

Data Warehousing Concepts

- A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing.
- It usually contains historical data derived from transaction data,
- but it can include data from other sources.

Data Warehousing Concepts

- It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources.
- In addition to a relational database, a data warehouse environment includes

Data Warehousing Concepts

- an extraction, transportation, transformation, and loading (ETL) solution,
- an online analytical processing (OLAP) engine,
- client analysis tools, and other applications that manage the process of gathering data and delivering it to business users.

Data Warehousing Concepts

- A common way of introducing data warehousing is to refer to the characteristics of a data warehouse as follows
- Subject Oriented
- Integrated
- Nonvolatile
- Time Variant

Subject Oriented

- Data warehouses are designed to help you analyze data.
- For example, to learn more about your company's sales data,
- you can build a warehouse that concentrates on sales.

Subject Oriented

- Using this warehouse, you can answer questions like
- "Who was our best customer for this item last year?"
- This ability to define a data warehouse by subject matter, sales in this case, makes the data warehouse subject oriented.



Integrated

- Integration is closely related to subject orientation.
- Data warehouses must put data from disparate sources into a consistent format.
- They must resolve such problems as naming conflicts and inconsistencies among units of measure.
- When they achieve this, they are said to be integrated.



Nonvolatile

- Nonvolatile means that, once entered into the warehouse, data should not change.
- This is logical because the purpose of a warehouse is to enable you to analyze what has occurred.



Time Variant

- In order to discover trends in business, analysts need large amounts of data.
- This is very much in contrast to online transaction processing (OLTP) systems, where performance requirements demand that historical data be moved to an archive.
- A data warehouse's focus on change over time is what is meant by the term time variant.



Data Warehouse Architectures

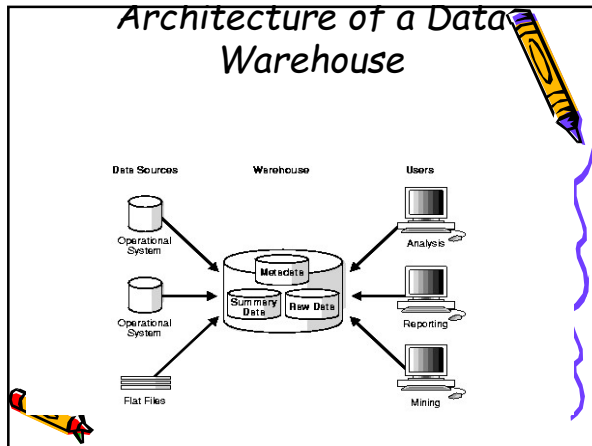
- Three common architectures are:
- Data Warehouse Architecture (Basic)
- Data Warehouse Architecture (with a Staging Area)
- Data Warehouse Architecture (with a Staging Area and Data Marts)



Data Warehouse Architecture (Basic)

- Figure below shows a simple architecture for a data warehouse.
- End users directly access data derived from several source systems through the data warehouse.





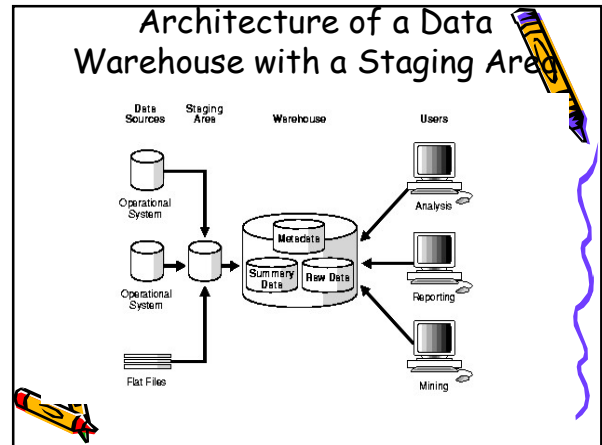
Data Warehouse Architecture (Basic)

- In Figure above, the metadata and raw data of a traditional OLTP system is present,
- as is an additional type of data, summary data.
- Summaries are very valuable in data warehouses because they pre-compute long operations in advance.

with a Staging Area

- In basic architecture, you need to clean and process your operational data before putting it into the warehouse.
- You can do this programmatically, although most data warehouses use a staging area instead.
- A staging area simplifies building summaries and general warehouse management.

Figure below illustrates this typical architecture.

