ST. ANNE'S COLLEGE OF ENGINEERING AND TECHNOLOGY
(Approved by AICTE, New Delhi. Affiliated to Anna University, Chennai)
Accredited by NAAC
ANGUCHETTYPALAYAM, PANRUTI – 607 106

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CCS341 DATA WAREHOUSING
QUESTION BANK

UNIT I INTRODUCTION TO DATA WAREHOUSE

PART - A

1. **What is a Data Warehouse?**

A data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision making process.

2. **What is the use of Data Warehouses in Organisations?**

To increasing customer focus which includes the analysis of customer buying patterns,

- To reposition products and managing product portfolios by comparing the performance of sales by quarter, by year, and by geographic regions, in order to fine-tune production strategies,
- To analysing operations and looking for sources of profit,
- To managing the customer relationships, making environmental corrections, and managing the cost of corporate assets, and
- It is also very useful from the point of view of heterogeneous database integration.

3. **Differences between Operational Database Warehouses.**

Feature	***OLTP***	***OLAP***
Characteristic	operational processing	informational transaction
User	clerk, DBA, DB professional	knowledge worker
Function	day-to-day operations	long-term informational requirements decision support
DB design	ER-based, application-oriented	star/snow flake, subject-oriented
Data	current, guaranteed up-to-date	historic, accuracy maintained over time

4. **List out the characteristics of Data Warehouse.**

- **Subject oriented:** A data warehouse is organized around major subjects. **Integrated:** A data warehouse is constructed by integrating multiple heterogeneous sources, such as relational databases, flat files, and online transaction records.
- **Time variant:** Data are stored to provide information from an historic perspective (e.g., the past 5–10 years).
- **Non-volatile:** A data warehouse does not require transaction processing, recovery, and concurrency control mechanisms.

5. **Define OLTP.**

The major task of online operational database systems is to perform online transaction and query processing. These systems are called online transaction processing (OLTP) systems. They cover most of the day-to-day operations of an organization such as purchasing, inventory, manufacturing, banking, payroll, registration, and accounting.

6. **Define OLAP.**

Data warehouse systems serve users or knowledge workers in the role of data analysis and decision making. These systems are known as online analytical processing (OLAP) systems.

7. **List out the Data Warehousing Components.**

- Data sourcing, clean-up, transformation, and migration tools
- Metadata repository
- Warehouse/database technology
    - Data marts
    - Data query, reporting, analysis, and mining tools
- Data warehouse administration and management
- Information delivery system

8. **List out the two approaches used to build a Data warehouse.**

- Top - Down Approach
- Bottom - Up Approach

9. **Define top-down approach of data warehouse.**

In the top down approach, we build a centralized repository to house corporate wide business data. This repository is called Enterprise Data Warehouse (EDW).The data in the EDW is stored in a normalized form in order to avoid redundancy.

10. **Define bottom-up approach of data warehouse.**

The bottom up approach is an incremental approach to build a data warehouse. We build the data marts separately at different points of time as and when the specific subject area requirements are clear. The data marts are integrated or combined together to form a data warehouse.

11. **Define conformed dimension.**

A Conformed dimension has consistent dimension keys, consistent attribute names and consistent values across separate data marts. The conformed dimension means exact same thing with every fact table it is joined.

12. **List out the nine-step method followed in the design of a data warehouse.**

1. Choosing the subject matter
2. Deciding what a fact table represents
3. Identifying and conforming the dimensions
4. Choosing the facts
5. Storing pre calculations in the fact table
6. Rounding out the dimension table
7. Choosing the duration of the db
8. The need to track slowly changing dimensions
9. Deciding the query priorities and query models

PART B

1. Explain the 3-tier data ware house architecture and its various components. ***Evaluate***
2. Differentiate Operational database versus data warehouse. **Understand**
3. Illustrate the various data warehouse components. **Understand**
4. Explain about oracle autonomous data warehouse. **Understand**
5. Design a data warehouse architecture for a hospital management system. **Apply**

UNIT II ETL AND OLAP TECHNOLOGY

PART A

1. **Define ETL**
   - EXTRACT data from its original source
   - TRANSFORM data by deduplication it, combining it, and ensuring quality, to then
   - LOAD data into the target database
   -
2. **What are ETL transformation types?**
   - Deduplication. Data deduplication is a technique used in data management to identify and eliminate duplicate data entries within a data set. ...
   - Derivation, Joining, Aggregating, Splitting, Cleaning, Sorting and ordering.
   - Mapping, Differentiate ETL and ELT. ETL, which stands for Extract, Transform, and Load, involves transforming data on a separate processing server before transferring it to the data warehouse. On the other hand, ELT, or Extract, Load, and Transform, performs data transformations directly within the data warehouse itself.

