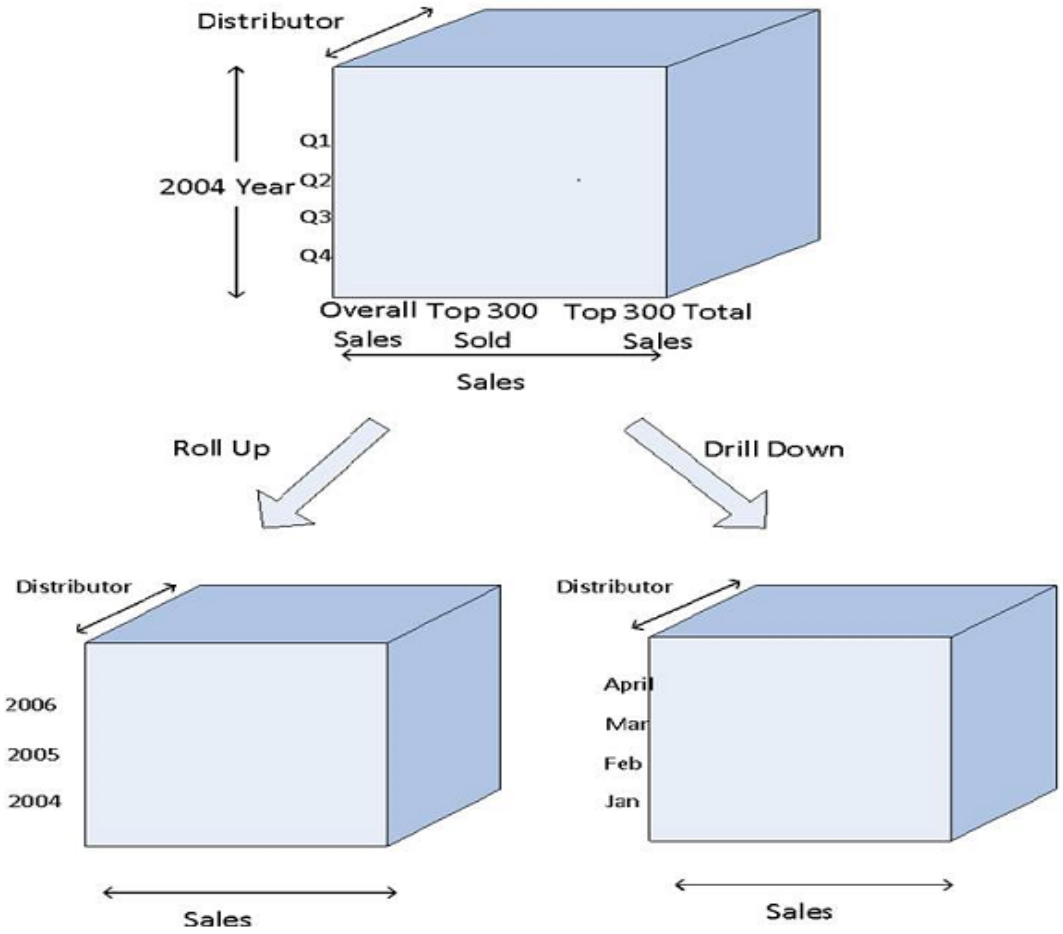


Figure 12: Rollup and Dice operations

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

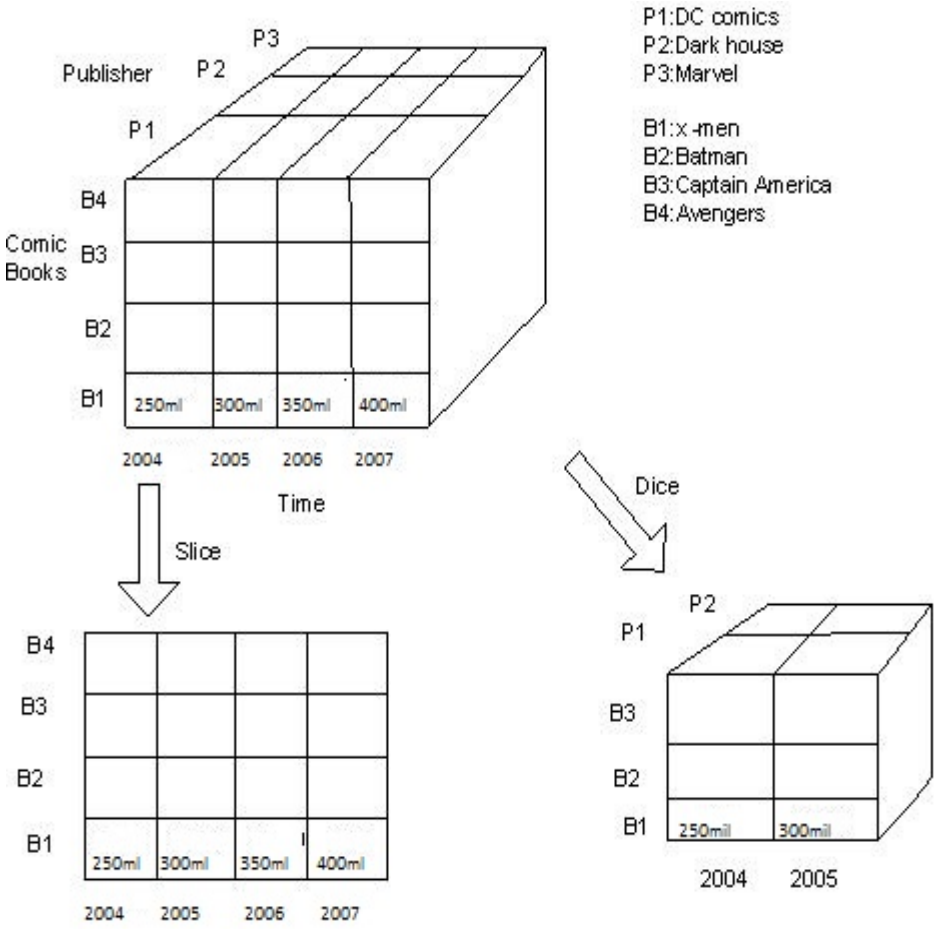


Figure 13 : Slice and Dice Operations

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.

