DEPARTMENT OF ARTIFICIAL INTELLIGECE AND DATA SCIENCE
QUESTION BANK

Year/Sem/Sec: III/VI/A, B, C
Course Code/ Course Name: U23AIV31/DATA WAREHOUSING

UNIT I INTRODUCTION TO DATA WAREHOUSE

Introduction - components- operational database Vs data warehouse – Data warehouse Architecture – Three - tier Data Warehouse Architecture - Autonomous Data Warehouse- Autonomous Data Warehouse Vs Snowflake - Modern Data Warehouse

S.No	PART A	CO	BTL
1	**Define Data Warehouse. (Nov/Dec 2019)** A Data Warehouse is a centralized repository used to store large volumes of historical data collected from multiple heterogeneous sources. It is specifically designed to support decision-making and analytical processing. Unlike operational databases, it is optimized for query and analysis. Data in a warehouse is subject-oriented, integrated, time-variant, and non-volatile. It helps management in strategic planning. Data Warehouses are widely used in business intelligence systems.	CO1	BTL1
2	**List the major components of a Data Warehouse. (Apr/May 2018)** The major components of a Data Warehouse include data sources, ETL tools, data storage, metadata, and front-end query tools. Data sources provide raw operational data. ETL tools extract, transform, and load data into the warehouse. Data storage maintains integrated and historical data. Metadata defines the structure and meaning of data. Query tools allow users to analyze the data.	CO1	BTL1
3	**What is an operational database? (Nov/Dec 2017)** An operational database is designed to support daily business operations. It stores current and detailed transactional data. These databases allow frequent insert, update, and delete operations. They are optimized for fast transaction	CO1	BTL1

processing. Operational databases support OLTP systems. Examples include banking and reservation systems.

4	**Mention any two characteristics of a Data Warehouse. (Apr/May 2019)** One characteristic of a Data Warehouse is subject-orientation, where data is organized around business subjects. Another characteristic is time-variance, which stores historical data for long-term analysis. Data is also integrated from multiple sources. It is stored in a consistent format. Data is non-volatile and stable. These characteristics support effective decision-making.	CO1	BTL1
5	**Differentiate OLTP and OLAP (any two points). (Nov/Dec 2018)** OLTP systems are designed for handling routine transactional operations. OLAP systems are designed for analytical and decision-support queries. OLTP uses current operational data, while OLAP uses historical data. OLTP focuses on speed and efficiency. OLAP focuses on complex queries and reports. Both systems serve different organizational needs.	CO1	BTL2
6	**What is subject-oriented data? (Apr/May 2017)** Subject-oriented data is organized around key business subjects such as sales, customers, or products. It helps users focus on specific areas of interest. This approach simplifies data analysis. It differs from application-oriented operational data. Subject orientation improves clarity and usability. It is a core feature of data warehouses.	CO1	BTL1
7	**Define time-variant property of a Data Warehouse. (Nov/Dec 2016)** Time-variant property means that data is associated with a specific time period. Historical data is stored for many years. Each record contains a time stamp. This enables trend analysis and forecasting. It supports long-term decision making. Time-variant data is essential for business analysis.	CO1	BTL1
8	**What is meant by non-volatile data? (Apr/May 2016)** Non-volatile data means that data is not frequently changed once it is stored. Data is mainly loaded in batch mode. Deletion and modification operations are rare. This ensures data consistency and stability. Analytical queries use this stable data. It improves the reliability of results.	CO1	BTL1
9	**State the purpose of metadata in a Data Warehouse. (Nov/Dec 2020)** Metadata describes the structure, meaning, and source of data. It acts as a directory for users and administrators. Metadata explains table names,	CO1	BTL2

