

ID 13698

Name Shariq

Q1. Differentiate between OLTP vs OLAP?

OLTP	OLAP
<ul style="list-style-type: none"><li>● It is an online transactional system. It manages database modification.</li></ul>	<ul style="list-style-type: none"><li>● OLAP is an online analysis and data retrieving process.</li></ul>
<ul style="list-style-type: none"><li>● It is characterized by large numbers of short online transactions.</li></ul>	<ul style="list-style-type: none"><li>● It is characterized by a large volume of data.</li></ul>
<ul style="list-style-type: none"><li>● OLTP is an online database modifying system.</li></ul>	<ul style="list-style-type: none"><li>● OLAP is an online database query management system.</li></ul>
<ul style="list-style-type: none"><li>● OLTP uses traditional DBMS.</li></ul>	<ul style="list-style-type: none"><li>● OLAP uses the data warehouse.</li></ul>
<ul style="list-style-type: none"><li>● Insert, Update, and Delete information from the database.</li></ul>	<ul style="list-style-type: none"><li>● Mostly select operations</li></ul>
<ul style="list-style-type: none"><li>● It's response time is in millisecond.</li></ul>	<ul style="list-style-type: none"><li>● Response time in seconds to minutes.</li></ul>
<ul style="list-style-type: none"><li>● OLTP database must maintain data integrity constraint.</li></ul>	<ul style="list-style-type: none"><li>● OLAP database does not get frequently modified. Hence, data integrity is not an issue.</li></ul>
<ul style="list-style-type: none"><li>● It helps to control and run fundamental business tasks.</li></ul>	<ul style="list-style-type: none"><li>● It helps with planning, problem-solving, and decision support.</li></ul>
<ul style="list-style-type: none"><li>● Allow read/write operations.</li></ul>	<ul style="list-style-type: none"><li>● Only read and rarely write.</li></ul>
<ul style="list-style-type: none"><li>● It is a market orientated process.</li></ul>	<ul style="list-style-type: none"><li>● It is a customer orientated process.</li></ul>
<ul style="list-style-type: none"><li>● Complete backup of the data combined with incremental backups.</li></ul>	<ul style="list-style-type: none"><li>● OLAP only need a backup from time to time. Backup is not important compared to OLTP</li></ul>

## Q2. Differentiate between Expert system and DSS?

Expert System	DSS
● To replicate humans	● To assist human
● Expertise transfer	● Decision-making
● Machine queries human	● Human queries machine
● Individual user	● Possible group user
● Narrow	● Complex, wide
● Problem is completely defines	● The problem is open-ended
● There is clear way for the solution method	● The evaluation required to solve it is also incompletely defined/ill-defined problems.
● An ES is a problem-solving computer program that achieves good performance in a specialized problem domain that is considered difficult <b>and</b> requires specialized knowledge <b>and</b> skill.	● A <b>DSS</b> is an interactive <b>system</b> that helps decision-makers utilize data <b>and</b> models to solve unstructured or semi-structured problems.
● An expert system is used to retrieve information about and/or answer questions relating to a specific subject area. The expert support system is assumed to contain a vast database of knowledge about the subject field.	● A decision support system is usually based on the mathematics associated with comparing/ contrasting two or more decisions and enabling the user to determine the ( <b>objectively</b> ) wisest possible decision; no subject matter knowledge is used or assumed.

**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.

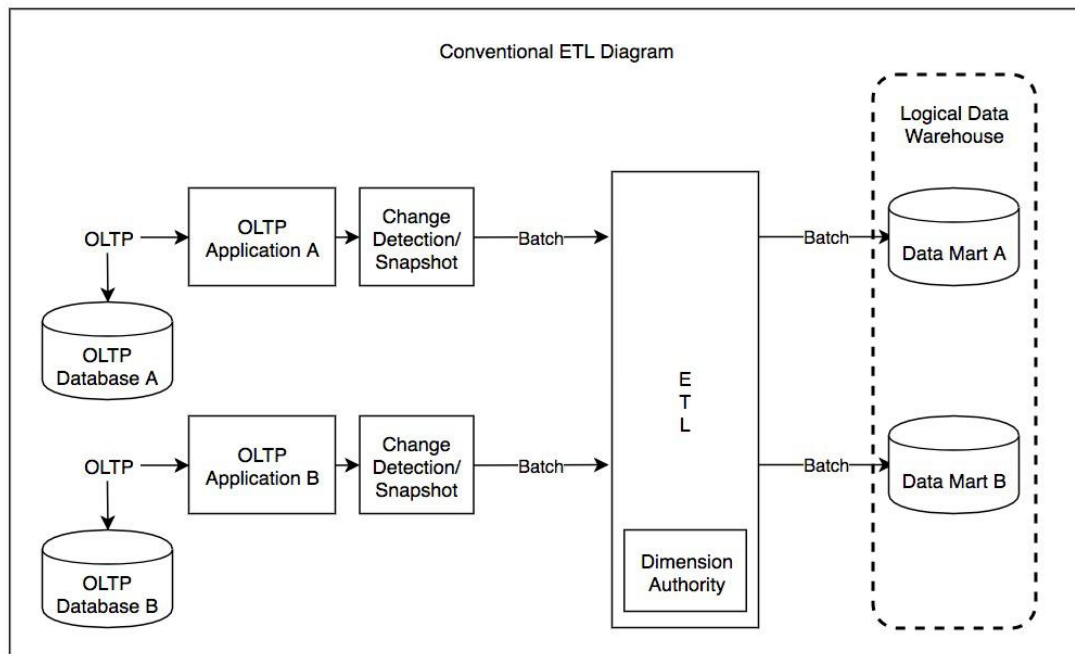
A data warehouse – where the data from the various sources is combined and stored – allows data mining to be used throughout the organization, from sales and marketing applications to research, product development and finance. Data warehousing and data mining are the cornerstones of modern business decisions: [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.****Definition:**

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.

Diagram for ETL:



**Process involved:**

Extraction

Transformation

Loading

### 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.