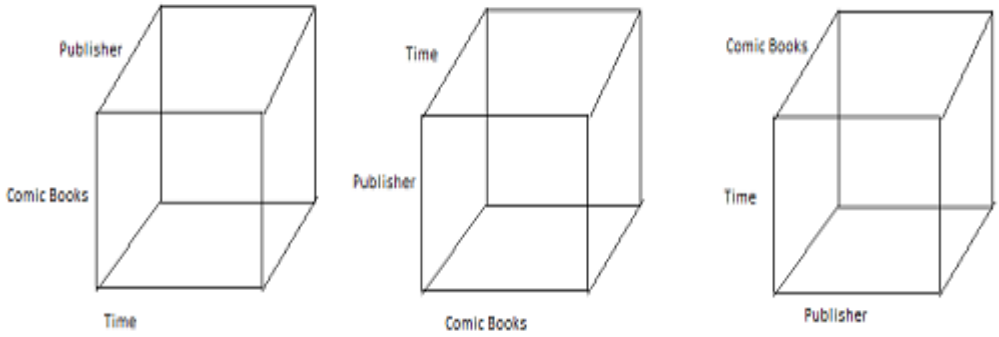


Figure 14 : OLAP Pivot Operations

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

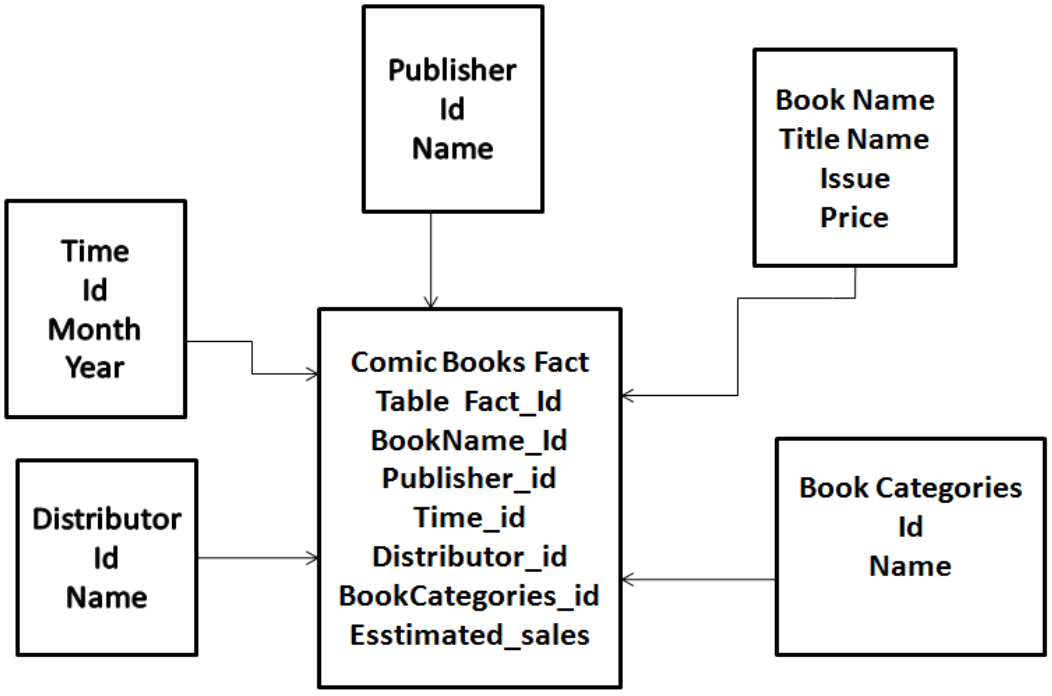


Figure 15 : Star Schema

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

Book_ID	Book_Name	Issue	Price	Publisher	Sales	Year	Month
1	Uncanny X-Men	378	1.99	Marvel	113703	2006	Jan
2	X-Men	98	1.99	Marvel	109676	2006	Jan
3	Wolverine	148	1.99	Marvel	87852	2006	Jan
4	JLA	39	1.99	DC	78273	2006	Jan
5	Avengers	26	1.99	Marvel	76535	2006	Jan
6	Spawn	94	1.95	Image	73517	2006	Jan
7	Soul Saga	1	2.5	Image	66273	2006	Jan
8	Earth X	11 (Res)	2.99	Marvel	64789	2006	Jan
9	Tomb Raider	3	2.5	Image	63902	2006	Jan
10	X-Men Unlimited	26	2.99	Marvel	62097	2006	Jan
11	Fantastic Four	27	1.99	Marvel	58575	2006	Jan
12	Cable	77	1.99	Marvel	57214	2006	Jan
13	Amazing Spider-Man	15	1.99	Marvel	56413	2006	Jan
14	Thor	21	1.99	Marvel	55965	2006	Jan
15	Peter Parker Spider-Man	15	1.99	Marvel	54318	2006	Jan
16	Iron Man	26	1.99	Marvel	52590	2006	Jan
17	Batman Gotham Knights	1	2.5	DC	51264	2006	Jan
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•

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

Book_id	Title Name	Issuence	Price
1111	Uncanny X-Men	378	1.99
1112	X-Men	98	1.99
1113	Wolverine	148	1.99
1114	JLA	39	1.99
1115	Avengers	26	1.99
