

## Module 5 (online Analytical Processing)

### \* Introduction to OLAP

- OLAP (online Analytical Processing) is a type of database technology which is designed to support complex queries and analysis of large amounts of data.
- OLAP is mainly used in data warehousing, business intelligence and decision support system.
- OLAP is characterized by its ability to store and manipulate data in a multi-dimensional format. Rather than using the traditional 2-D relational data models, OLAP utilizes a 3-D multi-dimensional cube format.

### \* The Need for OLAP

- ① Analyzing large volumes of data - OLAP is designed to handle and analyze large volumes of data quickly and efficiently.

for example - lets say a retail company wants to analyze its sales data from the past 5 years. The dataset would likely be very large, with a lot of rows and columns. With the multi-dimensional cube format of OLAP, analysis is made much

easier and identifying trends and patterns in the data and making informed business decisions.

- ② Multi-Dimensional Analysis - OLAP allows users to analyze data from multiple dimensions such as time, geography, product, customer making it easier to understand relations and co-relations b/w different data-points.
- ③ Flexible-reporting - OLAP enables users to create flexible reports and dashboards that can be customized to their specific needs. OLAP operations can be used for detailed and summarized view.
- ④ Improved decision making - By providing decision makers with the insights they need, to make informed decisions, OLAP helps businesses and organisations improve their operations and performance.

Data is analyzed from multiple dimensions leading to much accurate decisions.

③ Faster query response times - OLAP is optimized for analytical queries, making it faster to analyze, than traditional databases. Methods like aggregation and indexing are used to improve response times.

### \* Difference in OLTP and OLAP:

Criteria	OLTP	OLAP
① Purpose	OLTP is used to automate real-time transactions.	OLAP is used to analyze data for decision making.
② Design	OLTP is designed to be normalized, to minimize redundancy and maintain consistency.	OLAP is designed to be dimensional.
③ Data Volume	OLTP mainly handles low volumes of data as transactions occur in real-time.	OLAP occurs for high volumes of data for trend analysis and pattern generation.
④ Response Time	OLTP boasts fast response time as queries are simple for retrieving and updating individual records.	OLAP has a slower response time due to complex queries which involves aggregating data from multiple dimensions.

Criteria	OLTP	OLAP
⑤ Type of data	OLTP stores detailed data, such as individual data such as order or transaction.	OLAP stores summarized data such as monthly or yearly sales.

### \* OLAP and Multi-Dimensional Analysis.

→ Multi-Dimensional Analysis can be achieved by structuring data into a dimensional model. This dimensional model consists of :-

- ① Dimensions - These are for the different attributes and characteristics of data within the dataset for eg- time, geography, product etc.
- ② Hierarchies - Hierarchies define relationships b/w different levels of the model. For example, yearly, quarterly, monthly
- ③ Measures - These are the numerical values that are to be analyzed. Eg. sales and profit, revenue.

→ In OLAP, cube is the data structure which stores data in a multi-dimensional format. It can be thought of as a 3-D array, where each axis represents a different dimension.

## \* OLAP operations in Multi-dimensional Data Model.

→ In OLAP, there are 5 key operations used to manipulate and analyze data. Also known as the OLAP cube operations.

- ① Roll-up (Consolidation) - Roll-up is the operation which aggregates data along one or more dimensions. It involves moving up the hierarchy to a higher level dimension or a ~~finer~~ coarser level of granularity.
- ② Drill-down (Expansion) - Drill-down is an operation that disaggregates data along one or more dimensions. It involves moving to a lower-level dimension or a finer level of granularity.
- ③ Slice - The slice operation in OLAP is used to select a subset of data from the cube based on one or more criteria. It allows the user to slice through the cube and extract a particular view of data that meets specific criteria.
- ④ Dice - The dice operation is used to make a new sub-cube out of the original cube for multiple dimensions which satisfy a specific criteria. The resultant cube has a smaller size than the original cube.

⑤ Pivot (Rotation) - Pivot is the operation which re-organises the data in the cube along one or more dimensions. It involves changing the orientation of the cube to view it from a different perspective.

### \* Query Examples for OLAP operations

#### ① Slice Query

```
Select [Measures].[Sales Revenue] ON COLUMNS
[Product].[Category].&[1] ON ROWS
FROM [SALES DATA]
WHERE [TIME].[YEAR].&[2019]
```

#### ② Dice Query

```
Select [Measures].[Profit Margin] ON COLUMNS
[Product].[Category].[Category].[Members]* 
[Geography].[Region].[Region].&[East] ON ROWS
FROM [SALES DATA]
WHERE [TIME].[Quarter].&[Q1-2022]
```

→ The query slices sales data by selecting only data from year 2019 and product category F1.

- The query slices sales data cube by selecting only data for first quarter of 2022 and product categories and regions of east.

## \* OLAP Models.

There are mainly 3 OLAP models.

① MOLAP (Multi-Dimensional OLAP) : In MOLAP, the data is stored in multi-dimensional arrays called cubes. This model is suitable for accessing small-to medium data sets as it provides fast access of data due to pre-aggregate nature of cubes. MOLAP provides an intuitive and user-friendly interface for exploration and analysis.

→ Main advantages of MOLAP

- ① Fast query Performance
- ② Efficient use of Storage
- ③ Intuitive Data Exploration
- ④ Real-time data analysis.

→ Main disadvantages of MOLAP

- ① Difficulty in handling large amounts of data in datasets.
- ② Requires lot of processing power to operate.

② ROLAP (Relational OLAP) :- In ROLAP, data is stored in relational databases, making it easier to handle large volumes of data. ROLAP is designed to handle complex queries and provide flexible analysis. However, ROLAP is slower than MOLAP as data is accessed from a relational database.

- ③ Most dimensions are stored in separate tables and fact tables store measures.
- Some advantages are:

① Scalability

② Data consistency

③ Flexibility

④ Cost-effective.

- Some disadvantages are:-

① Slow Processing time

② Complex to set up.

③ HOLAP (Hybrid OLAP) :- HOLAP is a combination of MOLAP and ROLAP. It stores the summarized data in a cube format and detailed data in the relational database table.

→ This approach allows for faster processing data which is summarized while providing access to detailed data.

In summary, MOLAP is suited for small datasets. ROLAP is suitable for handling large amounts of data, while HOLAP is used for both small and large datasets.