DWH ASSIGNMENT

ASSIGNMENT TOPICS:
1. OLAP OPERATIONS AND ADVANTAGES
2. OLAP TECHNIQUES

By 19SW42

Zohaib Hassan Soomro

Data Ware Housing

1.   OLAP OPERATIONS AND ADVANTAGES

OLAP(Online Analytical Processing):

OLAP is a category of software that allows users to analyse information from multiple database systems at the same time. It is a technology that enables analysts to extract and view business data from different points of view.

Analysts frequently need to group, aggregate and join data. These OLAP operations in data mining are resource intensive. With OLAP data can be pre-calculated and pre-aggregated, making analysis faster.
OLAP databases are divided into one or more cubes. The cubes are designed in such a way that creating and viewing reports become easy. OLAP stands for Online Analytical Processing.

OLAP cube:
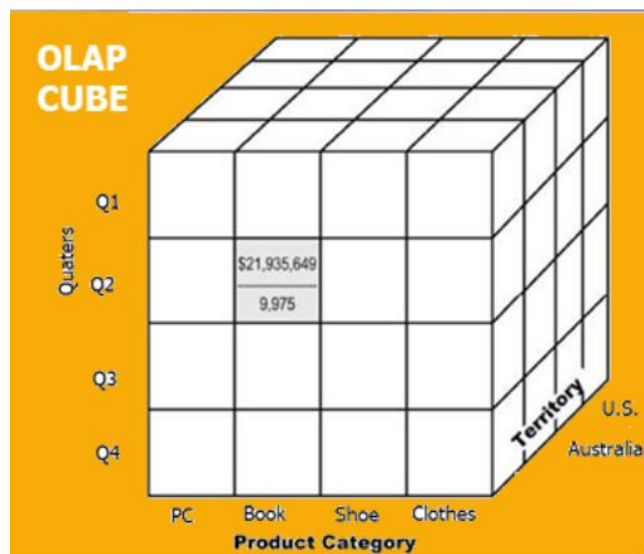
OLAP Cube

At the core of the OLAP concept, is an OLAP Cube. The OLAP cube is a data structure optimized for very quick data analysts.

The OLAP Cube consists of numeric facts called measures which are categorized by dimensions. OLAP Cube is also called the **hypercube**. Usually, data operations and analysis are performed using the simple spreadsheet, where data values are arranged in row and column format. This is ideal for two-dimensional data. However, OLAP contains multidimensional data, with data usually obtained from a different and unrelated source. Using a spreadsheet is not an optimal option. The cube can store and analyse multidimensional data in a logical and orderly manner.

How does it work?

A Data warehouse would extract information from multiple data sources and formats like text files, excel sheet, multimedia files, etc.

The extracted data is cleaned and transformed. Data is loaded into an OLAP server (or OLAP cube) where information is pre-calculated in advance for further analysis.

Basic analytical operations of OLAP

Four types of analytical OLAP operations are:

1. Roll-up
2. Drill-down
3. Slice and dice
4. Pivot (rotate)

1. Roll-up:

Roll-up is also known as "consolidation" or "aggregation." The Roll-up operation can be performed in 2 ways

1. Reducing dimensions
2. Climbing up concept hierarchy. Concept hierarchy is a system of grouping things based on their order or level.