1116	Spawn	94	1.95
1117	Soul Saga	1	2.5
1118	Earth X	11 (Res)	2.99
1119	Tomb Raider	3	2.5
1120	X-Men Unlimited	26	2.99
1121	Fantastic Four	27	1.99
1122	Cable	77	1.99
1123	Amazing Spider-M	15	1.99
1124	Thor	21	1.99
1125	Peter Parker Spid	15	1.99
•	•	•	•
•	•	•	•
•	•	•	•

- Book Categories

Sample data in Book_categories table is show in Table 6.

Table 6 : Data in Book Categories

Book_category_id	Book_category_Name
101	Comic Books
102	Graphics Novel

- Distributor

Sample data in Distributor table shown in Table 7.

Table 7 : Data in Distributor Table

Distributor_id	Distributor_name
100001	Diamond distributor

- Book Publisher

Sample data in Book Publisher table shown in Table 8.

Table 8 : Sample Data in Book Publisher

Publisher Name	Publisher_id
12 Gauge	2222
3 Finger Prints	2223
360ep	2224
3BP	2225
88 MPH	2226
A Silent	2227
AAA Milwaukee	2228
.	.
.	.

- Time

  Sample data in Time table shown in Table 9.

Table 9 : Sample table in Time Table

ID	Year	Month
105	2008	Sept
106	2008	Oct
107	2008	Nov
108	2008	Dec
109	2009	Jan
110	2009	Feb
111	2009	Mar
112	2009	April
113	2009	May
114	2009	June
115	2009	July
.	.	.
.	.	.

- 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

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

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.

