

**NAME HASNAAT AMIN**

**ID 6983**

**UBJECT DATA WAREHOUSE**

**SEMESTER 6<sup>th</sup>**

## **Q1. Differentiate between OLTP vs OLAP.**

Ans:

OLAP stands for On-Line Analytical Processing. It is used for analysis of database information from multiple database systems at one time such as sales analysis and forecasting, market research, budgeting and etc. Data Warehouse is the example of OLAP system.

OLTP stands for On-Line Transactional processing. It is used for maintaining the online transaction and record integrity in multiple access environments. OLTP is a system that manages very large number of short online transactions for example, ATM.

## **Q2: Differentiate between Expert system and DSS.**

**Ans:**

- **expert systems** emulate the knowledge of an expert in a narrow field/domain
- AI: Artificial Intelligence; the name given to a broad field where computer engineers and scientists try to mimic patterns of thinking and learning using computers

- knowledge base: a collection of facts and the relationships between them; much of the content may be a collection of if/then rules
- inference engine: software that combines data input by the user with a knowledge base to try to suggest a solution
- neural networks: hardware or software designed to mimic the way a brain works
- Turing Test: A test proposed by Alan Turing to determine if computers can think. To pass the Turing Test, a computer must have a dialog with humans and have the humans not be able to tell if they are talking to a computer or another person. To make the test a little more achievable, the dialog is almost always carried out as a text connection, like instant messaging or chat.
- intelligent agent: software designed to wait to perform particular operations when triggered by a specific event, such as automatically reordering an item when the stock level of that item falls below a certain value
- examples of ES given in text:
  - medical diagnosis: used to recognize patterns of diseases based on test results
  - medical management: used to suggest courses of action and help avoid bad interactions with medicines, procedures, etc.
  - credit evaluation
  - detection of insider securities trading
  - detection of common metals
  - irrigation and pest management
  - diagnosis and prediction of mechanical failure
  -
- **DSS**: an IS designed to help managers select one of many alternative solutions
- Components of a DSS
  - data management module: a database or data warehouse that holds and maintains data for the DSS; data may come

from a number of sources including such systems as SCM (supply chain management) and CRM (customer relationship management)

- model management module: contains a model or models to be used by the DSS; the model may be fixed (static), dynamic (able to change due to changes in the data), or it may be a collection of possible models from which the DSS or the user may select
- dialog module: this is the interface between the user and the DSS; this is what the user would interact with to enter data, query the system, produce reports, etc.
- sensitivity analysis module: this is used to determine what effect particular parameters have on the result; for example, you may be doing an analysis for a municipality where the amount of tax revenue generated is given great weight toward the outcome
- examples of DSS given in text:
  - production and retailing: used to project purchasing trends and help decide how much product to stock and where to purchase it
  - tax planning: used to make financial decisions to help reduce tax burden
  - web site planning and adjustment: used to analyze customer behavior and suggest changes to design
  - yield management: used to maximize overall revenue, often by using price discrimination
  - financial services: used to make decisions such as loan approvals
  - benefit analysis: used to help determine which package of benefits is best suited for someone's needs and budget

### **Q3:What is the relation between datamining and data warehousing?**

#### **Ans:**

Major challenge to exploit data mining is identifying suitable data to mine.

Data mining requires single, separate, clean, integrated, and self-consistent source of data.

A data warehouse is well equipped for providing data for mining.

Data quality and consistency is a pre-requisite for mining to ensure the accuracy of the predictive models. Data warehouses are populated with clean, consistent data.

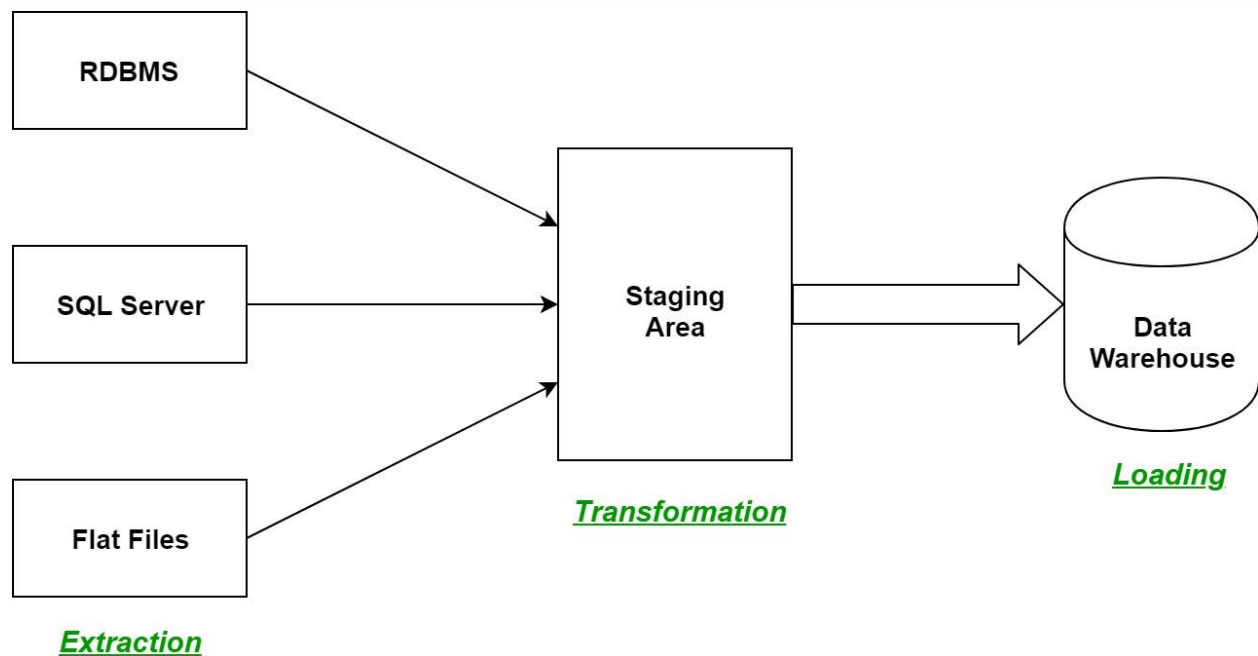
It is advantageous to mine data from multiple sources to discover as many interrelationships as possible. Data warehouses contain data from a number of sources.