Consider the following diagram
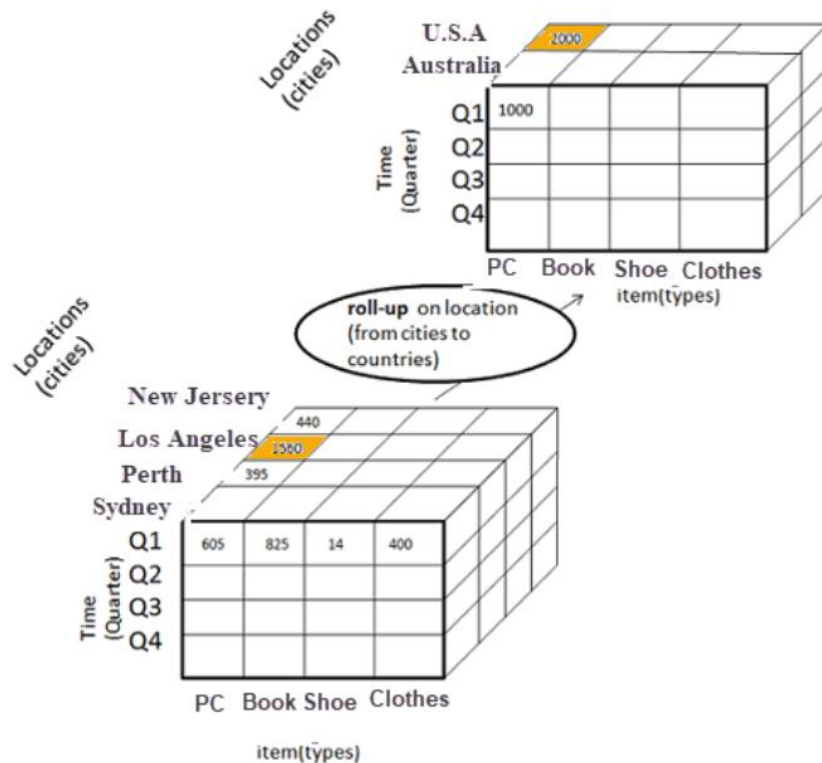
Roll-up operation in OLAP

- In this example, cities New jersey and Lost Angles and rolled up into country USA
- The sales figure of New Jersey and Los Angeles are 440 and 1560 respectively. They become 2000 after roll-up
- In this aggregation process, data is location hierarchy moves up from city to the country.
- In the roll-up process at least one or more dimensions need to be removed. In this example, Cities dimension is removed.

2. Drill-down
In drill-down data is fragmented into smaller parts. It is the opposite of the rollup process. It can be done via