	columns, and data types. It helps in understanding and managing data. Metadata supports ETL and query processing. It improves overall system efficiency.		
10	**What is Data Warehouse Architecture? (Apr/May 2021)** Data Warehouse Architecture defines how data flows within the warehouse system. It specifies data sources, storage, and access methods. Architecture ensures efficient data integration. It improves system scalability and performance. A good architecture supports analytical processing. It plays a key role in decision support systems.	CO1	BTL2
11	**List the layers in a Data Warehouse Architecture. (Nov/Dec 2019)** Data Warehouse Architecture consists of three layers. The bottom tier stores the data warehouse database. The middle tier contains OLAP servers. The top tier includes front-end tools. Each layer has a specific function. Together, they support data analysis.	CO1	BTL1
12	**What is a three-tier Data Warehouse Architecture? (Apr/May 2020)** Three-tier architecture divides the warehouse into bottom, middle, and top tiers. The bottom tier stores integrated data. The middle tier provides OLAP services. The top tier supports reporting and visualization. This architecture improves modularity. It enhances system scalability and performance.	CO1	BTL2
13	**Define bottom tier in three-tier architecture. (Nov/Dec 2018)** The bottom tier is the data warehouse server. It stores cleansed and integrated data. Data is loaded using ETL tools. It ensures data security and consistency. This tier forms the foundation of the warehouse. It supports analytical processing.	CO1	BTL1
14	**What is the role of middle tier in Data Warehouse architecture? (Apr/May 2019)** The middle tier acts as the OLAP layer. It processes complex analytical queries. It provides multidimensional views of data. This tier improves query performance. It supports fast data analysis. It connects storage and user layers.	CO1	BTL2
15	**What is front-end layer in Data Warehouse? (Nov/Dec 2017)** The front-end layer provides an interface for users. It includes reporting and visualization tools. Users access data through dashboards. It supports ad-hoc queries. This layer helps in decision making. It is user-centric in nature.	CO1	BTL1
16	**What is Autonomous Data Warehouse? (Apr/May 2022)**	CO1	BTL2

Autonomous Data Warehouse is a cloud-based data warehouse system. It automates tuning, scaling, and maintenance. Machine learning is used for optimization. It reduces human intervention. The system is self-securing and self-repairing. It improves reliability and efficiency.

17	**Mention any two features of Autonomous Data Warehouse. (Nov/Dec 2022)** Automatic scaling is a key feature of Autonomous Data Warehouse. Self-tuning using machine learning is another feature. It reduces administrative workload. Security updates are automated. Performance is continuously optimized. These features improve availability and efficiency.	CO1	BTL1
18	**What is Snowflake Data Warehouse? (Apr/May 2023)** Snowflake is a cloud-native data warehouse platform. It separates storage and compute resources. It supports scalable analytics. Snowflake uses SQL for querying. It supports structured and semi-structured data. It is widely used in modern analytics.	CO1	BTL1
19	**State one advantage of Autonomous Data Warehouse. (Nov/Dec 2021)** One major advantage is reduced maintenance cost. Automation eliminates manual tuning. System uptime is improved. Performance optimization is continuous. It saves time for administrators. This allows focus on analytics tasks.	CO1	BTL2
20	**Mention one limitation of traditional Data Warehouse. (Apr/May 2018)** Traditional data warehouses are difficult to scale. Hardware costs are high. Manual tuning is required. Maintenance effort is significant. Performance upgrades are complex. This limits flexibility and agility.	CO1	BTL1
21	**What is cloud-based Data Warehouse? (Nov/Dec 2020)** A cloud-based Data Warehouse is hosted on cloud platforms. It provides scalable storage and computing. Users pay based on usage. It supports remote access. Maintenance is handled by the provider. It improves cost efficiency and flexibility.	CO1	BTL1
22	**Define Modern Data Warehouse. (Apr/May 2021)** A Modern Data Warehouse integrates cloud, big data, and analytics. It supports structured and unstructured data. ELT is preferred over ETL. It enables real-time and batch processing. It supports advanced analytics. It improves business intelligence.	CO1	BTL2

