

NAME AHMAD ULLAH KHAN

ID 6958

SEMESTER 6th

Q1. Differentiate between OLTP vs OLAP.

Ans:

OLTP and OLAP: The two terms look similar but refer to different kinds of systems. Online transaction processing (OLTP) captures, stores, and processes data from transactions in real time. Online analytical processing (OLAP) uses complex queries to analyze aggregated historical data from OLTP systems

An OLTP system captures and maintains transaction data in a database. Each transaction involves individual database records made up of multiple fields or columns. Examples include banking and credit card activity or retail checkout scanning.

In OLTP, the emphasis is on fast processing, because OLTP databases are read, written, and updated frequently. If a transaction fails, built-in system logic ensures data integrity.

OLAP applies complex queries to large amounts of historical data, aggregated from OLTP databases and other sources, for data mining, analytics, and **business intelligence** projects. In OLAP, the emphasis is on response time to these complex queries. Each query involves one or more columns of data aggregated from many rows. Examples include year-over-year financial performance or marketing lead generation

trends. OLAP databases and **data warehouses** give analysts and decision-makers the ability to use custom reporting tools to turn data into information. Query failure in OLAP does not interrupt or delay transaction processing for customers, but it can delay or impact the accuracy of business intelligence insights

Q2: Differentiate between Expert system and DSS.

Ans:

EXPERT SYSTEM is an interactive and reliable computer-based decision-making system which uses both facts and heuristics to solve complex decision-making problems. It is considered at the highest level of human intelligence and expertise. The purpose of an expert system is to solve the most complex issues in a specific domain.

The Expert System can resolve many issues which generally would require a human expert. It is based on knowledge acquired from an expert. It is also capable of expressing and reasoning about some domain of knowledge. Expert systems were the predecessor of the current day artificial intelligence, deep learning and machine learning systems.

A **decision support system (DSS)** is an information system that supports business or organizational decision-making activities. DSSs serve the management, operations and planning levels of an organization (usually mid and higher management) and help people make decisions about problems that may be rapidly changing and not easily specified in advance—i.e. unstructured and semi-structured decision problems. Decision support systems can be either fully computerized or human-powered, or a combination of both.

DSSs include knowledge-based systems. A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from a combination of raw data, documents, and personal knowledge, or business models to identify and solve problems and make decisions

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

Ans:

Data mining is the process of exploring data to find the patterns and relationships that describe the data and to predict the unknown or future values of the data. The key value of data mining is the ability to

understand why some things happened in the past and the ability to predict what will happen in the future.

A **data warehouse** is a system that retrieves and consolidates data periodically from the source systems into a dimensional or normalized data store. It usually keeps years of history and is queried for business intelligence or other analytical activities.

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.

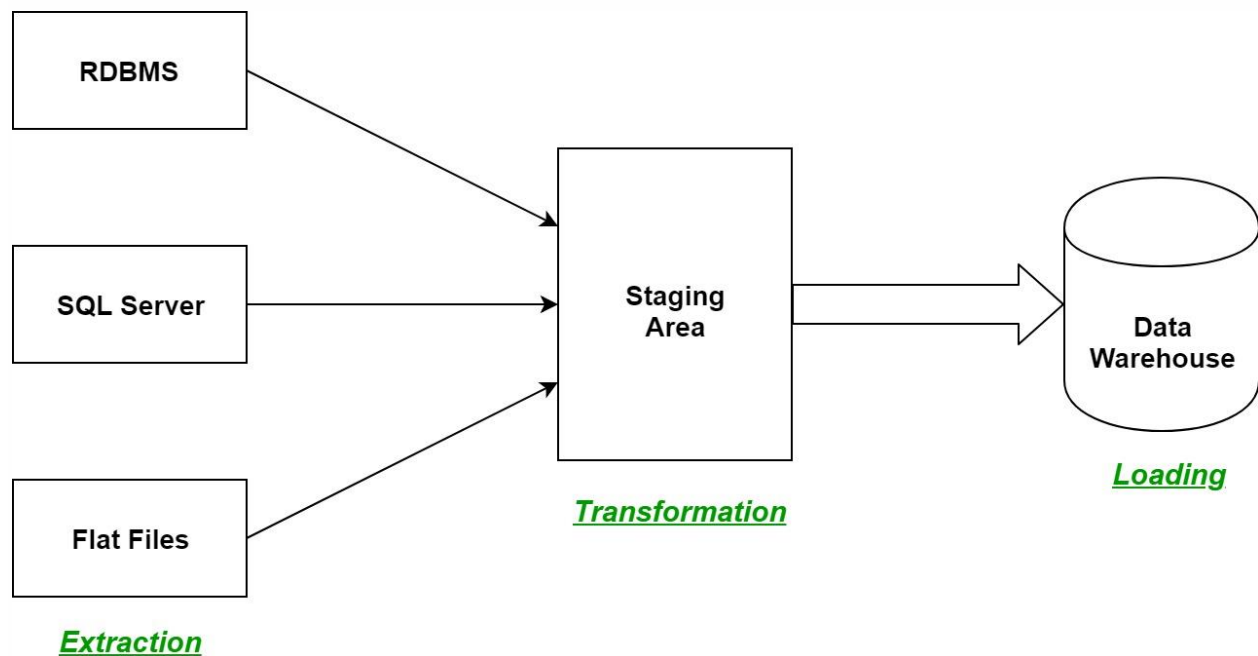
Data quality and consistency is a pre-requisite for mining to ensure the accuracy of the predictive models. Data warehouses are populated with clean, consistent data.