- Moving down the concept hierarchy
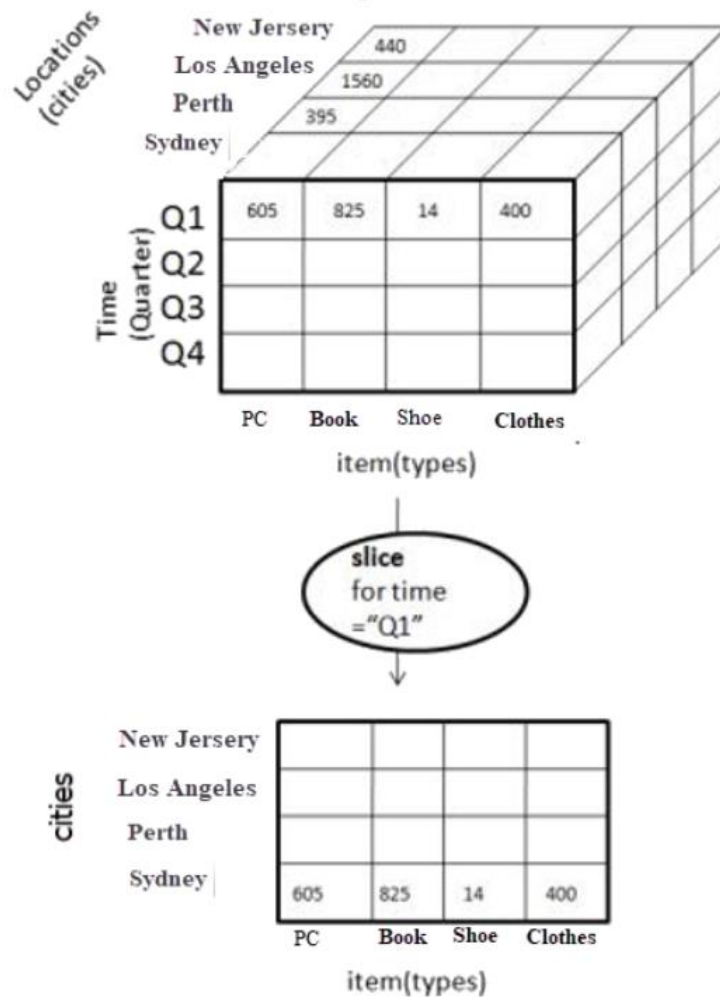- Increasing a dimension
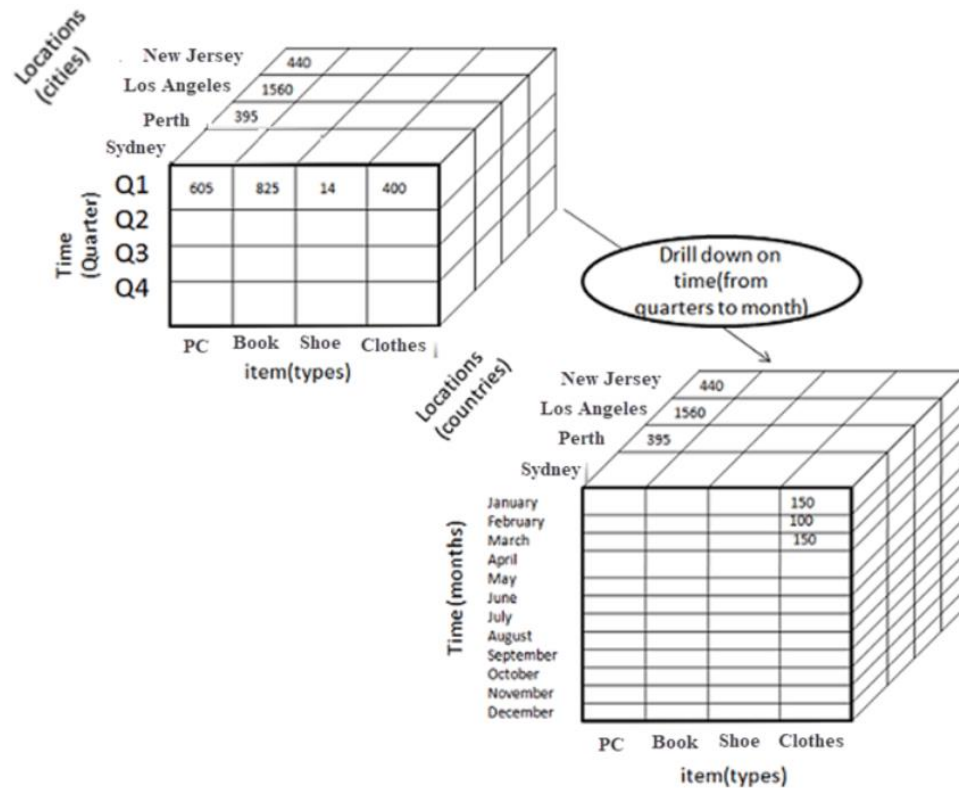
Drill-down operation in OLAP

Consider the diagram above

- Quater Q1 is drilled down to months January, February, and March. Corresponding sales are also registers.
- In this example, dimension months are added.

3. **Slice:**
   - Here, one dimension is selected, and a new sub-cube is created.
   - Following diagram explain how slice operation performed:

- Dimension Time is Sliced with Q1 as the filter.
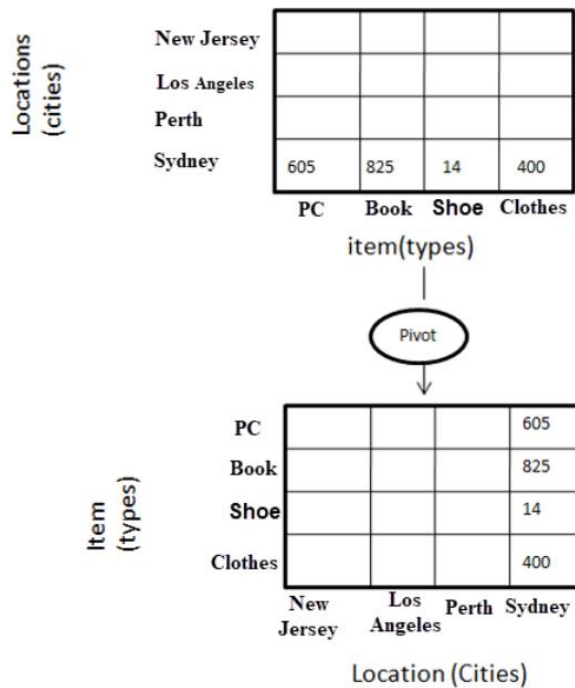- A new cube is created altogether.

