

Data Mining & Warehousing

Abhinav kumar Choudhary

Roll no - 21SCSE1011615

Sec - 5

* ----- * Assignment * ----- *

1) Difference between OLAP & OLTP?

Ans:

Difference between OLTP & OLAP are :-

OLTP	OLAP
(i) It stands for Online Transaction Processing.	It stands for Online Analytical Processing.
(ii) It is known as online database modifying system.	It is known as online database query management system.
(iii) Consist of only of operational current data.	consist of historical data from various database.
(iv) It makes use of DBMS.	It makes the use of warehouse
(v) It is application oriented.	It is subject oriented.
(vi) It is used for planning, day to day fundamental operation.	It is used for planning, problem-solving, decision-making, etc
(vii) Size of data is small as compared to OLAP (in MB or GB).	Size of data is very high. (in GB or TB).

OLTP

(viii) The data is managed by clerks, managers.

(ix) Both read & write operation

(x) It is focused to market

OLAP

The data is generally managed by CEO, MD, GM.

Supports only read & rarely write operation.

It is focused to the market.

2) Explain ROLAP, MOLAP & HOLAP ?

Ans → Comparison among ROLAP, MOLAP & HOLAP are :-

ROLAP

(i) It stands for Relational Online Analytical Processing.

(ii) Processing time of ROLAP is very slow.

MOLAP

It stands for Multi-dimensional Online Analytical Processing.

Processing time of MOLAP is fast.

HOLAP

It stands for Hybrid Online Analytical Processing.

Processing time of HOLAP is fast.

(vi) It is used as storage location for summary aggregation.

It is used as storage location for summary aggregation.

It is used as storage location for summary aggregation.

(iv) Large storage space required.

Medium storage space required.

Small storage space required.

(v) Low latency in ROLAP.

High Latency in MOLAP

Medium Latency in HOLAP.

(vi) Slow query response time in ROLAP.

Fast query response time in MOLAP.

Medium query response time in HOLAP.

3) Difference between star schema & snowflake schemas?
Also explain its advantage & disadvantage?

Ans → Difference between star schema and snowflake schemas

Star Schemas

Snowflake schemas

- | | |
|---|--|
| (i) The fact tables and the dimension tables are contained. | The fact table, dimension tables as well as sub dimension tables. |
| (ii) It is top-down model. | It is bottom-up model. |
| (iii) It uses more space. | It uses less space. |
| (iv) It takes less time for the execution of queries. | while it takes more time than star schemas for the execution of queries. |
| (v) Normalization is not used. | Normalization & denormalization are used. |
| (vi) Its design is very simple. | While its design is complex. |
| (vii) The query complexity is low. | Query complexity is high. |
| (viii) It is very simple. | It is complex than star schemas. |
| (ix) It has less number of foreign keys. | while it has more number of foreign keys. |
| (x) It has high data redundancy. | While it has low data redundancy. |

Advantage of Star schemas:-

- 1) Simpler Queries
- 2) Simplified Business Reporting Logic
- 3) Feeding Cubes

Disadvantage of Star schemas:-

- (i) Data integrity is not enforced well since in a highly de-normalized schemas state.
- (ii) Not flexible in terms of analytical needs as a normalized data model.
- (iii) Star schemas don't reinforce many-to many relationships within business entities - at least not frequently.

Snowflake Schemas:-

• Advantage:

- (i) It provides structured data which reduces the problem of data integrity.
- (ii) It uses small disk space because data are highly structured.

• Disadvantage:

- (i) Snowflake reduces space consumed by dimension tables but compared with the entire data warehouse the saving is usually insignificant.
- (ii) Multiple hierarchies that can belong to the same dimension have been designed at the lowest possible detail.

4) Difference between Datamart & Relational database ?
 Ans → Difference between datamart & database are :-

Data mart

Relational Database

1) A data mart is an analytical data Repository (OLAP)

A database is a transactional data.

2) It will house data from multiple subjects.

It captures all the aspects & activities of one subject in particular.

3) The data in it will be processed & validated for greater reporting ease.

The data in it will be raw and unprocessed (not cleaned).

4) It is the last step in the ETL process

It is the first step in the data ETL Process.

5) Users directly interact with data from a data mart

Users do not interact with data in a database.

5) What is data-cube materialization? List out some & identify which one is important & why?