Selecting the relevant subsets of records and fields for data mining requires the query capabilities of the data warehouse.

The results of a data mining study are useful if there is some way to further investigate the uncovered patterns. Data warehouses provide the capability to go back to the data source.

#### **Q4: Explain ETL process**

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 systems, transforms it in the staging area and then finally, loads it into the Data Warehouse system.



Let us understand each step of the ETL process in depth:

## 1. **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 databases, 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. Hence loading it directly into the data warehouse may damage it and rollback will be much more difficult. Therefore, this is one of the most important steps of ETL process.

## 2. **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.
- Cleaning – filling up the NULL values with some default values, mapping U.S.A, United States and America into USA, etc.
- Joining – joining multiple attributes into one.

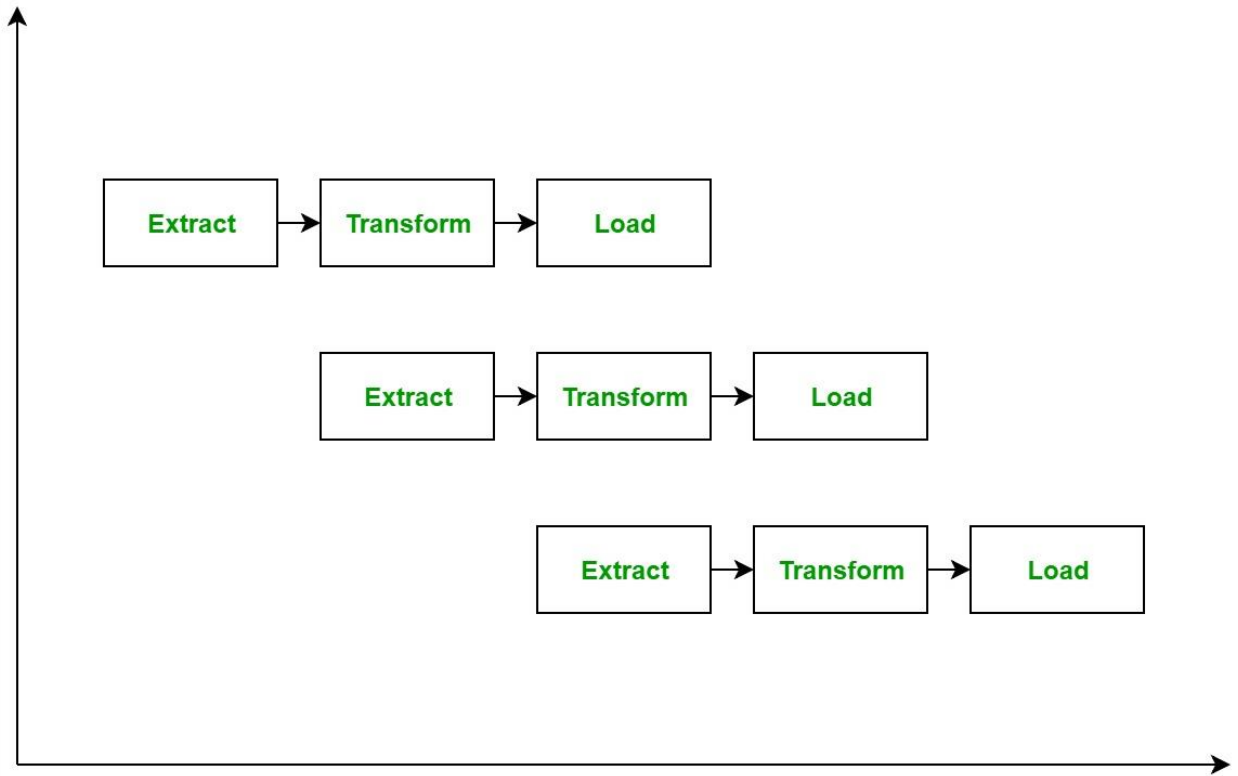
### **3. Loading:**

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 depends on the requirements and varies from system to system.

ETL process can also use the pipelining concept i.e. as soon as some data is extracted, it can be transformed and during that period some new data can be extracted. And while the transformed data is being loaded into the data warehouse, the already extracted data can be transformed. The block diagram of the pipelining of ETL process is shown below:





**ETL Tools:** Most commonly used ETL tools are Sybase, Oracle Warehouse builder, CloverETL and MarkLogic