Dice:

- This operation is similar to a slice. The difference in dice is you select 2 or more dimensions that result in the creation of a sub-cube.

Dice operation in OLAP
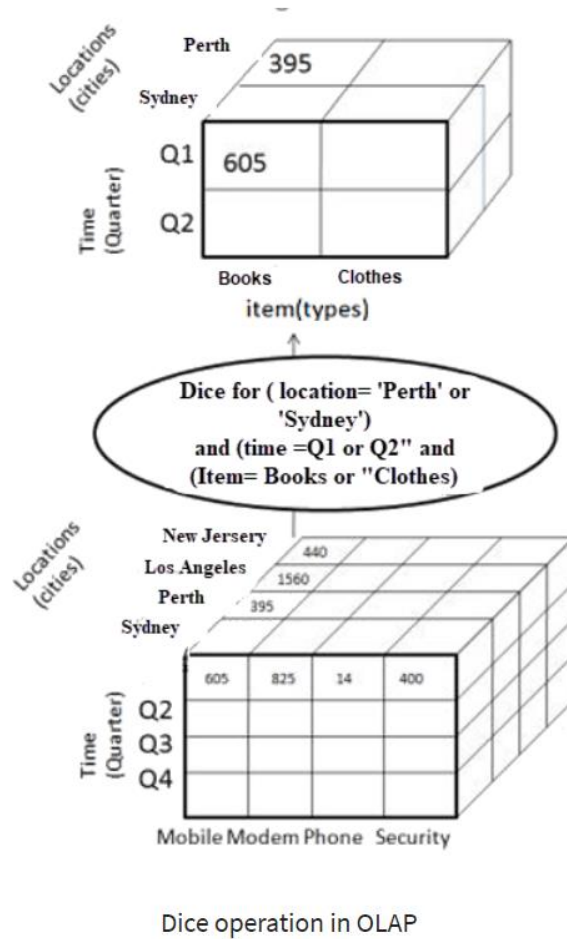
4) Pivot

In Pivot, you rotate the data axes to provide a substitute presentation of data.

In the following example, the pivot is based on item types.

Pivot operation in OLAP

Advantages of OLAP:

- OLAP is a platform for all type of business includes planning, budgeting, reporting, and analysis.
- Information and calculations are consistent in an OLAP cube. This is a crucial benefit.
- Quickly create and analyze "What if" scenarios
- Easily search OLAP database for broad or specific terms.
- OLAP provides the building blocks for business modeling tools, Data mining tools, performance reporting tools.
- Allows users to do slice and dice cube data all by various dimensions, measures, and filters.
- It is good for analysing time series.
- Finding some clusters and outliers is easy with OLAP.
- It is a powerful visualization online analytical process system which provides faster response times.

Disadvantages of OLAP

- OLAP requires organizing data into a star or snowflake schema. These schemas are complicated to implement and administer
- You cannot have large number of dimensions in a single OLAP cube
- Transactional data cannot be accessed with OLAP system.
- Any modification in an OLAP cube needs a full update of the cube. This is a time-consuming process

Reference:

https://www.guru99.com/online-analytical-processing.html

2.   OLAP TEHNIQUES

There are different OLAP types based upon the storage modes which are utilized for data discovery with complex analytical calculations, limitless record viewing, and predictive "what if" scenarios.

i)   Multi-dimensional OLAP (MOLAP):

These are broadly known as the classic OLAP type. This server utilizes a multi-dimensional Database (MDDB) for storing and analysing information. MDDB can proficiently store summaries, giving a method for quick questioning and recovering information from the database for processing. Multi-dimensional database management system and users visualize the held data as a 3-D Cube. These Data blocks are put away in memory called CubeCache.