3. **List out the three data warehouse models.**
   - Enterprise warehouse, - Data mart, and - Virtual warehouse

4. **Define Enterprise warehouse.**

   An enterprise warehouse collects all of the information about subjects spanning the entire organization. It provides corporate-wide data integration. An enterprise data warehouse may be implemented on traditional mainframes, computer super servers, or parallel architecture platforms.

5. **Define Data mart.**

   A data mart contains a subset of corporate-wide data that is of value to a specific group of users. The scope is confined to specific selected subjects.

6. **List out the types of data marts and explain.**
   - ***Independent*** data marts are sourced from data captured from one or more operational systems or external information providers, or from data generated locally within a particular department or geographic area.
   - ***Dependent*** data marts are sourced directly from enterprise data warehouses.

7. **Define Virtual warehouse.**

   A virtual warehouse is a set of views over operational databases. A virtual warehouse is easy to build but requires excess capacity on operational database servers.

8. **Write the functions of back-end tools and utilities used in datawarehouse systems.**
   - **Data extraction** gathers data from multiple, heterogeneous, and external sources
   - **Data cleaning** detects errors in the data and rectifies them
   - **Data transformation** converts data from legacy or host format to warehouse format.
   - **Load** sorts, summarizes, consolidates, computes views, checks integrity, and builds indices and partitions.
   - **Refresh** propagates the updates from the data sourcesto the warehouse.

- **What are the different OLAP servers?** ○ Relational OLAP (ROLAP) servers ○ Multidimensional OLAP (MOLAP) servers ○ Hybrid OLAP (HOLAP) servers

10. Define ROLAP.

ROLAP are the intermediate servers that stand in between a relational Back-end server and client front-end tools. They use a *relational* or *extended-relational DBMS* to store and manage warehouse data, and OLAP middleware to support missing pieces.

Example: DSS server of Micro strategy

11. Define MOLAP.

MOLAP servers support multidimensional data views through *array based multidimensional storage engines*.
They map multidimensional views directly to data cube array structures.

12. Define HOLAP.

The hybrid OLAP approach combines ROLAP and MOLAP technology, benefiting from the greater scalability of ROLAP and the faster computation of MOLAP.
Example: Microsoft SQL Server 2000 supports a hybrid OLAP server.

13. **What is drill-down operation?**

Drill-down is the reverse of roll-up operation. It navigates from less detailed data to more detailed data. Drill-down operation can be taken place by stepping down a concept hierarchy for a dimension.

14. **What is slice operation?**

The slice operation performs a selection on one dimension of the cube resulting in a sub cube.

15. **What is dice operation?**

The dice operation defines a sub cube by performing a selection on two (or) more dimensions.

16. **What is pivot operation?**

Pivot operation is a visualization operation that rotates the data axes in an alternative presentation of the data.

PART B

1. Illustrate data modeling life cycle with neat sketch.

2. Explain about various OLAP operations in detail.

3. Discuss the various types of OLAP techniques.

4. Discuss various delivery process followed in data warehouse.

UNIT III META DATA, DATA MART AND PARTITIONING STRATEGY

1. Define data warehouse metadata.

Data warehouse **metadata** are data defining the warehouse objects. A metadata repository provides details regarding the warehouse structure, data history, the algorithms used for summarization, mappings from the source data to the warehouse form, system performance, and business terms and issues.

2. What are the types of partition in data warehouse?

There are three typical strategies for partitioning data:

- Horizontal partitioning (often called sharing). In this strategy, each partition is a separate data store, but all partitions have the same schema. ...
- Vertical partitioning. ...
- Functional partitioning.

3. What are the 3 characteristics of data mart?

Characteristics of data marts

Typically uses a dimensional model and star schema. Contains a curated subset of data from the larger data warehouse. The data is highly structured, having been cleansed and confirmed by the enterprise data team to make it easy to understand and query.

4. What is a metadata repository?

A metadata repository is a software tool that stores descriptive information about the data model used to store and share metadata. Metadata repositories combine diagrams and text, enabling metadata integration and change.

