

## Data Warehouse

A data warehouse exhibits the following characteristics to support the management's decision-making process –

- **Subject Oriented** – Data warehouse is subject oriented because it provides us the information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue, etc. The data warehouse does not focus on the ongoing operations, rather it focuses on modelling and analysis of data for decision-making.
- **Integrated** – Data warehouse is constructed by integration of data from heterogeneous sources such as relational databases, flat files etc. This integration enhances the effective analysis of data.
- **Time Variant** – The data collected in a data warehouse is identified with a particular time period. The data in a data warehouse provides information from a historical point of view.
- **Non-volatile** – Nonvolatile means the previous data is not removed when new data is added to it. The data warehouse is kept separate from the operational database therefore frequent changes in operational database is not reflected in the data warehouse.

## Data Warehousing

Data warehousing is the process of constructing and using the data warehouse. A data warehouse is constructed by integrating the data from multiple heterogeneous sources. It supports analytical reporting, structured and/or ad hoc queries, and decision making.

Data warehousing involves data cleaning, data integration, and data consolidations. To integrate heterogeneous databases, we have the following two approaches –

- Query Driven Approach
- Update Driven Approach

## Query-Driven Approach

This is the traditional approach to integrate heterogeneous databases. This approach is used to build wrappers and integrators on top of multiple heterogeneous databases. These integrators are also known as mediators.

## Process of Query Driven Approach

Mentioned below is process of query driven data warehousing approach –

- When a query is issued to a client side, a metadata dictionary translates the query into the queries, appropriate for the individual heterogeneous site involved.
- Now these queries are mapped and sent to the local query processor.
- The results from heterogeneous sites are integrated into a global answer set.

## Disadvantages

This approach has the following disadvantages –

- The Query Driven Approach needs complex integration and filtering processes.
- It is very inefficient and very expensive for frequent queries.
- This approach is expensive for queries that require aggregations.

## Update-Driven Approach

Today's data warehouse systems follow update-driven approach rather than the traditional approach discussed earlier. In the update-driven approach, the information from multiple heterogeneous sources is integrated in advance and stored in a warehouse. This information is available for direct querying and analysis.

### Advantages

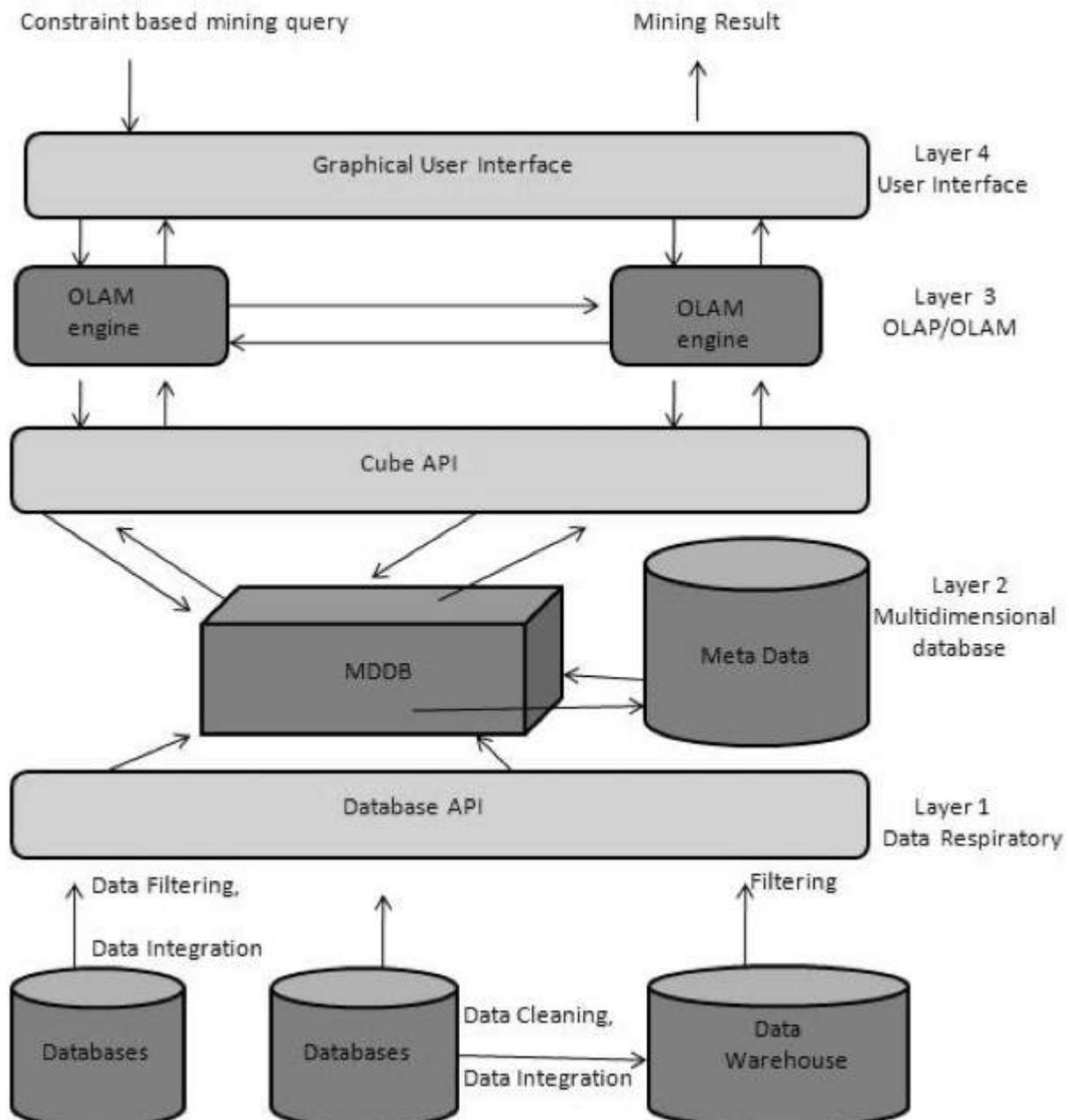
This approach has the following advantages –

- This approach provides high performance.
- The data can be copied, processed, integrated, annotated, summarized and restructured in the semantic data store in advance.

Query processing does not require interface with the processing at local sources.

### From Data Warehousing OLAP to Data Mining OLAM

Online Analytical Mining integrates with Online Analytical Processing with data mining and mining knowledge in multidimensional databases. Here is the diagram that shows the integration of both OLAP and OLAM –



## Importance of OLAM

OLAM is important for the following reasons –

- **High quality of data in data warehouses** – The data mining tools are required to work on integrated, consistent, and cleaned data. These steps are very costly in the preprocessing of data. The data warehouses constructed by such preprocessing are valuable sources of high quality data for OLAP and data mining as well.
- **Available information processing infrastructure surrounding data warehouses** – Information processing infrastructure refers to accessing, integration, consolidation, and transformation of multiple heterogeneous databases, web-accessing and service facilities, reporting and OLAP analysis tools.
- **OLAP–based exploratory data analysis** – Exploratory data analysis is required for effective data mining. OLAM provides facility for data mining on various subset of data and at different levels of abstraction.
- **Online selection of data mining functions** – Integrating OLAP with multiple data mining functions and online analytical mining provide users with the flexibility to select desired data mining functions and swap data mining tasks dynamically.

Loading [MathJax]/jax/output/HTML-CSS/jax.js