

Data warehouse & Data warehousing provides architectures and tools for business objectives to systematically organised. understand and use their data to make strategies decisions.

Data warehouse system are very valuable tools in today's competitive, fast evolving world. Data warehouse refers to a data repository that is maintained separately from an organisation operational data bases. Data warehouse system allows for integration of a variety of applications systems.

following features distinguish data warehouse from other data repositories such as RDBMS.

③ Subject oriented: A data warehouse is organized around major subjects such as customer, supplier, product and sales. Rather than concentrating on day-to-day operations and transactions processing of an organisation, a data warehouse focuses on the modeling and analysis of data for decision makers. Hence, data warehouses typically provide a simple and concise view of particular subject issues by excluding data that are not useful in the decision support process.

④ Integrated: A data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files and online transaction records. Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures and so on.

- ③ Time-variant : Data are stored to provide information from an historic perspective. Every key structure in the data warehouse contains either implicitly or explicitly a time element.
- ④ Non-volatile : A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment. Due to this separation, a data warehouse does not require transaction processing, recovery and concurrency control mechanisms. It usually requires only two operations in data processing initial loading of data and access of data.

How to organisation used information :

Many organisation use the information from data warehouse to support business decision making process including

- ① Increasing customer focus which includes the analysis of customer buying pattern such as buying preference, buying time, buying cycle, budget cycle etc.
- ② Forecasting & Repositioning products and Managing product portfolio's by

Comparing the performance of sales by ~~year~~ by geographic regions in order to fine tune production strategies.

Analysing operation and looking for sources of profit.

Managing customer relationship, making environment corrections etc.

Data warehousing is also very useful from the point of view of heterogeneous database integration.

Difference between operational DB system and Data warehousing.

The Major tasks of online operational DB system is to perform online transactions and query processor.

Such as system are called online transaction processing system (OLTP)

They cover most of the day-to-day operations of an organisation such as purchasing, inventory, manufacturing, Banking, payroll, registration and accounting. Data w.h. system on the other hand

serves its users in the role of data analysis and decision making such systems can organised and present data in various format in order to fulfill the different needs of different users. These systems are known as online analytical processing system(OLAP).

Difference b/w OLTP and OLAP. (MI)

features	OLTP.	OLAP
Characteristic	Operational processing system	Information processing system
orientation	Customer , transaction	Market Analysis
User	clerk , DBA , database professional	knowledge , worker , manager , executive analyst .
function	day to day operations	long - term informational requirements decision support .
DB Design	ER-based , application oriented	star / snowflake subject - oriented
manages	current data	historical data

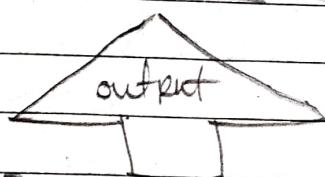
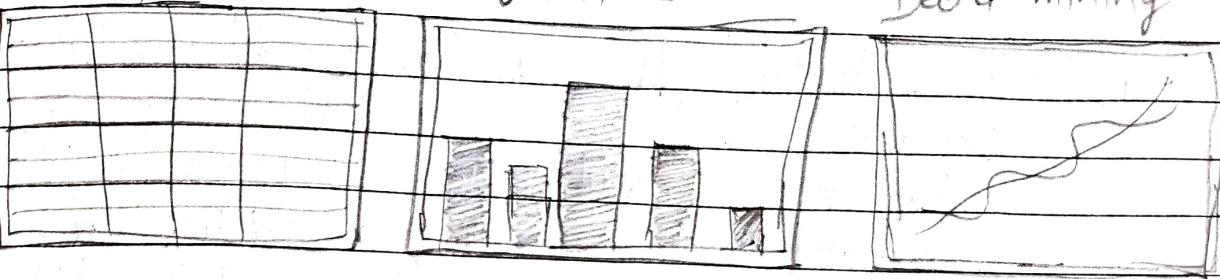
Data	guaranteed up-to-date	maintained over time
summization	primitive, highly detailed	summarized, detailed
view	detailed, fact relational	summarized consolidated
Unit of work	short, simple transaction	Complex query
Access	read/write	mostly read
Focus	data in	data out
Operations	index / hash on primary key	lots of scans
Number of records accessed	tens	millions
No of users	thousands	hundreds
DB size	GB to higher order GB	\geq DB
Priority	high performance high availability	high flexibility end-user autonomy