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.

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

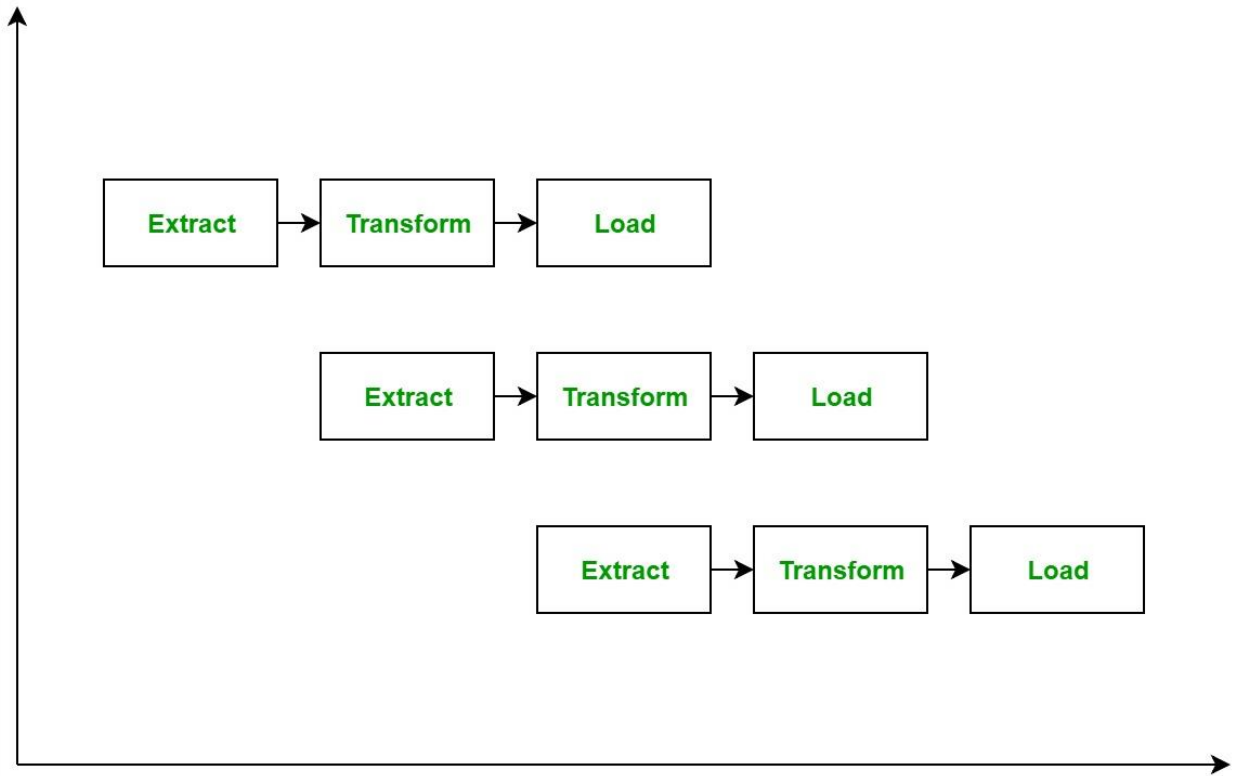
- Filtering – loading only certain attributes into the data warehouse.
- Cleaning – filling up the NULL values with some default values, mapping U.S.A, United States and America into USA, etc.
- Joining – joining multiple attributes into one.
- Splitting – splitting a single attribute into multiple attributes.
- Sorting – sorting tuples on the basis of some attribute (generally key-attribute).

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