Abstract: In today’s scenario, large amount of data is available for taking strategic decisions inspite of the fact that large amount of data is available, this data is not sufficient for strategic decision making. SQL (Structured Query Language) is not answerable to complex queries. Then OLAP (Online Analytical Processing) tool can solve complex queries in high quality and highly efficient manner. This research Paper discusses characteristics of OLAP and how OLAP completes the responsibility of SQL by solving these complex queries.

Keywords: OLAP, data warehouse, complex queries, SQL

1. Introduction:-

Although sometimes used interchangeably, the terms data warehousing and online analytical processing (OLAP) apply to different components of systems often referred to as decision support systems or business intelligence systems. Components of these types of systems include databases and applications that provide the tools analysts need to support organizational decision-making.

A data warehouse is a database containing data that usually represents the business history of an organization. This historical data is used for analysis that supports business decisions at many levels, from strategic planning to performance evaluation of a discrete organizational unit. Data in a data warehouse is organized to support analysis and integrity rather than to process real-time transactions as in online transaction processing systems (OLTP).

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. The answer to a query into historical data often leads to subsequent queries as the analyst searches for answers or explores possibilities. OLAP systems provide the speed and flexibility to support the analyst in real time.

Why we use OLAP

An analyst may drill down into data to see. For example, how an individual salesperson’s performance affects monthly revenue numbers. At the same time, the drill-down procedure may help the analyst discover certain patterns in sales of given products? This discovery can force another set of questions of similar or greater complexity. Technically, all these analytical questions can be answered by a large number of rather complex queries against the set of detailed and pre summarized data views. In reality however if the analyst could quickly and accurately formulate SQL statements of this complexity, the response time and resource consumption problem would still persist and the analyst productivity would be seriously impacted. So we introduce concept of OLAP in place of SQL.

Drawbacks of SQL (Structured Query Language)

1. SQL: A query may translate into a number of complex SQL statements, each of which may involve full table scan, multiple joins, aggregations and sorting, and large temporary tables for storing intermediate results. The resulting query may require significant computing resources that may not be available at all times and even then may take a long time to complete.

2. Another drawback of SQL is its weakness in handling time series data and complex mathematical functions. Time series calculations such as 3 month moving average or net present value calculations typically require extensions to ANSI SQL rarely found in commercial products

Characteristics of OLAP:-

- Fast: Data is organized for rapid query and analysis. The database structure uses efficiency multidimensional or tuned relational approaches.
- Visual: Tools enable the analyst to navigate and view results through graphics such as bar charts, pie charts and tree structures.
- Multidimensional: Supports slicing and dicing along multiple dimensions such as product, customer and location. In addition, supports “pivot” or “cross tabs” where the investigator changes the direction of the analysis.
- Aggregation: Supports both drill down to details as well as roll up. Some aggregations may be pre-calculated to save analysis time. This pre-calculation is where MOLAP provides value.
- Time Series: Supports trend analysis. Most data must include a calendar dimension. This dimension supports time horizons: year, quarter, month, week, day of week, and day.
- Ranking: Find the top, bottom or quartile members of a group, such as the top 20 most profitable products or the 10 least profitable sales territories.
- Clusters and Outliers: Gain an understanding of groupings of items with common characteristics (clusters) as well as items with unusual characteristics (outliers).

What data stored in data ware house

In simple words: Subject(s) per Dimension

Example: If our subject/measure is ‘quantity sold’ and if the dimensions are: Item Type, Location and Period then, Data warehouse stores...
the items sold per type, per geographical location during the particular period. This data is represented by Data Cube as shown in Fig 1. 

This figure 1 shows a 3-Dimensional data model. X Dimension: Item type, Dimension: Time/Period Z – Dimension: Location Each cell represents the items sold of type ‘x’, in location ‘z’ during the quarter ‘y’. This is easily visualized as Dimensions are 3. What if we want to represent the store where it was sold too? We can add more dimensions. This makes representation complex. Data cube is thus a n-dimensional data model. Data is extracted from this cube using OLAP operations.

Operations of OLAP which helps to solve complex queries.

These operations helps to find out the data from data marts. A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as Sales, Finance, or Marketing. Data marts are often built and controlled by a single department within an organization.

1. Roll-up

Performs aggregation on a data cube, either by climbing up a concept hierarchy for a dimension or by dimension reduction. See fig 2.7

2. Drill-down

Can be realized by either stepping down a concept hierarchy for a dimension or introducing additional dimensions.

3. Slice and Dice

Slice performs a selection on one dimension of the given cube, resulting in a sub cube. Dice defines a sub cube by performing a selection on two or more dimensions.

4. Pivot (rotate)

It’s a visualization operation that rotates the data axes in view in order to provide an alternative presentation of the data.

Queries which are effectively solved by OLAP in place of SQL.

Query 1. List the maximum number of age of persons at which crime occurs in particular days in particular country. See fig 2.

Query 2. In data warehouse shows the affected area of data marts. Or Decision table that shows affected attributes during user queries in drill down mode. See fig 3.

Above solved queries will take a lot of time and resources to be solved in SQL and still we cannot sured that we will get expected results because the process and query used in SQL both are complex.

Conclusion:

This research paper discusses features of OLAP and how OLAP can be used in place of SQL to solve complex queries which are time taken or which can be proof wrong with respect to the expected result. This research paper also discusses how OLAP can be used where SQL fails.

References:


