

Name: # Asad Amin

ID: # 6962

Paper: # Data warehousing

Department: # BS (CS) 6th Semester

Date: # 29/06/2020

Q no 1: Differentiate between OLTP vs OLAP?

Ans: OLTP:

Online transaction processing provides transaction-oriented application in a 3- tier architecture. OLTP administers day to day transaction of an organization.

Example :

Uses of OLTP are as follow:

- 1) ATM center is an OLTP application.
- 2) OLTP handles the ACID properties during data transaction via the application.

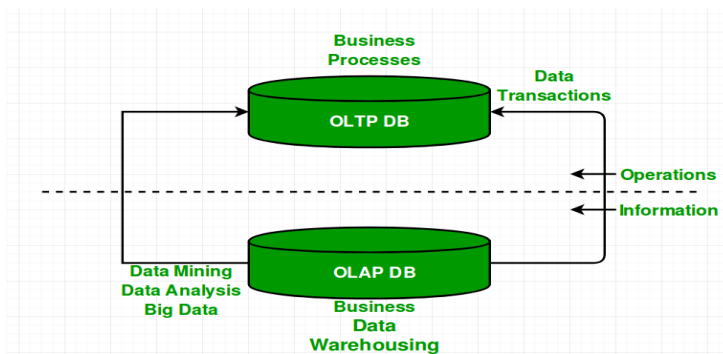
OLAP:

Online Analytical processing consists of a type of software tool that are used for data analysis for business decisions. OLAP provides an environment to get insights from database retrieved from multiple database system at one time.

Examples:

Any type of Data warehouse system in an OLAP system. Uses of OLAP are as follows:

- 1) Spottily analyzed songs by users to come up with the personalized homepage of their songs and playlist.
- 2) Netflix movie recommendation system.



Q No 2: Differentiate between Expert system and DSS?

Ans: Expert System:

1. ES is based on simple rule-based logic.
2. Problem is completely defines.
3. ES represent precisely what is needed, the extraction of expertise from those who know and making that knowledge available to those who don't know, with very positive additional connotations of top- down technology transfer within organizations.

Example:

Locating critical areas for non-point leakage of nitrogen and phosphorus. The principle is to use GIS and expert systems to integrate landscape concept which consider hydrological process into account.

Decision Support System (DSS):

- 1) The problem is open-ended.
- 2) The evolution required to solve it is also incompletely defined/ill-defined problems. Its characteristic: the solution involving a mixture of methods and dependent on the perspective of the user.
- 3) One way of method: Multi Criteria Evolution in IDRISI Andes Software.

Example:

Use MCE in IDRISI Andes to determine the best site location for village.

Q No 3: What is the relation between data mining and data warehousing?

Ans: Following the relation between Data warehousing and Data Mining.

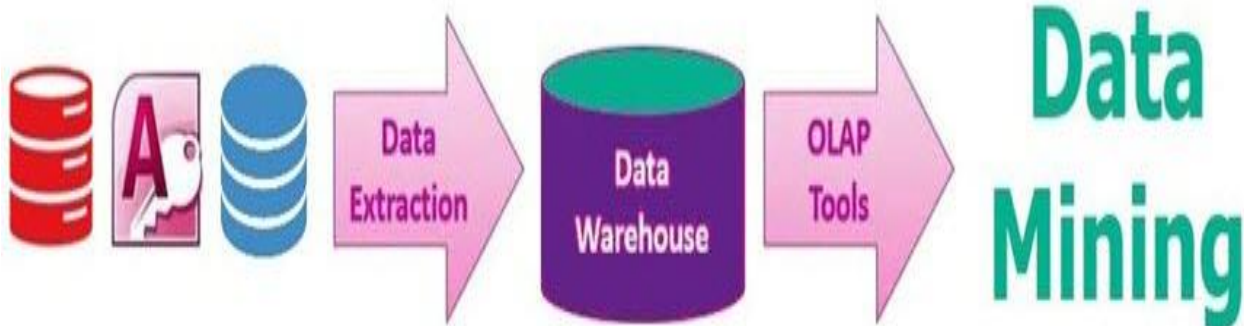
Data warehousing:

A Data warehouse is an environment where essential data from multiple sources is stored under a single schema. It is then used for reporting and analysis. 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. While a data warehouse is built to support management functions.

Data Mining:

Data mining is used to extract useful information and patterns from data. The data mining can be carried with any traditional database, but since a data warehouse contains quality data, it is good to have mining over the data warehouse system. Data mining supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction.

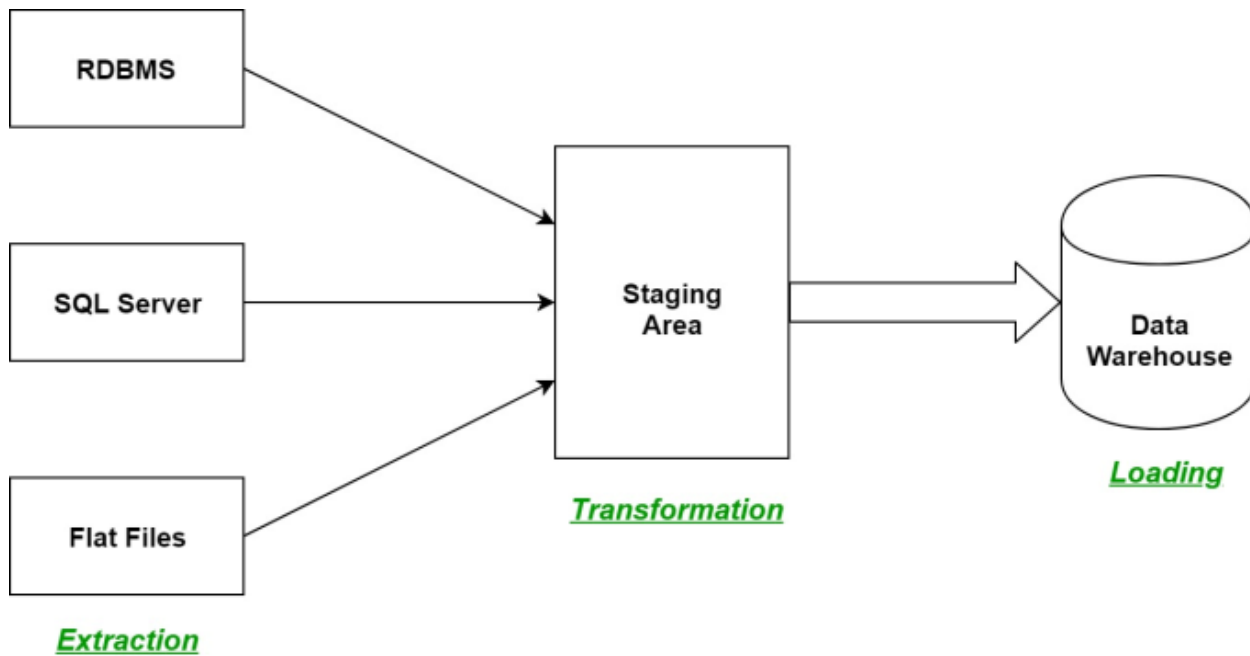
Data Warehousing & Data mining



Q No 4: Explain ETL process?

Ans: ETL process in Data Warehouse:

ETL is a process in Data warehousing and it stands for Extract, Transform and Load. It is a process in which an ETL tool extracts the data from various data source system, transforms it in the staging area and then finally, Loads it into the Data Warehouse system.



Extraction:

The first step of the ETL process is extraction. In this step, data from various source systems is extracted which can be in various formats like relational database, No SQL, XML and flat files into the staging area. It is important to extract the data from various source systems and store it into the staging area first and not directly into the data warehouse because the extracted data is in various formats and can be corrupted also. Therefore, this is one of the most important steps of ETL process.

Transformation:

The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involve following processes/tasks:

Filtering: Loading only certain attributes into the data warehouse.

Joining: Joining multiple attributes into one.

Sorting: Sorting tuples on the basis of some attributes (generally key-attribute).

Loading:

The third and final step of the ETL process is loading. In this step, the transformed data is finally loaded into the data warehouse. Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals. The rate and period of loading solely depend on the requirements and various from system to

Finally ETL process can also use the pipelining concepts i.e. as soon as some data is extracted, it can transformed and during that period some new data can be extracted. The block diagram of the pipelining of ETL process is shown below:

