1. What is data-warehouse?

* A data warehouse is a large collection of business data used to help an organization make decisions.
* The large amount of data in a warehouse comes from different business applications used by different parts of the organization.
* A ware house pulls data from these business application and then processes it and stores t so that the data can be made available when required.
* Benefits of data warehousing
* Easily interpretable and better data.
* It helps organization to make data driven decision which can benefit the organization.

1. Difference between OLTP vs OLAP

* **An OLTP** system captures and maintains transaction data in a database.
* **OLAP** applies queries to large amounts of historical data, aggregated from OLTP databases for data mining, analytics and business intelligence.
* Some key differences can be tabulated as below

|  |  |  |
| --- | --- | --- |
|  | **OLTP** | **OLAP** |
| **Characteristics** | Handles a large number of small transactions | Handles large volumes of data with complex queries |
| **Query types** | Simple standardized queries | Complex queries |
| **Operations** | Based on INSERT, UPDATE, DELETE commands | Based on SELECT commands to aggregate data for reporting |
| **Response time** | Milliseconds | Seconds, minutes, or hours depending on the amount of data to process |
| **Source** | Transactions | Aggregated data from transactions |
| **Purpose** | Control and run essential business operations in real time | Plan, solve problems, support decisions, discover hidden insights |
| **Space requirements** | Generally small if historical data is archived | Generally large due to aggregating large datasets |
| **Backup and recovery** | Regular backups required to ensure business continuity and meet legal and governance requirements | Lost data can be reloaded from OLTP database as needed in lieu of regular backups |
| **Database design** | Normalized databases for efficiency | Denormalized databases for analysis |

1. What are the various characteristics of data-warehouse?

* Data Warehouse is aggregated transactional data, transformed and stored for analytical purposes.
* It is optimized for aggregation and retrieval of large data sets
* Data Warehouse stores data from multiple databases which makes it easier to analyze.
* It layers on top of transactional databases to allow for analytics.
* Data Warehouse is subject oriented as it provides information about a theme instead of ongoing transactions of the organization.
* Data in a data warehouse is nonvolatile, hence data warehouses are huge in size.
* The data is maintained via different intervals of time.

1. What is Star-Schema?

* **Star Schema** in data warehouse, in which the center of the star can have one fact table and a number of associated dimension tables
* It is known as star schema as its structure resembles a star.
* The Star Schema data model is the simplest type of Data Warehouse schema
* It is also known as Star Join Schema and is optimized for querying large data sets.
* Every dimension in a star schema is represented with the only one-dimension table.
* The dimension table should contain the set of attributes.
* The dimension table are not joined to each other
* Fact table would contain key and measure
* provides optimal disk usage.

1. What do you mean by SETL?

* SETL stands for Semantic ETL(Extract, Transform, Load).
* Extract: extract data from warehouse
* Transform: transform extracted data.
* Load: load transformed data back to warehouse.
* SETL is more efficient than traditional ETL techniques in dealing with the semantic issue that normally are encountered while working with unstructured data.
* Current ETL tools neither support processing semantic data nor create a semantic Data Warehouse (DW), a repository of semantically integrated data.