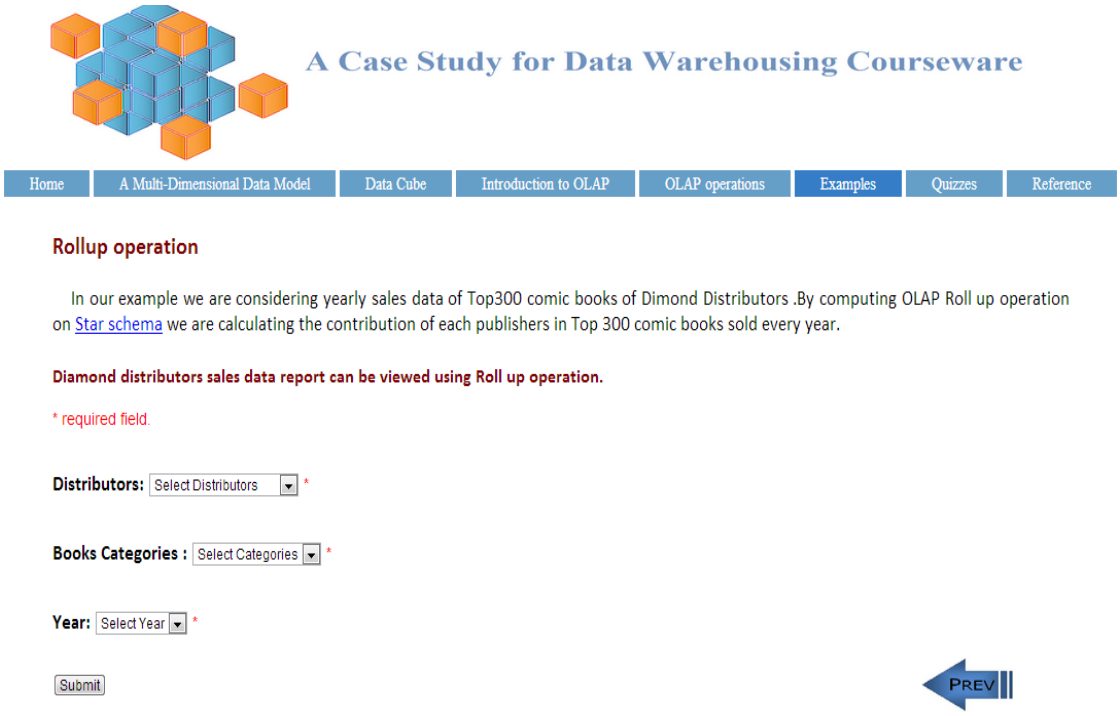


Figure 16 : Web page to generate Roll up operation

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

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

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.

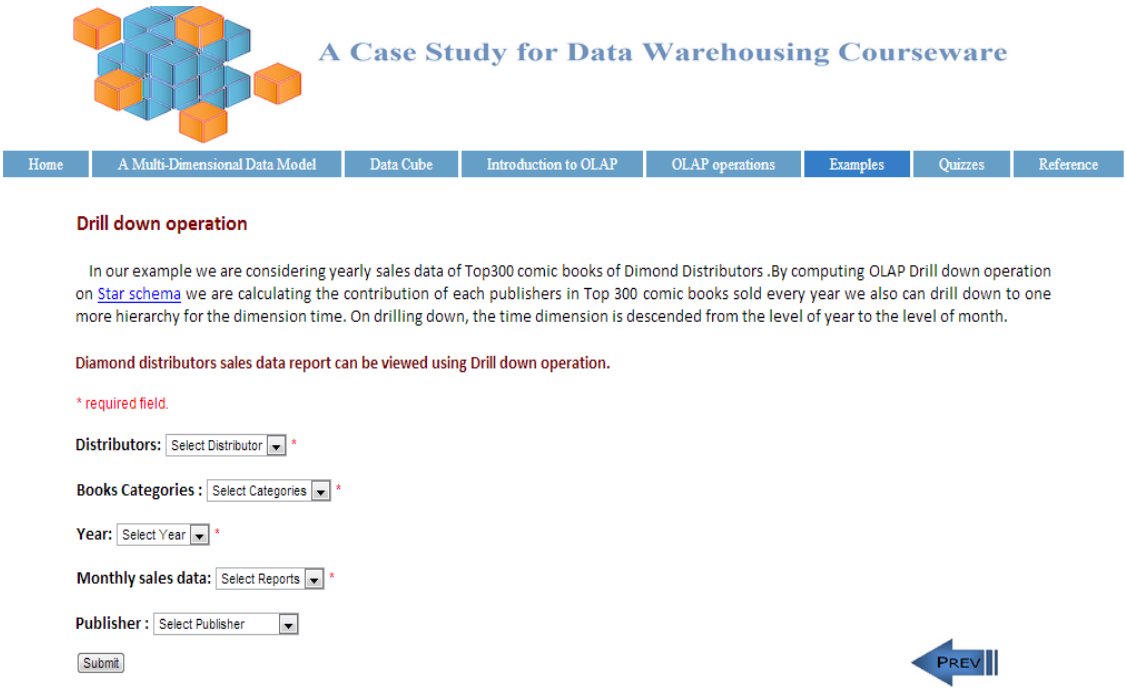


Figure 17 : Web page to generate Drill down results

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.

