



NAME: YASIR FAHEEM

ID # 6991

BS COMPUTER SCIENCE

DATA WARE HOUSING

FINAL EXAM

DATE: 29/06/20

Q NO:1 Differentiate between OLTP vs OLAP.

Ans:

OLAP

OLAP stands for **Online Analytical Processing**. The database here, stores the historical data and historical data processing are done. It allows users to analyze the database information from multiple databases at the same time. This analysis is then used for decision making and planning for companies and organizations.

Example 1: If any organization wants to launch any new policy then it can use past information about how the customers have responded to any similar policy. This information will help them to make their policy more user-friendly and successful among their customers.

OLTP

OLTP stands for **Online Transaction Processing**. This system captures and maintains transaction in the database. This mainly focuses on the **fast processing of the database** as the read and write operations are performed here frequently.

Example 1: The ATM machines are an OLTP application. The online transactions that we perform are stored and updated here.

Difference

1. **Characteristics:** OLTP handles small transaction systems whereas the OLAP is a historical data retrieval and processing system.
2. **Users:** The OLTP systems are mainly used by hundreds of users like clerks, Database Admin, online shoppers, etc, whereas OLAP systems are used by data analysts, business analysts, and executives.
3. **Functions:** The OLTP systems provide day to day operations for organizations. These operations are simple and repetitive. The bank clients mainly perform operations like money deposit and withdrawal on a

repetitive basis. OLAP systems help organizations in making decisions, planning, and problem-solving for them.

4. **Operations allowed:** OLTP allows both read and write operations as the user performs insert, update and delete tasks. But, OLAP allows only read operations and rarely write operation as the main task here is to extract the data for analysis.
5. **Source:** Transactions are the source of OLTP systems. OLTP databases become the source for OLAP systems.
6. **Amount of Data:** OLAP systems stores a large amount of data usually in TB and PB (PetaByte) whereas OLTP systems stores relatively less amount of data usually in MB and GB.
7. **Backup:** OLTP systems backups the complete data regularly. OLAP needs to back up the data from time to time and is not as important as backups in OLTP systems.
8. **Response Time:** OLTP systems respond within milliseconds whereas OLAP systems may take from seconds to minutes to respond depending on the amount of data being processed.
9. **Tables:** Normalized tables(3NF) are used in OLTP whereas the tables may not be normalized in OLAP.

Q NO 2: Differentiate between Expert system and DSS.

Ans:

A decision support system (DSS) is a computerized program used to support determinations, judgments, and courses of action in an organization or a business. A DSS sifts through and analyzes massive amounts of data, compiling comprehensive information that can be used to solve problems and in decision-making.

Typical information used by a DSS includes target or projected revenue, sales figures or past ones from different time periods, and other inventory- or operations-related data.

KEY TAKEAWAYS

- A decision support system (DSS) is a computerized system that gathers and analyzes data, synthesizing it to produce comprehensive information reports.
- A decision support system differs from an ordinary operations application, whose function is just to collect data.
- Decision support systems allow for more informed decision-making, timely problem-solving, and improved efficiency in dealing with issues or operations, planning, and even management.

while

Expert systems (ES)

Artificial Intelligence is a piece of software that simulates the behavior and judgment of a human or an organization that has experts in a particular domain is known as an expert system. It does by acquiring relevant knowledge from its knowledge base and interpreting it according to the user's problem. The data in the knowledge base is added by humans that are expert in a particular domain and this software is used by a non-expert user to acquire some information. It is widely used in many areas such as medical diagnosis, accounting, coding, games etc. An expert system is an AI software that uses knowledge stored in a knowledge base to solve problems that would usually require a human expert thus preserving a human expert's knowledge in its knowledge base. They can advise users as well as provide explanations to them about how they reached a particular conclusion or advice.

Q NO 3: 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.

A data warehouse is a description for specific server and storage capacities, mostly used to store big and/or unstructured data. The idea is that data is stored in a easy to find and easy to extract way - like goods in the shelf's of a warehouse.

Data mining is a collection of techniques to find meaningful relationships in large data sets. These relationships, such as correlations, can be used for diverse means.

So, you'd first store data in a data warehouse (or somebody does that for you). Now this data can be downloaded and used for data mining.

Starting by loading data into data warehouse, (you can build your data warehouse in any DB engine, only you should respect some norms to build your star or snowflake, or constellation schema the schema (schema = Fact tables linked into Dimension tables).

Than you can extract the meaning of this data, or the information behind it by using Data mining. Data warehouse is not mandatory to do data mining; you can extract information using DM only by using Excel file, or any other data format.

The link between "Data Mining" and "Data Warehousing" is the same as the link between metal mining and gathering metal bearing ore in a place and format conducive to easy processing.

