NAME   MEHBOOB SHABBIR

ID   13916

SUBMITTED TO   SIR ZAIN

SUBJECT   DATA WAREHOUSING

- Plagiarized assignment will not be acceptable.

- Make a Proper Word Document/ PDF or PowerPoint Presentation of this assignment, Picture will not be acceptable.

- Must submit before deadline

Q1: Differentiate between OLTP vs OLAP

We can divide IT systems into transactional (OLTP) and analytical (OLAP). In general, we can assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

- **OLTP (On-line Transaction Processing)**

 . Is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE).

 . The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second.

. In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model

- **OLAP (On-line Analytical Processing)**

. Is characterized by relatively low volume of transactions.

. Queries are often very complex and involve aggregations.

. For OLAP systems a response time is an effectiveness measure.

. OLAP applications are widely used by Data Mining techniques.

. In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema).

. For example, a bank storing years of historical records of check deposits could use an OLAP database to provide reporting to business users.

Q2: Differentiate between Expert system and DSS

DSS

. A DSS is an interactive system that helps decision-makers utilize data and models to solve unstructured or semi-structured problems.

.it is used for facilitates decision making

.it is use for unstructured environment

. alternatives may not be fully realized yet

.it used to extract or gain knowledge from a computer system

.it is use goal and the system data to established alternatives and outcomes, so a good decision can be made

EXPER SYSTEM

A specialized problem domain that is considered difficult and requires specialized. An ES is a problem-solving computer program that achieves good performance in zed knowledge and skill

. it is Automate decision making

. the decision environment have structure

. the alternatives and goals are often established in advance

. inject expert knowledge in to a computer system

. the expert system can eventually replace the human decision maker

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

Ans: While closely related, both concepts have their own specific roles.

Data mining is the automated process of analyzing large data sets to find these patterns, relationships and trends and ultimately to generate business insights – which will be used to solve challenges and identify new opportunities, so organizations can use past patterns to predict future behaviors and results.

A data warehouse – where the data from the various sources is combined and stored – allows data mining to be used throughout the organization, from sales and marketing applications to research, product development and finance. Data warehousing and data mining are the cornerstones of modern business decisions: How Your Data Warehouse Can Make Data Mining Easier and More Efficient

Data mining techniques can be carried with any traditional database, but because a data warehouse contains quality data that has already been sanitized and tested, it makes sense to have data mining over a data warehouse system.

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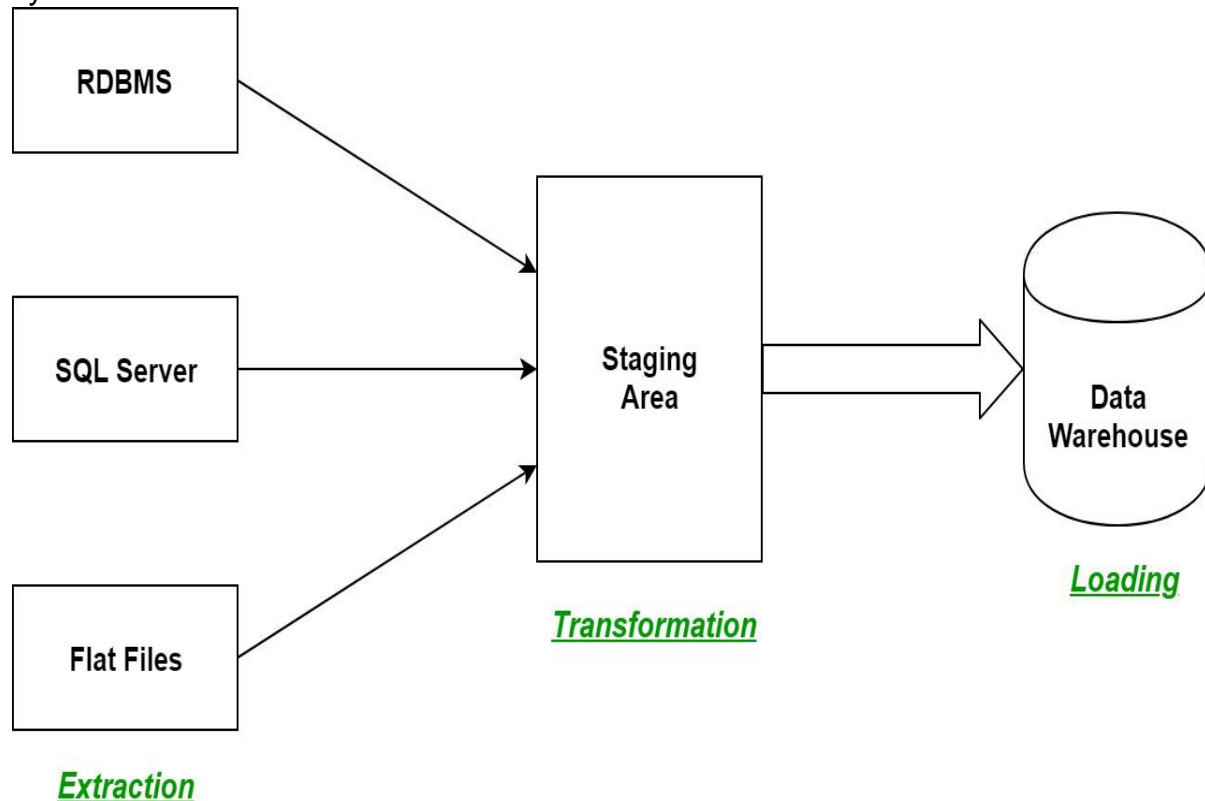
Let us understand each step of the ETL process in depth:

