

DATA WAREHOUSING
(CSEN 3237)

Time Allotted : 2½ hrs

Full Marks : 60

Figures out of the right margin indicate full marks.

*Candidates are required to answer Group A and
any 4 (four) from Group B to E, taking one from each group.*

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve:

12 × 1 = 12

Choose the correct alternative for the following

- (i) Which is considered the heart of the data warehouse?
 - (a) Relational database server
 - (b) Data Mart database server
 - (c) Data warehouse database server
 - (d) All of the above
- (ii) Reading from the database is synonymous with which process?
 - (a) Extraction
 - (b) Transformation
 - (c) Loading
 - (d) All of the above
- (iii) On which of these factors do OLTP and OLAP differ?
 - (a) Database size
 - (b) Complexity of queries
 - (c) Types of business tasks
 - (d) All of the above
- (iv) Which architecture is suited for analytical processing and complex queries on large datasets?
 - (a) ETL
 - (b) CRM
 - (c) OLTP
 - (d) OLAP
- (v) What are the components of metadata?
 - (a) Data structure
 - (b) Summarization algorithm
 - (c) Mapping connecting the data warehouse with the operational environment
 - (d) All of the above
- (vi) What is the significance of ETL for businesses?
 - (a) Analysis of business data
 - (b) Repository of data
 - (c) Facilitation of data relocation
 - (d) All of the above
- (vii) Data scrubbing is _____.
 - (a) a process to reject data from the data warehouse and to create the necessary indexes.
 - (b) a process to load the data in the data warehouse and to create the necessary indexes.
 - (c) a process to upgrade the quality of data after it is moved into a data warehouse.
 - (d) a process to upgrade the quality of data before it is moved into a data warehouse.

- (viii) Business Intelligence and data warehousing is used for _____.
 (a) Forecasting (b) Data Mining.
 (c) Analysis of large volumes of product sales data. (d) All of the above.
- (ix) What is the main purpose of a data mart?
 (a) To store historical data for reporting and analysis
 (b) To integrate data from multiple sources into a centralized repository
 (c) To provide a subset of data for a specific department or user group
 (d) To perform online transaction processing
- (x) Which of the following is NOT a key factor to consider when designing an ETL pipeline?
 (a) Data quality (b) Performance optimization
 (c) Data duplication (d) Color scheme of the reports

Fill in the blanks with the correct word

- (xi) _____ is a process of converting logical data model into physical database design structures such as tables, indexes, and views.
- (xii) The _____ dimension signifies the time-related attributes such as year, quarter, month, and day.
- (xiii) _____ is a technique used to optimize query performance by limiting the amount of data scanned during query execution.
- (xiv) _____ is a characteristic of a data warehouse.
- (xv) _____ is used to identify and resolve duplicate records in data cleansing processes.

Group - B

2. Suppose that a data warehouse consists of the four dimensions *product, time, customer, and salesrep*, and two measures *SalesQuantity and SalesRupees*.

Dimensions and measures are explained as follows:

Dimension-Product: ProductKey - number(10) Primary Key, ProductName - varchar2(50) not null, ProductCode - varchar2(10) not null, BrandName - varchar2(50) not null.

Dimension-Time: TimeKey - number(10) Primary Key, Date date not null, Month number(2) not null, Quarter number(1) not null, Year number(2) not null.

Dimension-Customer: CustomerKey - number(10) Primary Key, CustomerName - varchar2(50), CustomerCode - varchar2(20), MaritalStatus - varchar2(2), Address - varchar2(255), State - varchar2(5), Zip - varchar2(10).

Dimension-Salesrep: SalesrepKey - number(10) Primary Key, SalepersonName - varchar2(50), TerritoryName varchar2(50), RegionName varchar2(50).

SalesQuantity and SalesRupees are measures of products sold by Salesrep to Customers on Quarter basis.