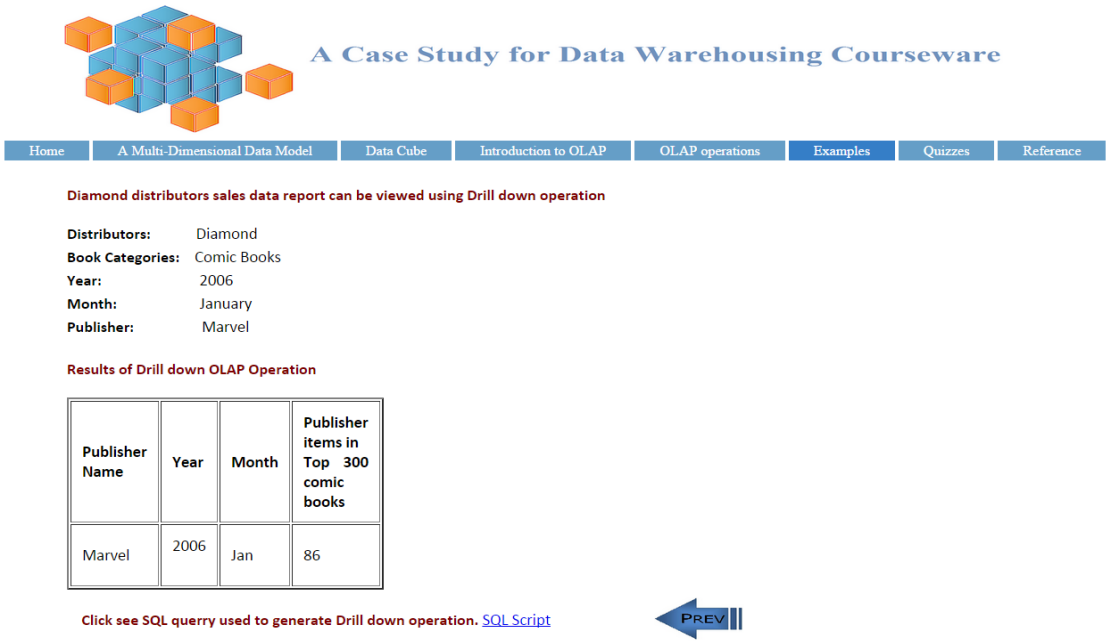


Figure 18 : Results of Drill down OLAP operations

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

```
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
```

t.month = 'jan' and

cbf.Time_id =t.time_id;

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.

Table 12 : Results of OLAP Slice operations

Comic Books Name	Issuence	Price	Year	Month
All Star Superman	2	$2.99	2006	Jan
New Avengers	15	$2.50	2006	Jan
Amazing Spider-Man	528	$2.50	2006	Jan
Green Lantern	7	$2.99	2006	Jan
Ultimate Extinction	1	$2.99	2006	Jan
Uncanny X-Men	468	$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
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

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

```
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

Books Title	Issuence	Price	Publisher	Month	Year
Star Wars Knights of the	1	$2.99	Dark Hors	Jan	2006
Star Wars Purge One Shc	0	$2.99	Dark Hors	Jan	2006
Star Wars Republic	81	$2.99	Dark Hors	Jan	2006
Star Wars Republic	82	$2.99	Dark Hors	Jan	2006
BPRD Black Flame	6	$2.99	Dark Hors	Jan	2006
Revelations	6	$2.99	Dark Hors	Jan	2006
Aeon Flux	4	$2.99	Dark Hors	Jan	2006
Usagi Yojimbo	90	$2.99	Dark Hors	Jan	2006
13th Son Worse Thing W	3	$2.99	Dark Hors	Jan	2006
Blade of the Immortal	109	$2.99	Dark Hors	Jan	2006

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.

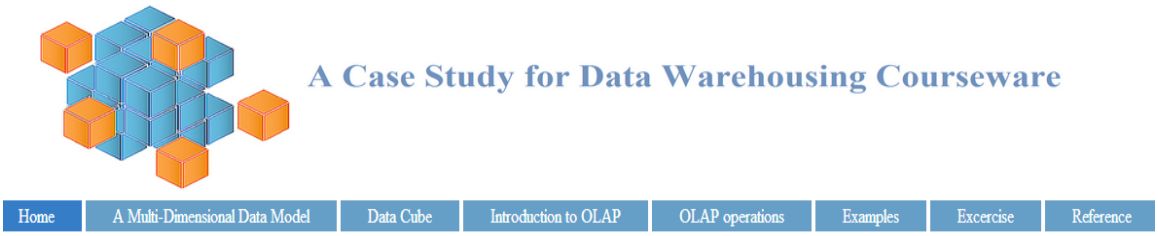


Figure 19: Exercise Page

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;
}
```

```
    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">
    <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>
```

```
    <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="quiz.php">Excercise</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-weight:bold;">
        <p>Introduction
      </font>
    </td>
  </tr>
  <tr>
    <td colspan="2" >
         
```

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

</td>

</tr>

<tr>

<td >   <img src="http://athena.ecs.csus.edu/~biligers/msproject/DWOLAP.png" align="left" > </td>

</tr>

<tr>

<td>

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.