Query/report

Analysis

Data mining

Top tier
front -
end tools

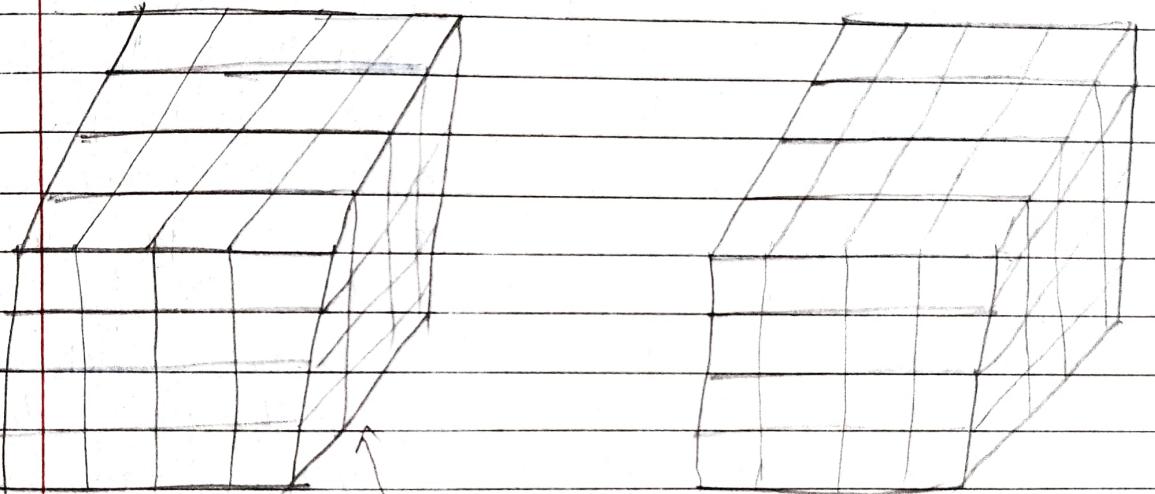


OLAP server

OLAP server

Middle
tier:

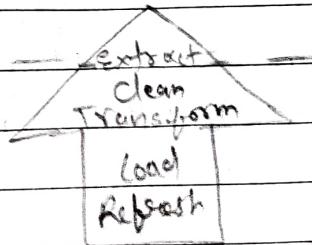
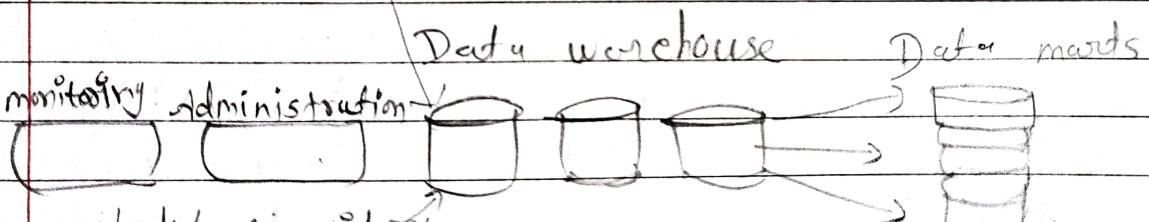
OLAP
server



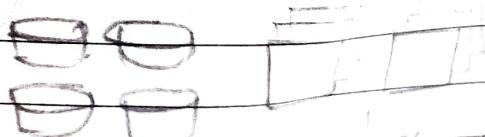
Bottom
tier:

metadata repository

Data ware-
house -
server



Operational databases



External sources

① The bottom tier is a warehouse DB server that is almost always a relational database system. Back-end tools and utilities are used to feed data into the bottom tier from operational databases or other external sources. These tools and utilities perform data extraction, cleaning and transformation as well as load and refresh functions to update the datawarehouse. The data are extracted using application program interfaces / known as gateway. A gateway is supported by the underlying DBMS and allows client programs to generate SQL code to be executed at a server. Examples of gateways include ODBC (open DB connection) and OLEDB (object linking and embedding Database) by Microsoft and JDBC (Java Database connection). The tier also contains a metadata repository, which stores information about the data warehouse and its contents. The metadata repository is further described in section 2.14.

(2) The Middle tier is an OLAP server that is typically implemented using either (1) a relational OLAP (ROLAP) model (i.e. an extended relational DBMS that maps operational & on Multidimensional data to standard relational operations.) or (2) a multi-dimensional OLAP (modOLAP) model (i.e. a special-purpose server that directly implements multidimensional data and operations.) OLAP servers are discussed in section 4.4.

(3) The top tier is a front-end client layer which contains query and reporting tools, analysis tools, and data mining tools.

~~Types of Data Warehouse :-~~
~~from the architecture point of view~~
~~there are 3 Data warehouse models :-~~

(1) Enterprise Warehouse :- This kind of warehouse collects all of the information about the subjects spanning the entire organisation.
 It provides corporate wide data usually from one or more operational systems or external information providers and it is cross functional in scope. It is

typically contains detailed data as well as summarize data whose size from few GB to 100 GB or TB.

It may be implemented on traditional main frame comp. servers or parallel architecture platform.

It requires extensive business modeling and may take many years to design & build.

② Data Mart : It contains a subset of corporate wide data that is valued to specified group of user. The scope is confined to specific selected subject, example:

A marketing DM may be limited subject wise to its customer, item and sales.

The Data contain in DM is summarized DM are implemented on low cost Departments all server that are units Linux / windows.

Implementation cycle of DM take 8-10 weeks depending upon the source of Data. DM can be categorized as independent or dependent.

Independent : Independent DM are sourced from Data captured from 1 or more operational systems or external

information providers or from data generators locally within a particular department or geographical area.

They have clear dependent DM sourced directly from enterprises DM house.

- ⑥ Virtual warehouse: It is a set of views over operational DB system for efficient query only sum of the possible summary remaining user need. It is easy to build but require excess capacity on ~~operational~~ operational DB server.