- (i) Enumerate STAR schema that is popularly used for modeling data warehouses. [CO1(Learn and understand/LOCQ)]
- (ii) Draw a STAR schema diagram for the above data warehouse. [CO4(Understand/IOCQ)]
- (iii) Define OLTP and OLAP. [CO1(Learn and understand/LOCQ)]

[2 + 4 + (3 + 3)] = 12

3. (a) What are the key components of Data Warehousing Architecture? What are the different types of Data Warehousing Architectures? [CO1 (Learn and Understand/LOCQ)]
 (b) What is the role of ETL in Data Warehousing Architecture? How does Data Warehousing Architecture support decision-making? [CO2 (Understand/IOCQ)]
 (c) What are the key phases of a data warehouse life cycle? What is the difference between Conceptual, Logical, and Physical Data Models? [CO1 (Learn and Understand/LOCQ)]

4 + 4 + 4 = 12

Group - C

4. (a) Suppose that a data warehouse for Big University consists of four dimensions: *student*, *course*, *semester*, and *instructor*, and two measures *count* and *avg_grade*. Draw a snowflake schema diagram for the data warehouse. [CO4 (Understand/IOCQ)]
 (b) The sales for New York, Los Angeles, Chicago, and Houston are shown for the time dimension (organized in quarters) and the product dimension (classified according to the types of products sold) are displayed in the 2D table below. The fact or measure is displayed in revenue_sold (in thousands). Create slices of locations to view the data.

Quarter	Product Type	New York	Los Angeles	Chicago	Houston
Q1	Electronics	250	180	200	150
Q2	Electronics	300	220	230	180
Q3	Electronics	280	200	210	160
Q4	Electronics	320	240	250	190
Q1	Apparel	150	100	120	90
Q2	Apparel	180	120	140	110
Q3	Apparel	160	110	130	100
Q4	Apparel	200	140	160	120

[[CO4](Remember/LOCQ)]

- (c) What is a star schema in data warehousing?

[[CO2](Apply/LOCQ)]

6 + 4 + 2 = 12

5. (a) What is a Conceptual Model in Data Warehousing? [[CO3](Analyse/HOCQ)]
 (b) What is the difference between Measurement and Context in Dimensional Design? [[CO4](Remember/LOCQ)]
 (c) What are Facts and Dimensions in Dimensional Modeling? [[CO2](Apply/IOCQ)]
 (d) What is the significance of Grouping Dimensions and Facts? [[CO2](Apply/IOCQ)]

3 + 3 + 3 + 3 = 12

Group - D

6. (a) What are the benefits of partitioning in a data warehouse? [[CO3](Analyse/LOCQ)]
 (b) What is partition pruning? [[CO4](Remember/LOCQ)]

- (c) What is a materialized view in Oracle? [[CO3](Analyse/LOCQ)]
 (d) What are some benefits of using materialized views in a data warehouse? [[CO4](Remember/LOCQ)]
3 + 3 + 3 + 3 = 12

7. (a) What is Slowly Changing Dimension (SCD), and how do you implement it? [[CO3](Analyse/HOCQ)]
 (b) What are the parameters that characterize the basic functions of a computer network? [[CO4](Remember/LOCQ)]
 (c) Describe briefly the functions of the Transport layer of the ISO/OSI reference architecture. [[CO2](Apply/IOCQ)]
 (d) How do you create a materialized view? [[CO2](Apply/IOCQ)]
3 + 3 + 3 + 3 = 12

Group - E

8. (a) What factors should be considered when implementing cleansing and transformation processes? [[CO3](Analyse/LOCQ)]
 (b) Why is transformation important for dimension attributes in a data warehouse? [[CO4](Remember/LOCQ)]
 (c) What are common challenges in integrating data from multiple sources? [[CO2](Apply/LOCQ)]
 (d) Why is normalization a crucial transformation step? [[CO2](Apply/IOCQ)]
3 + 3 + 3 + 3 = 12
9. (a) What is incremental extraction? What are some common data extraction techniques? [[CO3](Analyse/LOCQ)]
 (b) What is an example of query-based extraction? What is the advantage of log-based extraction? [[CO4](Remember/LOCQ)]
 (c) Why is data cleansing important in ETL? [[CO2](Apply/IOCQ)]
 (d) What is the difference between ETL and ELT? [[CO2](Apply/LOCQ)]
3 + 3 + 3 + 3 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	60.42	33.33	6.25