

Data warehousing

A data warehousing separate from DBMS it stores huge amount of data which is collected from multiple heterogeneous data sources such as file ,dbms etc.The aim of data warehousing is to produce statistical results for good decision making . An ordinary database can stores data in mb to gb. So there management is easily done by DBMS as well as RDBMS. But this traditional database system doesn't offer itself to analyze the data for statistical results to overcome this problem upon a database management system we need an effective platform to analyze data through of central data storage place to closely study about the business or data warehouse comes in existence .

These are the following advantages of data warehousing.

1. A data warehouse is based on analytic processing
2. A data warehouse maintains historical data with data in present time, the historical data , if the data kept over years and uses to analyze present trends of business , future prediction and dynamic decision support .
3. It is integrated generally at the organization level by combining the data from the different sources. For ex- A data warehouse of school result decide the best school in a city , state country and across the worldwide.
4. The redundancy data from central warehouse reduces at atomic level and leaved by clustering tools .

Disadvantage of data warehousing

1. It requires huge space which is expensive.
2. A expert professional requires to manage a data warehouse.The expense of professionals generally not bearable.
3. A warehouse sub divided into partitions which requires individual users to manage them.
4. The implementation of data warehouse needs extra hardware and network components which make a warehouse very expensive ,not approachable for small industries.

Q. Difference between database and data warehousing



KEY DIFFERENCE

Database is a collection of related data that represents some elements of the real world whereas Data warehouse is an information system that stores historical and commutative data from single or multiple sources.

Database is designed to record data whereas the Data warehouse is designed to analyze data.

Database is application-oriented-collection of data whereas Data Warehouse is the subject-oriented collection of data.

Database uses Online Transactional Processing (OLTP) whereas Data warehouse uses Online Analytical Processing (OLAP).

Database tables and joins are complicated because they are normalized whereas Data Warehouse tables and joins are easy because they are denormalized.

ER modeling techniques are used for designing Database whereas data modeling techniques are used for designing Data Warehouse.



Parameter	Database	Data Warehouse
Purpose	Is designed to record	Is designed to analyze
Processing Method	The database uses the Online Transactional Processing (OLTP)	Data warehouse uses Online Analytical Processing (OLAP).
Usage	The database helps to perform fundamental operations for your business. ER modeling techniques are used for designing.	Data warehouse allows you to analyze your business. Data modeling techniques are used for designing.
Tables and Joins	Tables and joins of a database are complex as they are normalized.	Table and joins are simple in a data warehouse because they are denormalized.
Orientation	Is an application-oriented collection of data	It is a subject-oriented collection of data

Storage limit	Generally limited to a single application	Stores data from any number of applications
Availability	Data is available real-time	Data is refreshed from source systems as and when needed
Technique	Capture data	Analyze data
Data Type	Data stored in the Database is up to date.	Current and Historical Data is stored in Data Warehouse. May not be up to date.
Storage of data	Flat Relational Approach method is used for data storage.	Data Ware House uses dimensional and normalized approach for the data structure. Example: Star and snowflake schema.
Query Type	Simple transaction queries are used.	Complex queries are used for analysis purpose.
Data Summary	Detailed Data is stored in a database.	It stores highly summarized data.

Data mining

The word mining is the process of extraction of some valuable information in data warehouse known as data mining. The data mining extract the useful information from a bulk data or warehouse by applying some statistical tools. In case of data mining the extraction of data is in the form of data pattern. The data pattern combines together by follows some grammatical rules which absolutely return knowledge at the end of the extraction process. So we can also say that data mining is known as knowledge discovery or knowledge extraction.

The term knowledge discovery in database first of all proposed by "Shapiro" in 1989 . Today this term became most popular in the business and so many communities such as research , medical and healthcare etc. Nowadays data mining is used in almost extract data from large storage place for a purpose of decision making.

Since data mining is a process of data extraction, so as a whole process we need to follow the following three steps.

1. Data preprocessing (Cleaning , integration, selection , transformation)
2. Data extraction (statistical tools)
3. Data evaluation and presentation.