Ans → Data cube materialization is the process of precomputing & storing the aggregates of a data cube, which allows for faster query response times. The different types of data cube materialization are :-

- Full Materialization
- No Materialization
- Half Materialization

- Full Materialization - All the possible combinations of dimensions, and measures are precomputed and stored in the data cube.
- Partial Materialization - Only a subset of data are precomputed and stored in the data cube.
- No Materialization - The data cube aggregates are not precomputed and stored, based on the and most frequently are instead computed at query time.

The most common type of data cube materialization depends on the specific use case & requirements of the organization. Full materialization is useful for environments with sufficient storage capacity & the data is relatively static, as it provides best query performance. However, it may not be easy for larger dataset for the huge amount of storage required. Here partial materialization perfect balance b/w storage & query performance. No materialization may be used for scenarios where the dataset is small & infrequent.

Q) Difference between multi-dimensional cube & Multidimensional database? Also, explain star-net query model?

Ans) Both are closely related concepts, but they are slightly different things.

A multidimensional cube is a data structure used in OLAP system to analyze large volumes of data from different perspective. It consist of a set of measures that are organised along multiple dimensions. For example - a sales cube might have dimensions such as - time region, & measures such as revenue, profit. The cube allows users to drill down, slice & dice.

On the other hand, a multi-dimensional database is a type of database that is optimized for storing & querying multi-dimensional data. It is designed to handle complex queries that involve multiple dimensions & measures.

Star Schemas is the fundamental schemas among the data mart schemas & it is simplest. The schemas is widely used to develop or build a data warehouse & dimensional data marts. It includes one or more fact tables indexing any number of dimensional tables. It is the necessary cause of the snowflake schemas. It is also efficient for handling basic queries. It is said to be 'star' as its physical model resembles to the star shape having a fact table a center & the dimension tables at its peripheral representing the star's point.

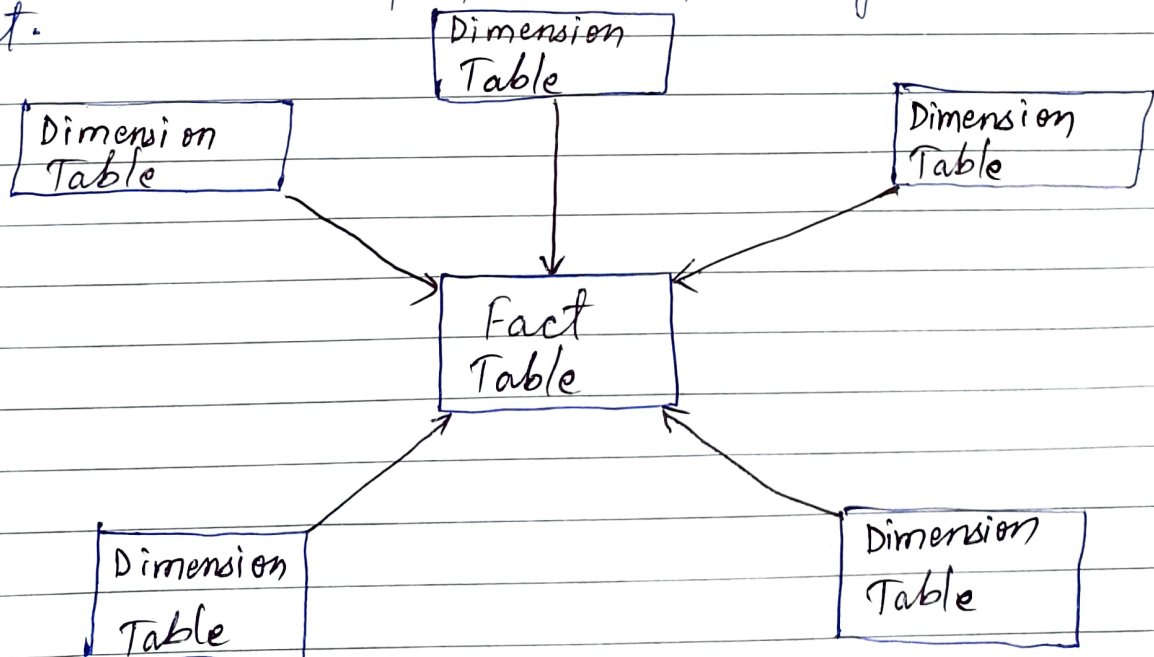


Diagram of Star Schemas