23	**Mention any two features of Modern Data Warehouse. (Nov/Dec 2022)** Cloud scalability is an important feature. Support for real-time analytics is another feature. It integrates with big data tools. It supports self-service analytics. Performance is highly optimized. These features enhance decision making.	CO1	BTL1
24	**What is scalability in Data Warehousing? (Apr/May 2019)** Scalability refers to the ability to handle increasing data volumes. Resources can be added easily. It supports business growth. Performance remains stable with growth. Scalability is essential for modern systems. It ensures long-term usability.	CO1	BTL2
25	**Give one difference between Autonomous Data Warehouse and Snowflake. (Nov/Dec 2023)** Autonomous Data Warehouse is fully automated using machine learning. Snowflake requires some manual configuration. ADW focuses on automation and self-management. Snowflake focuses on flexibility and performance. Both are cloud-based solutions. They serve modern analytical needs.	CO1	BTL3

S. No	PART B	CO	BTL
1	Explain the concept of Data Warehouse and discuss its characteristics in detail.*(Nov/Dec 2019)*	CO1	BTL2
2	Describe the major components of a Data Warehouse with a neat diagram.(Apr/May 2018)	CO1	BTL2
3	Differentiate Operational Database and Data Warehouse with suitable examples.(Nov/Dec 2017)	CO1	BTL4
4	Explain Data Warehouse Architecture and discuss its various layers.(Apr/May 2021)	CO1	BTL2
5	Describe the Three-Tier Data Warehouse Architecture and explain the function of each tier.(Apr/May 2020)	CO1	BTL2
6	Explain OLTP and OLAP systems and compare them with suitable examples.(Nov/Dec 2018)	CO1	BTL4
7	Discuss the role and importance of metadata in a Data Warehouse.(Nov/Dec 2020)	CO1	BTL3

S. No	PART B	CO	BTL
8	Explain Autonomous Data Warehouse and its key features.(Apr/May 2022)	CO1	BTL3
9	Compare Autonomous Data Warehouse and Snowflake Data Warehouse.(Nov/Dec 2023)	CO1	BTL4
10	Describe the concept of Modern Data Warehouse and explain its features.(Nov/Dec 2022)		

UNIT II ETL AND OLAP TECHNOLOGY

ETL Vs ELT – Types of Data warehouses - Data warehouse Design and Modeling - Delivery Process - Online Analytical Processing (OLAP) - Characteristics of OLAP - Online Transaction Processing (OLTP) Vs OLAP - OLAP operations- Types of OLAP- ROLAP Vs MOLAP Vs HOLAP.

S.No	PART A	CO	BTL
1	**What is ETL? (Nov/Dec 2019)** ETL stands for Extract, Transform, and Load. It is a process used to move data from multiple source systems into a data warehouse. Data is first extracted from operational databases. It is then transformed by cleaning, filtering, and converting formats. Finally, the transformed data is loaded into the warehouse. ETL ensures data quality and consistency.	CO2	BTL1
2	**What is ELT? (Apr/May 2021)** ELT stands for Extract, Load, and Transform. In this approach, data is first extracted from source systems. The data is then directly loaded into the data warehouse. Transformation is performed inside the warehouse. ELT leverages powerful cloud processing engines. It improves scalability and performance.	CO2	BTL1
3	**Differentiate ETL and ELT. (Nov/Dec 2020)** ETL transforms data before loading it into the warehouse. ELT transforms data after loading it into the warehouse. ETL uses external transformation engines. ELT uses the warehouse's processing power.	CO2	BTL2

S.No	PART A	CO	BTL
	ETL is common in traditional systems. ELT is widely used in modern cloud platforms.		
4	**What are the types of Data Warehouses? (Apr/May 2019)** Data Warehouses are classified into Enterprise Data Warehouse, Data Mart, and Virtual Data Warehouse. Enterprise Data Warehouse supports organization-wide analysis. Data Mart supports department-level analysis. Virtual Data Warehouse provides logical views of data. Each type serves specific business needs. They improve decision-making efficiency.	CO2	BTL1
5	**What is an Enterprise Data Warehouse? (Nov/Dec 2018)** An Enterprise Data Warehouse stores integrated data for the entire organization. It supports enterprise-wide decision making. Data is collected from multiple departments. It ensures data consistency and accuracy. It supports complex analytical queries. Enterprise warehouses are highly scalable systems.	CO2	BTL1
6	**What is a Data Mart? (Apr/May 2017)** A Data Mart is a subset of a data warehouse. It is designed for a specific department or business function. Data marts are smaller in size. They provide faster access to relevant data. They reduce system complexity. Data marts support focused analysis.	CO2	BTL1
7	**What is Data Warehouse Design? (Nov/Dec 2019)** Data Warehouse design defines how data is structured and organized. It includes schema design and dimensional modeling. Good design improves query performance. It ensures data consistency across tables. Design supports analytical requirements. It is a critical development stage.	CO2	BTL2
8	**What is Data Modeling in Data Warehousing? (Apr/May 2020)** Data modeling defines how data entities are represented in the warehouse. It uses fact tables and dimension tables. It supports multidimensional analysis. Proper modeling improves query	CO2	BTL2