</td>

</tr >

<tr>

<td colspan="2">

<p> <a href="differencebw.php"><font style="font-family:'Calibri';

font-size:18px;

color:maroon;

```
font-weight:bold;">Difference between Data warehouse and other Operational
Database Systems</font></strong></a>
<br><font size="18"></font>
</p>
</td>
</tr>
</table>
//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;
```

```
    }
    #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">
  <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-weight:bold;">
        <p>Multi dimensional Data Model</p>
      </font>
    </td>
  </tr>
  <tr>
    <td >
          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.</font>
    </td>
  </tr>
  <tr>
    <td>
         
    </td>
  </tr>
  <tr>
    <td>
         
    </td>
```

```
        </tr>
        <tr>
          <td>
            <img src="http://athena.ecs.csus.edu/~biligers/msproject/mdfig.jpg" alt=" "
align="left" width="750"  height="300" hspace="20" >
          </td>
        </tr >
        <tr>
          <td>
               
          </td>
        </tr>
        <tr>
          <td>
               
          </td>
        </tr>
        <tr>
          <td>
            <font style="font-family:'Calibri';
              font-size:18px;
              color:maroon;
```

```
        font-weight:bold;">Cube </font><br />
      </td>
    </tr>
    <tr>
      <td >
            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.
      </td>
    </tr>
    <tr>
      <td>
        <font style="font-family:'Calibri';
        font-size:18px;
        color:maroon;
        font-weight:bold;">Measures </font><br />
      </td>
    </tr>
    <tr>
      <td >
```

    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.</p>

</td>

</tr>

<tr>

<td>

<font style="font-family:'Calibri';

font-size:18px;

color:maroon;

font-weight:bold;">Dimensions

</td>

</tr>

<tr>

<td >

    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.

</p>

</td>

```
</tr>
<tr>
  <td>
    <font style="font-family:'Calibri';
      font-size:18px;
      color:maroon;
      font-weight:bold;">Hierarchy  </font><br />
  </td>
</tr>
<tr>
  <td >
         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.
    </p>
  </td>
</tr>
<tr>
  <td>
    <font style="font-family:'Calibri';
      font-size:18px;
```

```
        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>
    <tr >
      <td>
        <font style="font-family:'Calibri';
```

```
            font-size:22px;
            color:maroon;
            font-weight:bold;"> The Relational Implementation of the Model </font>
          </td>
        </tr>
        <tr>
          <td >
              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</p>
          </td>
        </tr>
        <tr>
          <td>
            <font style="font-family:'Calibri';
            font-size:18px;
            color:maroon;
            font-weight:bold;"> SnowFlake Schema</font>
              
          </td>
        </tr>
```

<tr>
<td>

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.</p>
</td>
</tr>
<tr>
<td>
<font style="font-family:'Calibri';
font-size:18px;
color:maroon;
font-weight:bold;">Star Schema
</td>
</tr>
<tr>
<td >

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.

```
            </p>
          </td>
        </tr>
        <tr>
          <td>
            <img src="http://athena.ecs.csus.edu/~biligers/msproject/Star.png" alt=" "
align="left" width="750"  height="300"  >
          </td>
        </tr >
      </table>
      //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" />
          <!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">
<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>
```

```
<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" >
<td>
<font style="font-family:'Calibri';
 font-size:22px;
 color:maroon;
 font-weight:bold;"><p>Data cube </font><br />
</td></tr>
<tr><td  >
   
```

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 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.

```
</p></td></tr>
```

```
<tr>
<td>
<font style="font-family:'Calibri';
 font-size:22px;
 color:maroon;
 font-weight:bold;"> Representation </font>
</td>
</tr>
<tr>
<td>
   
Consider comic books distributor company sales fact table in North California
region.</p>
</td>
</tr>
<tr>
<td>
Example for  two Dimensional representation of Data Cube </p>
</td></tr>
<tr>
<td>
   
```

```
</td>
</tr>
<tr>
<td>
<img align="left"    src="http://athena.ecs.csus.edu/~biligers/msproject/cube1.jpg"
alt=" "  >
</td></tr >
<tr>
<td >
<ul>
<li>Multi-dimensional databases are a compact and easy-to-understand way of
visualizing and manipulating data elements that have many inter-relationships. </li>
<li>The cube can be expanded to include another dimension, for example, sales
fact compared with SouthCalifornia region</li>
</ul>
</td>
</tr>
<tr>
<td>
     Example for  three Dimensional representation of Data
Cube </td>
</tr>
```

```
<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">
```

```
<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 class ="current" 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>
<tr>
<td>
<font style="font-family:'calibri';
  font-size:22px;
  color:maroon;
  font-weight:bold;"> <p>On-line Analytical Processing (OLAP)</p></font>
```

</td></tr>

<tr><td >

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.