5. What is metadata used for?

Simply defined, metadata is the summary and the description about your data that is used to classify, organize, label, and understand data, making sorting and searching for data much easier. Without it, companies can't manage the huge amounts of data created and collected across an enterprise

PART B

1. Discuss the role of Meta data in data warehousing.

2. Explain in detail about horizontal partitioning techniques.

3. Illustrate the two ways of vertical portioning.
4. Summarize the challenges for Meta management.

UNIT IV DIMENSIONAL MODELLING AND SCHEMA

PART A

1. Define multi-dimensional data model.

- A multidimensional data model is used for the design of corporate data warehouses and departmental data marts.
- Multidimensional model is the data cube, which consists of a large set of facts (or measures) and a number of dimensions.

2. What is a data cube?

A **data cube** allows data to be modeled and viewed in multiple dimensions. It is defined by dimensions and facts.

3. Define dimensions.

Dimensions are the perspectives or entities with respect to which an organization wants to keep records.

Each dimension may have a table associated with it, called a **dimension table**, which further describes the dimension.

Example: A dimension table for *item* may contain the attributes *item name*, *brand*, and *type*.

4. Define Fact Table.

Facts are numeric measures. The **fact table** contains the names of the *facts*, or measures, as well as keys to each of the related dimension tables.

5. List out the Schemas for Multidimensional Data M

- Star Schema, - Snowflake Schema, - Fact Constellation Schema.

6. Define star schema.

In the star schema, the data warehouse contains,

(1) a large central table (**fact table**) containing the bulk of the data, with no redundancy, and

(2) a set of smaller attendant tables (**dimension tables**), one for each dimension.

The schema graph resembles a starburst, with the dimension tables displayed in a radial pattern around the central fact table.

7. Define snowflake schema.

The snowflake schema is a variant of the star schema model in which the dimension tables are *normalized*, thereby further splitting the data into additional tables. The resulting schema graph forms a shape similar to a snowflake.

8. Define fact constellation schema.

Sophisticated applications may require multiple fact tables to *share* dimension tables. This kind of schema can be viewed as a collection of stars, and hence is called a **galaxy schema** or a **fact constellation**.

9. Define conceptual hierarchies.

Concept hierarchies organize the values of attributes or dimensions into gradual abstraction levels. They are useful in mining at multiple abstraction levels.

10. Define data cube measure.

A data cube **measure** is a numeric function that can be evaluated at each point in the data cube space. A measure value is computed for a given point by aggregating the data corresponding to the respective dimension– value pairs defining the given point.

PART B

1. ***Design*** a star-schema, snow-flake schema and Fact-constellation schema for the following data warehouse that consist of the following four dimensions: (Time, Item, Branch and Location) .Include the appropriate measures required for the schemas. ***Create***
2. Discuss about multidimensional database, data mart and data cube? Explain schemas for multi-dimensional database.
3. Explain the **star schema**, **snowflake schema** and **fact constellation schema** with examples.
4. Construct a star schema for hospital management system

UNIT 5 SYSTEM AND PROCESS MANAGER

PART-A

1. Identify the role of system manager in data warehousing.

The system configuration manager is responsible for the management of the setup and configuration of data warehouse. The structure of configuration manager varies from one operating system to another. In UNIX structure of configuration, the manager varies from vendor to vendor.

2**. What is the role of data warehouse manager**?

A Data Warehousing Manager manages the daily activities of the team responsible for the design, implementation, maintenance, and support of data warehouse systems and projects. Oversees data design and the creation of database architecture and data repositories.

3. **What is backup and recovery in data warehouse?**

Backup refers to storing a copy of original data separately. Recovery refers to restoring the lost data in case of failure. 02. So we can say Backup is a copy of data which is used to restore original data after a data loss/damage occurs

Part B

1. Describe in detail about working of system scheduling manager.
2. Summarize the role of load manager and warehouse manager.
3. Discuss about query manager process in detail.

PART – C

1. Suppose that a data warehouse for Big University consists of the four dimensions student, course, semester, and instructor, and two measures count and avg grade. At the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination.
   (a) Draw a snowflake schema diagram for the data warehouse.
   (b) Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations (e.g., roll-up from semester to year) should you perform in order to list the average grade of CS courses for each Big University student.
   (c) If each dimension has five levels (including all), such as "student < major < status < university < all", how many cuboids will this cube contain
   (including the base and apex cuboids)? ***Create***

2. A data warehouse can be modeled by either a star schema or a snowflake schema. Briefly describe the similarities and the differences of the two models, and then analyze their advantages and disadvantages with regard to one another with example. ***Understand***