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 transform 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:
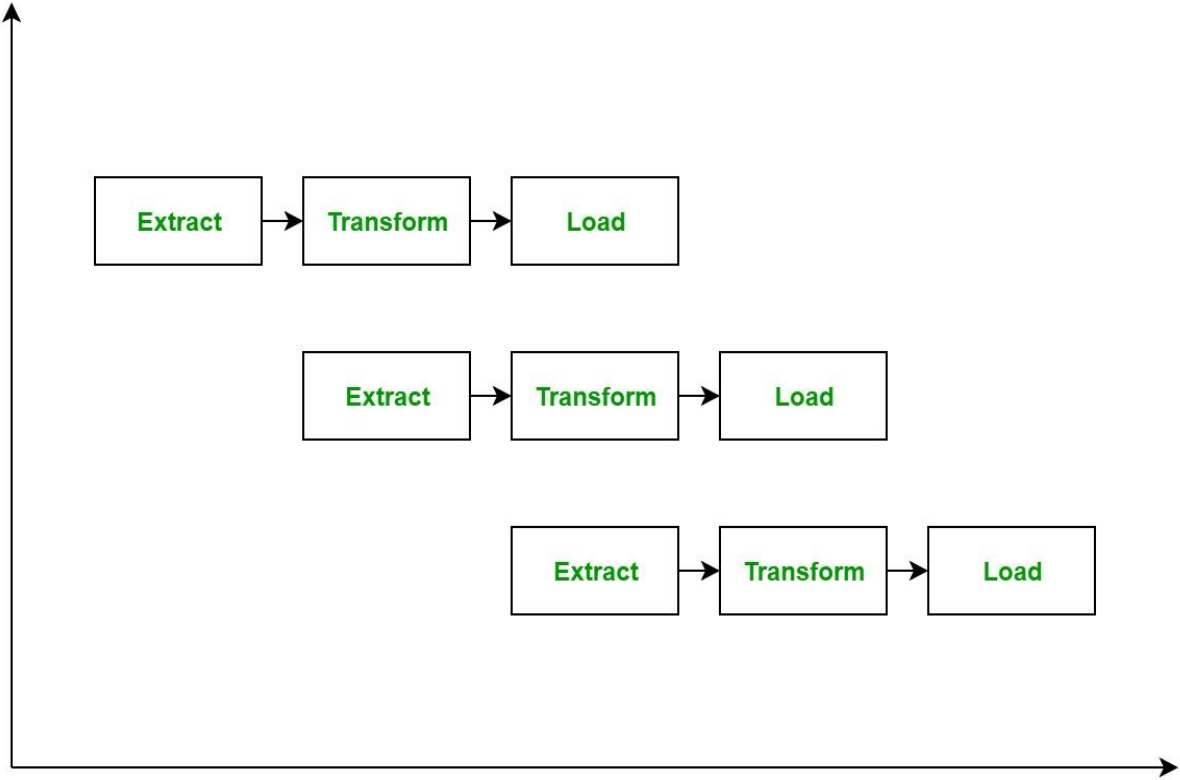
ETL Tools: Most commonly used ETL tools are Sybase, Oracle Warehouse builder, Clover ETL and Mark Logic.