</td></tr>

<tr>

<td >

<p>Click to know about OLAP Architecture

</td>

</tr>

<tr>

<td>

<font style="font-family:'calibri';

font-size:18px;

color:maroon;

```
  font-weight:bold;">  OLAP Categories </font>
</td></tr>
<tr><td >
   
```

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.</p>

```
</td>
</tr
<tr>
<td>
<font style="font-family:'calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">
Relational OLAP(ROLAP)
</font>
</td>
</tr>
<tr>
<td>
   
```

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>

</td>

</tr>

<tr>

<td>

<font style="font-family:'calibri';

font-size:18px;

color:maroon;

font-weight:bold;">

Multidimensional OLAP (MOLAP)

</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>

```
<tr>
<td>
<a href="ROLAPandMOLAP.php"><strong>Click know about more about
ROLAP and MOLAP categories  </strong></a></p>
<br>
</td>
</tr>
<tr>
<td>
<font style="font-family:'calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">
Hybrid OLAP (HOLAP)
</font>
</td>
</tr>
<tr>
<td>
   
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.</p>

</td>

</tr>

<tr>

<td>

<font style="font-family:'calibri';

 font-size:18px;

 color:maroon;

 font-weight:bold;">

Desktop OLAP (DOLAP)

</td>

</tr>

<tr>

<td>

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</p>

</td>

```
</tr>
<tr>
<td>
<a href="HOLAPandDOLAP.php"><strong>Click  know about more about HOLAP and DOLAP categories  </strong></a></p>
<br><font size="18"></font>
</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;
}
.currentLink {
color:red;
}
#navcontainer li a.current{
```

```
color: #fff;
background-color: #357EC7;
}
#header {
background-color:#996633;
color:white;
text-align:center;
padding:1px;
}
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">
<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 class ="current" 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>
<tr>
<td>
<font style="font-family:'calibri';
```

```
 font-size:22px;
 color:maroon;
 font-weight:bold;">OLAP Operations</font>
</td></tr>
<tr><td>
   
</br>
```

OLAP provides user with the flexibility to view data from different perspectives hence we will discuss the OLAP operations in multidimensional data.

```
<tr>
</tr>
</td></tr>
</tr>
<tr>
<td>
<font style="font-family:'calibri';
 font-size:18px;
 color:maroon;
 font-weight:bold;">ROLL UP </font>
</td></tr>
<td width="100%"style="vertical-align:top">
<ul>
```

 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 then one or more dimensions from the data cube are removed.

</td>

</tr>

<tr>

<td>

<font style="font-family:'calibri';

font-size:18px;

color:maroon;

font-weight:bold;">Drill-down

</td></tr>

<tr><td width="100%"style="vertical-align:top">

Drill down is the reverse of rollâ€ up. Navigates fr more detailed data it can achieved by any of the following way.

Stepping down a concept hierarchy for a dimension

Introduces additional dimensions


```
</td>
</tr>
<tr>
</tr>
<tr>
<td style="vertical-align:top" width="100%">
Below figure shows the demo of Roll up and Drill down operations</font>
</br></br>     <img align="left"
src="http://athena.ecs.csus.edu/~biligers/msproject/Rollup.jpg" alt=" "  >
</td> </tr>
<tr>
<tr>
<td>
<font style="font-family:'calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">Slice </font>
</td></tr>
<tr><td width="100%"style="vertical-align:top">
<ul>
<li>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

</td>

</tr>

<tr>

<td>

<font style="font-family:'calibri';

font-size:18px;

color:maroon;

font-weight:bold;">Dice

</td></tr>

<tr><td width="100%"style="vertical-align:top">

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

</td>

</tr>

<tr>

<td style="vertical-align:top" width="100%">

```
Below figure shows the demo of slice and Dice operations</font>
</br></br>     <img align="left"
src="http://athena.ecs.csus.edu/~biligers/msproject/sliceanddice.jpg" alt=" "  >
</td> </tr>
</br>
<tr>
<td>
<font style="font-family:'calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">Pivot (or rotate)</font>
</td></tr>
<td width="100%"style="vertical-align:top">
<ul>
<li>Rotates the data axis to view the data from different perspectives.
<li>Groups data with different dimensions
</ul>
</td>
</tr>
<tr>
<td style="vertical-align:top" width="100%">
Below figure shows the demo of OLAP Pivot operations</font>
```

```
</br></br>     <img align="left"
src="http://athena.ecs.csus.edu/~biligers/msproject/Pivot.png" alt=" "  >
</td> </tr>
<tr>
<td>
<font style="font-family:'calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">Some more OLAP operations are listed below</font>
</td></tr>
<tr>
<td>
<font style="font-family:'calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">Drill-across</font>
</td>
</tr>
<tr>
<td style="vertical-align:top">
<ul>
<li>An additional drilling operation.
```

```
<li>Executes queries involving more than one fact table
</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>Uses relational SQL facilities to drill through the bottom level of a data cube down to its back end relational tables
</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 {
```

```
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">
<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">OLAP operations</a></li>
<li><a class ="current" 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>
<font style="font-family:'Calibri';
  font-size:22px;
  color:maroon;
  font-weight:bold;"><p>Examples of OLAP Operations</p></font>
</td></tr>
<tr><td >
   
```