S.No	PART A	CO	BTL
	efficiency. It simplifies complex queries. It is essential for OLAP operations.		
9	**What is the delivery process in Data Warehousing? (Nov/Dec 2018)** The delivery process defines the steps to build a data warehouse. It includes requirement analysis and system design. Development and testing are carried out. Deployment makes the system operational. Maintenance ensures long-term reliability. The process follows an incremental approach.	CO2	BTL2
10	**What is Online Analytical Processing (OLAP)? (Apr/May 2019)** OLAP is a technology used for analytical data processing. It supports complex queries on large datasets. OLAP provides multidimensional data views. It enables fast query response. It supports trend and pattern analysis. OLAP helps in decision making.	CO2	BTL1
11	**List the characteristics of OLAP. (Nov/Dec 2020)** OLAP supports multidimensional data analysis. It provides fast query performance. It supports complex calculations. OLAP handles large volumes of historical data. It enables interactive querying. These characteristics support decision support systems.	CO2	BTL1
12	**What is OLTP? (Apr/May 2018)** OLTP stands for Online Transaction Processing. It supports day-to-day business operations. OLTP systems handle large numbers of short transactions. They ensure data accuracy and integrity. OLTP focuses on speed and reliability. These systems support operational databases.	CO2	BTL1
13	**Differentiate OLTP and OLAP. (Nov/Dec 2018)** OLTP systems handle routine transactional operations. OLAP systems handle analytical queries. OLTP uses current operational data. OLAP uses historical data. OLTP focuses on speed. OLAP focuses on analysis and reporting.	CO2	BTL2
14	**What are OLAP operations? (Apr/May 2021)** OLAP operations are used to analyze multidimensional data. Common operations include roll-up and drill-down. Slice and dice operations	CO2	BTL2

S.No	PART A	CO	BTL
	filter data. Pivot changes the orientation of data. These operations enable interactive analysis. They improve data exploration.		
15	**Explain roll-up operation in OLAP. (Nov/Dec 2019)** Roll-up operation summarizes data to a higher level. It reduces data detail. It follows dimension hierarchies. It aggregates data values. Roll-up provides a summarized view. It supports strategic decision making.	CO2	BTL2
16	**Explain drill-down operation in OLAP. (Apr/May 2020)** Drill-down operation moves from summarized data to detailed data. It increases data granularity. It follows hierarchy levels. It helps identify root causes. Drill-down supports detailed analysis. It improves analytical depth.	CO2	BTL2
17	**What is slice operation in OLAP? (Nov/Dec 2017)** Slice operation selects a single dimension value. It creates a sub-cube. It reduces dimensionality. It focuses on specific data. Slice improves analysis clarity. It supports quick insights.	CO2	BTL1
18	**What is dice operation in OLAP? (Apr/May 2018)** Dice operation selects multiple dimension values. It creates a smaller data cube. It supports complex filtering. Dice enables comparative analysis. It improves flexibility. It is widely used in OLAP queries.	CO2	BTL1
19	**What are the types of OLAP? (Nov/Dec 2019)** The types of OLAP are ROLAP, MOLAP, and HOLAP. ROLAP uses relational databases. MOLAP uses multidimensional cubes. HOLAP combines both approaches. Each type has advantages. They support analytical processing.	CO2	BTL1
20	**What is ROLAP? (Apr/May 2017)** ROLAP stands for Relational OLAP. It stores data in relational databases. It uses SQL queries. It supports large datasets. ROLAP is highly scalable. It is flexible in design.	CO2	BTL1
21	**What is MOLAP? (Nov/Dec 2018)** MOLAP stands for Multidimensional OLAP. It stores data in cube structures. Data is pre-aggregated. It provides fast query response.	CO2	BTL1

