Data warehousing Short Questions Past Paper 2023

i. Define Data Warehousing

Data Warehousing is a technology that collects, stores, and manages large volumes of data from various sources to support decision-making, reporting, and data analysis. Data warehouses are optimized for read access and analytical processing, allowing users to derive insights from historical data.

ii. What do you mean by Operational Database?

An Operational Database supports daily operations of an organization, storing current data used in real-time processes. It is typically transactional, supporting CRUD operations (Create, Read, Update, Delete), and focuses on day-to-day operations rather than historical data.

iii. Explain the Term Metadata

Metadata is data that describes other data, providing context, structure, and meaning. In data warehousing, metadata includes details about data sources, data transformations, structures, and processes, helping users understand data lineage and relationships.

iv. Differentiate Between OLAP and MOLAP. What do you mean by Semantic Conflict?

- **OLAP (Online Analytical Processing)** enables complex queries and multi-dimensional data analysis, often used in data warehouses.
- **MOLAP (Multidimensional OLAP)** is a type of OLAP that uses multidimensional cube structures to store and retrieve data, providing faster query responses for certain complex analyses.
- **Semantic Conflict** occurs when data definitions, terminologies, or meanings differ across data sources, leading to inconsistencies in data integration.

vi. Briefly Explain the Term Minimality

Minimality in data warehousing refers to maintaining only the essential data required for analysis, ensuring data storage efficiency and avoiding redundancy.

vii. Explain the Term HOLAP

HOLAP (Hybrid Online Analytical Processing) combines MOLAP and ROLAP (Relational OLAP) advantages, offering both high-performance multidimensional analysis and flexibility in storing large data sets.

viii. Explain the Purpose of Analysis and Reconciliation of Data Sources

The analysis and reconciliation of data sources involve evaluating, cleaning, and standardizing data from various origins to ensure consistency, accuracy, and compatibility before loading it into the data warehouse.

ix. What is the Purpose of Optional Arcs?

Optional Arcs in data modeling define mutually exclusive relationships between entities. They ensure that only one of multiple related paths is chosen, providing flexibility and logical structuring in relational designs.

x. How Can Data be Accessed in a Data Warehouse?

Data in a data warehouse can be accessed through OLAP tools, SQL queries, BI (Business Intelligence) tools, and reporting software. These tools support slicing, dicing, drilling down, and rolling up data to provide insights.

xi. What are the Risk Factors Associated with the Data Warehouse System Lifecycle?

Some **risk factors** include:

- High initial setup and maintenance costs
- Complexity in data integration
- Data quality issues
- Performance challenges with increasing data volume
- Security and privacy risks

xii. Compare and Contrast the Top-Down and Bottom-Up Approaches in Data Warehousing

- **Top-Down Approach**: Develops a centralized data warehouse first, followed by the creation of data marts. It is ideal for large organizations needing consistent, company-wide reporting.
- **Bottom-Up Approach**: Begins with data marts focused on specific business areas, eventually integrating them into a larger warehouse. It is faster to implement and better suited for smaller organizations.

xiii. What are the Key Phases in Data Mart Design, and How Are Data Marts Tested?

Key phases include:

1. Requirements gathering and data modeling
2. ETL (Extract, Transform, Load) process implementation
3. Building and deploying data marts

Testing involves validating data integrity, accuracy, and performance to ensure the data mart meets analytical needs.

xiv. How Do You Inspect and Normalize Schemata in the Analysis and Reconciliation of Data Sources?

Inspection involves identifying data anomalies and inconsistencies, while normalization reduces redundancy by organizing data into related tables and ensuring referential integrity in the data warehouse.

xv. What Are the Different Approaches to Conceptual Design in Data Warehousing?

Common approaches are:

- **Entity-relationship modeling**: Focuses on data relationships.
- **Dimensional modeling**: Structures data into facts and dimensions, suited for analytical queries.

xvi. What Factors Are Considered in Workload and Data Volume Analysis in Data Warehousing?

Key factors include:

- Query frequency and complexity
- User concurrency requirements
- Data storage capacity
- Processing power required for ETL and analytical tasks
- Future scalability for data growth