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.</p>  </td></tr>
<tr><td> <font style="font-family:'Calibri';
```

```
 font-size:18px;
 color:maroon;
 font-weight:bold;"> Dimension Tables</font></td></tr>
<tr><td>
<ul>
<li><a  href="BooksNameTable.png"><b>BooksName</a></li>
<li><a href="TimeTable.png"><b>Time</b></a></li>
<li><a href="BookCategories.png"><b>BooksCategories</a></li>
<li><a href="DistributorsTable.png"><b>Distributors</a></li>
<li><a href="Publisher.png"><b>Publishers</a></li>
</ul>
</td>
</tr>
<tr><td> <font style="font-family:'Calibri';
 font-size:18px;
 color:maroon;
 font-weight:bold;"> Fact Tables</font></td></tr>
<tr><td>
<ul>
<li><a href="Facttable.png"><b>ComicBooksFactTable</a></li>
</ul>
</td>
```

```
</tr>
<tr><td>
   
The below Star schema shows the way Data has organized in the data Model</p>
</td></tr>
<tr>
<td>     <img
src="http://athena.ecs.csus.edu/~biligers/msproject/Stardataschema.png" alt=" "
align="left" width="650"  height="400" >
</td>
</tr>
<tr>
<td>
<font style="font-family:'Calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">Roll up operation</font>
</td></tr>
<tr><td >
   
```

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</p>
</td></tr>
<tr>
<td vertical-align:Top; font-family:'Calibri';text-align: left width="10%">
<a href="rollup.php"><strong>Rollup Operation example</strong></a></p>
<font size="18"></font>
</td>
</tr>
<tr><td>
</td>
</tr>
<tr>
<td>
<font style="font-family:'Calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">Drill down operation</font>
</td></tr>
<tr><td >
```


Drilling down is analytical technique where 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.</p>

</td></tr>

<tr>

<td vertical-align:Top; font-family:'Calibri';text-align: left width="10%">

Drill down Operation example</p>

</td>

</tr>

<tr>

<td>

<font style="font-family:'Calibri';

font-size:18px;

color:maroon;

font-weight:bold;">Slice operation

</td></tr>

<tr><td >

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 </p>

```
</td></tr>
<tr>
<td vertical-align:Top; font-family:'Calibri';text-align: left width="10%">
<a href="sliceex.php"><strong>Slice Operation example</strong></a></p>
<font size="18"></font>
</td>
</tr>
<tr>
<td>
<font style="font-family:'Calibri';
  font-size:18px;
  color:maroon;
  font-weight:bold;">Dice operation</font>
</td></tr>
<tr><td >
   
   Dice selects two or more dimensions from a given cube and provides a new sub-
  cube. </p>
</td></tr>
```

```
<tr>
<td vertical-align:Top; font-family:'Calibri';text-align: left width="10%">
<a href="diceex.php"><strong>Dice Operation example</strong></a></p>
<font size="18"></font>
</td>
</tr>
/table>
```

REFERENCES

[1] W. H. Inmon, *Building the Data Warehouse*, John Wiley & Sons Inc., New York,USA, 2002.

[2] Microsoft Sql Server, Data Warehousing and OLAP [On-line serial], https://technet.microsoft.com/en-us/library/aa197903(v=sql.80).aspx.

[3] Ralph Kimball, Margy Ross, *The Data Warehouse Toolkit: the Complete Guide to Dimensional Modeling*, John Wiley & Sons Inc., New York, USA, 2013.

[4] S. Chaudhuri, U. Dayal, "An overview of data warehousing and OLAP technology" in *Proceedings of ACM Sigmod Record*, New York, USA, 1997.

[5] W. Giovinazzo, *Object-Oriented Data Warehouse Design: Building a Star Schema*, Prentice-Hall, New Jersey, USA, 2000.

[6] ICV2'S TOP 300 COMICS & TOP 300 GN'S INDEX, http://icv2.com/articles/news/view/1850/icv2s-top-300-comics-top-300-gns-index.

[7] Oracle® OLAP Application Developer's Guide, http://web.stanford.edu/dept/itss/docs/oracle/10g/olap.101/b103/multimodel.html.

[8] MOLAP, ROLAP, And OLAP, http://www.1keydata.com/datawarehousing/molap-rolap.html.