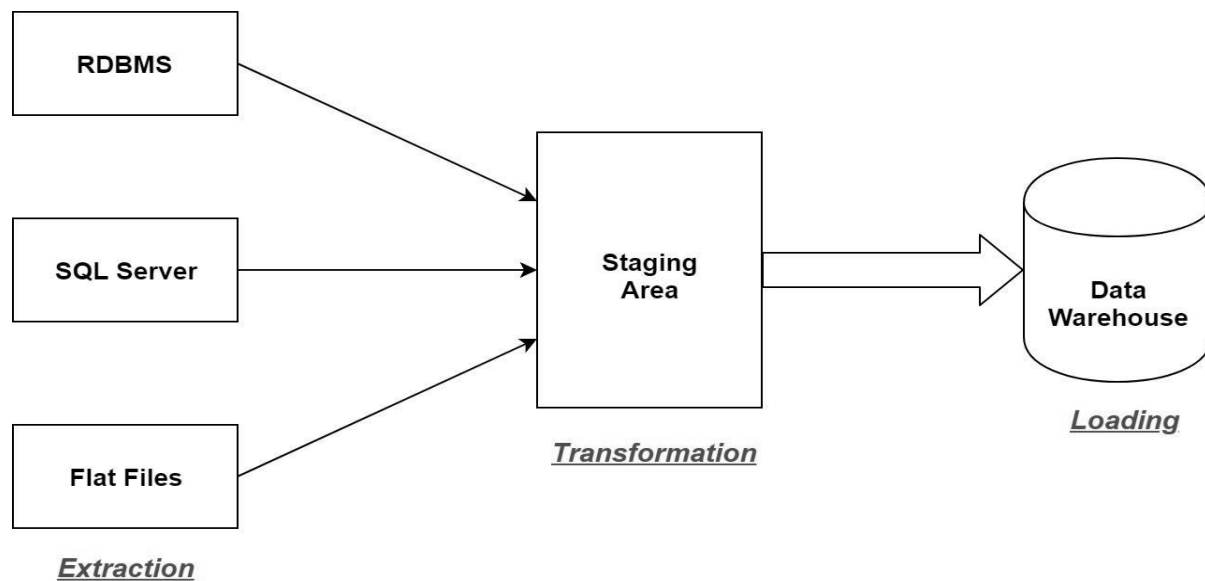
Data mining requires *data preparation*. A data warehouse (DW) supports and performs the *data preparation* processes to support the one or more data mining projects.

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 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:**

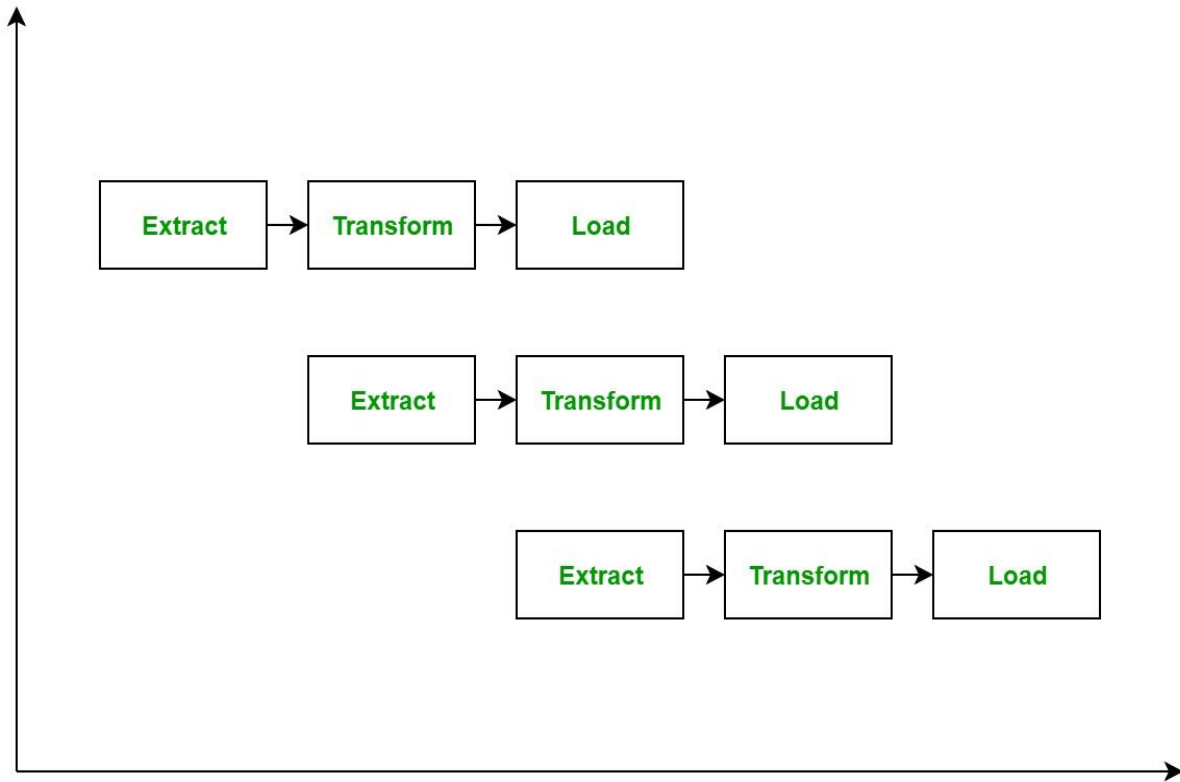
The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involve following processes/tasks:

1. Filtering – loading only certain attributes into the data warehouse.
2. Cleaning – filling up the NULL values with some default values, mapping U.S.A, United States and America into USA, etc.
3. Joining – joining multiple attributes into one.
4. Splitting – splitting a single attribute into multiple attributes.
5. 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 Mark Logic.