Name: Shahab ud din ID: 13791 datawarehousing

Questions

1. **Differentiate between OLTP vs OLAP.**

Ans: **OLTP Points:**

1.OLTP is basically a transactional processing.

2. OLTP is also known as online database modifying system.
3. OLTP focus on insert, update, delete information from the database.
4. OLTP has a short transactions.
5. The processing time of a transaction in OLTP is comparatively less then OLTP.

OLAP points:

1. Its is an online data retrieving and data analysis system.
2. The big difference between is that its an database query answering system.
3. It focuses on extract data for analyzing that helps in decision making.
4. Different OLTP's database becomes the source of data for OLAP.
5. OLAP takes longer transaction time then compare to OLTP.

2.Differentiate between Expert system and DSS.

Ans**: DSS points:**

1. A DSS is an interactive system that helps decision-makers to utilize data and models to solve different unstructured problems.
2. A DSS uses goals and the system data to establish alternatives and outcomes, so a good decision can be made.
3. DSS extract or grain knowledge from a computer system.
4. DSS is a systems that can be either fully computerized or human –powered or can be a combination of both.
5. DSS focus on the less well structured, underspecified problem that upper level mangers typically face.

ES points:

1. An ES is a problem solving computer program that achieves good performance in a specialized problem domain that is considered difficult and requires specialized knowledge and skill.
2. ES has a more structured environment for problem solving rather than DSS.

3. The expert system can eventually replace the human decision maker.
4. Rather than extracting or gain knowledge like DSS, ES Inject expert knowledge in to a computer system.
5. In ES alternatives and goals are often established in advance.

3.What is the relation between data mining and data warehousing?

Ans: both concepts have their own specific roles. But this is the basic relation between both:

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 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.

Data mining is a process of statistical analysis. Analysts use technical tools to query and sort through terabytes of data looking for patterns. Running a query on the relevant data to prove or disprove this theory is data mining. Businesses then use this information to make better business decisions based on how they understand their customers' and suppliers' behaviors.

Data warehousing describes the process of designing how the data is stored in order to improve reporting and analysis. Data warehouse experts consider that the various stores of data are connected and related to each other conceptually as well as physically. A business's data is usually stored across a number of Databases. However, to be able to analyze the broadest range of data, each of these databases needs to be connected in some way. This means that the data within them need a way of being related to other relevant data and that the physical databases themselves have a connection so their data can be looked at together for reporting purposes.

So the relationship between data mining and data warehousing is that data, properly warehoused, is easier to mine. If a data mining query has to run through terabytes of data spread across multiple databases, which sit on different physical networks - - that is not an efficient query and getting results will take a long a time.

4.Explain ETL process.

Ans: ETL is a process in data warehouse and its stands for Extact, transform and load. It is a process in which an ETL tool extracts the data from different data source systems and transforms it in the staging area and then finally loads it into the data warehouse system.

Explaining Extract, transform and load aka ETL:

1. **Extract:**

   well basically in this step data from different system is extracted which can be in different formats like relational databases, no SQL ,XML and flat files into the staging area. Its is important to extract the data from different systems and store it into the staging area first and not directly into data warehouse because the extracted data is in different formats and can be corrupted .therefore, this is one of the most important steps of ETL process.

2. **Transform:**

the second step is transform. In this step a set of rules or functions are applied on the extracted data to convert it into a single standard format. It involve the following tasks:

*Filtering *Cleaning *Joining * Splitting *Sorting

3. **LOAD:**

The final step in ETL process is LOAD. In this step the transformed data is finally loaded into the data warehouse. Sometimes that 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.