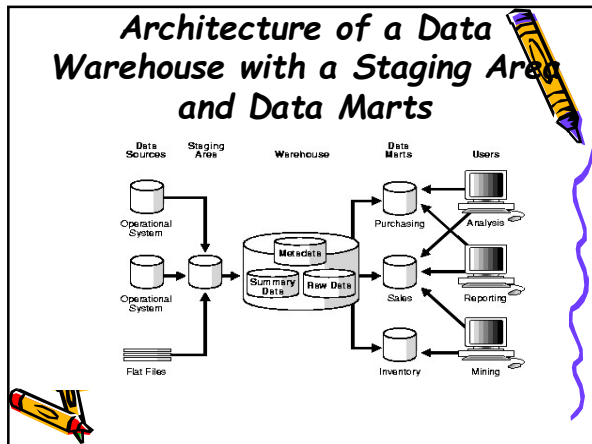


with a Staging Area and Data Marts

- Although the architecture in Figure above is quite common,
- you may want to customize your warehouse's architecture for different groups within your organization.
- You can do this by adding **data marts**,

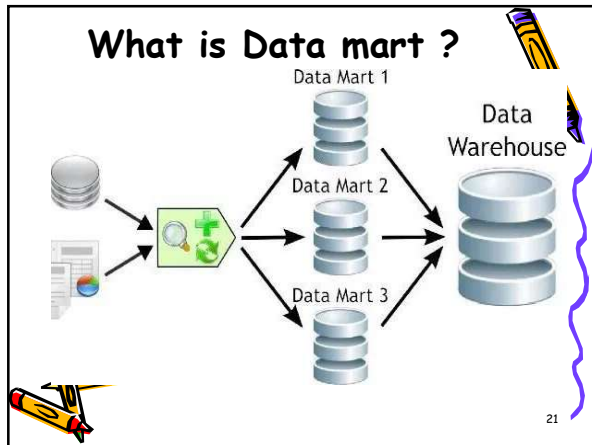
with a Staging Area and Data Marts

- which are systems designed for a particular line of business.
- Figure below illustrates an example where purchasing, sales, and inventories are separated.
- In this example, a financial analyst might want to analyze historical data for purchases and sales.



What is Data mart ?

- Data mart is a subset of an enterprise Data Warehouse and it is a subject oriented database which supports the business needs of department specific to users (middle level management).
- It represents a single subject.



Data mart types.

- **Dependent Data mart** : In a top-down approach, a Data mart development depends on enterprise DWH. Such data marts are known as dependent data marts
- **Independent Data mart** : In a Bottom-Up approach a data mart development is independent of Enterprise Data Warehouse (EDWH), such data marts are known as Independent Data marts.

Data mart types



- Independent data marts are marts that are fed directly by external sources and do not use the Data Warehouse.
- **Embedded Data marts** : Embedded Data marts are the marts that are stored within the central DW.

Data mart types

- What is To-Down DWH approach?
- According to W.H Inmon, first we need to design an enterprise specific database known as Data Warehouse from the Enterprise Data Warehouse (EDWH) department specific data base called "Data mart".

Data mart types



- What is Bottom-Up approach?
- According to Kimball, first design department specific databases known as Data marts, Integrated data marts into an Enterprise data warehouse.

25

Data Mart features.



- Low cost.
- Contain less information than the warehouse.
- Easily understood and navigated than an enterprise data warehouse.
- Within the range of divisional or departmental budgets.

26

Data Mart advantages.



- Typically single subject area and fewer dimensions.
- Focused user needs.
- Limited scope.
- Optimum model for Data Warehouse construction.
- Very quick time to market.

27

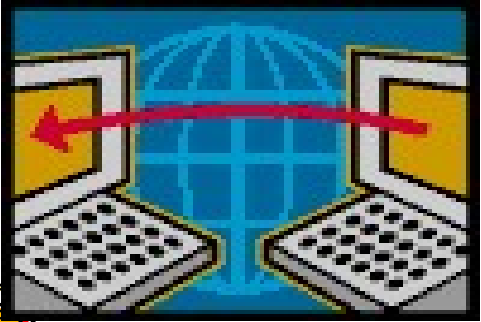


Data Mart disadvantages.

- Does not provide integrated view of business information.
- More number of data marts are complex to maintain.
- Scalability issues for large number of users and increased data volume.

28

Questions

29