Information is moved from an information source into the multidimensional database which sits between the customer and the server, and after that, the database is aggregated. MOLAP stores the data summaries in binary files and keep it away from the relational database. It is imperative to comprehend that MOLAP additionally makes a duplicate copy of the fact and dimension information in a unique binary file.

It offers fast indexing to precomputed aggregations, thus making it faster for computations. It effectively stores and works with numeric information.

In any case, the constraints for the utilization of MOLAP incorporates:

Less versatility as it can just deal with a restricted measure of information, the speed of MOLAP is quicker for little to medium informational indexes yet normal for bigger informational indexes, the MOLAP storage may likewise incorporate excess information.

MOLAP Applications: Essbase, Express Server, Yellowfin, Clear Analytics, SAP Business Intelligence.

ii) ROLAP; Relational On-Line Analytical Processing:

The term ROLAP indicates that the OLAP server involves storage that is relational in nature. They utilize a relational database management system to keep and control the data. These are the servers that exist between the database and the user. ROLAP systems work on the information that resides in a relational database.

ROLAP builds indexed views to store the data summaries in those views in the relational database. It additionally leaves both, fact and dimension, the information in the relational table.

ROLAP servers support large amounts of information than MOLAP servers do. It productively oversees both numeric and textual information. It allows clients to drill down to the most minimal level of a hierarchy structure

DSS server of Microstrategy embraces the ROLAP approach

The main argument against RDBs is that querying a massive database with SQL to get information usually brings about complex queries. Subsequently, ROLAP applications show a slower performance and may not be perfect for the performance of certain calculations.

iii) HOLAP; Hybrid OLAP:

It is a blend of MOLAP and ROLAP. By utilizing both **ROLAP AND MOLAP** information stores, Hybrid OLAP offers the qualities of both techniques. HOLAP stores data summaries in the binary files

or in the pre-calculated cubes. It leaves the quantities of fact and dimension information in the relational database.

HOLAP approach can be commonly executed if any of the accompanying circumstances exists:

If there exists a massive amount of data, if there is performance congestion on a server or if you are making use of saved information sources which are summarized.

The HOLAP servers store information in the most functional way possible. The HOLAP component offers flexibility in designing, accessing, and maintaining HOLAP data groups. HOLAP integrates both multi-dimensional and relational information storage that can be utilized to address issues with scalability and performance.

Apart from the previously mentioned three sorts which are generally identified and utilized, we additionally have a couple of more kinds of OLAP which are not all that prevalent yet unquestionably worth referencing and understanding.

iv) WOLAP:

A Web OLAP which is known as Web-enabled OLAP is utilized through the internet browser. It very well may be utilized on the off chance that you are thinking about something on a truly minimal effort spending plan since it requires just a web connection and an internet browser to get to the information. In contrast with different kinds of OLAP, WOLAP functionality and performance are undermined.

v) DOLAP:

Desktop On-Line Analytical Processing (DOLAP) is a single-tier, desktop-based OLAP technology. Functionality is limited in contrast to other OLAP applications. It has a more cost-effective value and is beneficial for mobile clients who can't generally connect to the data warehouse.

vi) SOLAP:

The wide use of geo-referenced information has added the need to upgrade OLAP with spatial analysis tools. Spatial OLAP (SOLAP) devices have been implemented to tackle issues in the field of Geo-Business Intelligence. SOLAP tools help perform spatial analysis of information. These devices combine OLAP analysis with GIS frameworks for spatial visualization. There are nevertheless lookup going on to enhance technological know-how for improved optimization of complicated queries and data visualization.

Reference:

https://www.educba.com/types-of-olap/