S.No	PART A	CO	BTL
	MOLAP supports complex calculations. It is suitable for smaller datasets.		
22	**What is HOLAP? (Apr/May 2019)** HOLAP stands for Hybrid OLAP. It combines ROLAP and MOLAP features. Aggregated data is stored in cubes. Detailed data is stored in relational databases. HOLAP balances speed and scalability. It supports efficient analysis.	CO2	BTL1
23	**Differentiate ROLAP and MOLAP. (Nov/Dec 2020)** ROLAP uses relational databases for storage. MOLAP uses multidimensional cubes. ROLAP handles large datasets efficiently. MOLAP provides faster query performance. ROLAP is scalable. MOLAP is storage intensive.	CO2	BTL2
24	**Differentiate MOLAP and HOLAP. (Apr/May 2021)** MOLAP stores all data in cubes. HOLAP stores aggregated data in cubes. MOLAP offers high speed. HOLAP balances speed and scalability. HOLAP handles large datasets better. Both support OLAP queries.	CO2	BTL2
25	**Differentiate ROLAP, MOLAP and HOLAP. (Nov/Dec 2023)** ROLAP uses relational storage. MOLAP uses cube storage. HOLAP combines both methods. ROLAP is scalable. MOLAP is fast. HOLAP balances performance and storage.	CO2	BTL3
S. No	**Part B**	CO	BTL
1	Explain the architecture of a Data Warehouse with a neat diagram. (Nov/Dec 2019)	CO2	BTL4
2	Discuss the components of a Data Warehouse and their functions in detail. (Apr/May 2018)	CO2	BTL4
3	Compare OLTP and OLAP systems with suitable examples. (Nov/Dec 2018)	CO2	BTL4
4	Describe the three-tier Data Warehouse architecture and explain each tier. (Apr/May 2020)	CO2	BTL4

S.No	PART A	CO	BTL
5	Explain the ETL process in detail with suitable examples. (Nov/Dec 2020)	CO2	BTL4
6	Discuss the various OLAP operations such as roll-up, drill-down, slice, and dice. (Apr/May 2019)	CO2	BTL4
7	Explain Data Warehouse design and dimensional modeling concepts. (Nov/Dec 2019)	CO2	BTL4
8	Describe the types of OLAP servers (ROLAP, MOLAP, HOLAP) with advantages and limitations. (Apr/May 2021)	CO2	BTL5
9	Explain the delivery process of a Data Warehouse with suitable stages. (Nov/Dec 2018)	CO2	BTL4
10	Discuss Modern Data Warehouse concepts and explain how they differ from traditional warehouses. (Nov/Dec 2022)	CO2	BTL5

UNIT III META DATA, DATA MART AND PARTITION STRATEGY

Meta Data – Categories of Metadata – Role of Metadata – Metadata Repository – Challenges for Meta Management - Data Mart – Need of Data Mart- Cost Effective Data Mart- Designing Data Marts- Cost of Data Marts- Partitioning Strategy – Vertical partition – Normalization – Row Splitting – Horizontal Partition

S.No	PART A	CO	BTL
1	**Define Metadata. (Nov/Dec 2019)** Metadata is data about data that describes the structure, meaning, origin, and usage of data stored in a data warehouse. It helps users understand the data content and format. Metadata supports data management and query processing. It improves data consistency and usability.	CO3	BTL1
2	**List the categories of Metadata. (Apr/May 2018)** The main categories of metadata are technical metadata, business metadata, and operational metadata. Technical metadata describes database structures. Business metadata defines business rules and meanings. Operational metadata contains ETL and process information.	CO3	BTL1
3	**What is technical metadata? (Nov/Dec 2017)**	CO3	BTL1

S.No	PART A	CO	BTL
	Technical metadata describes the physical structure of data. It includes table names, column names, data types, and indexes. It supports developers and administrators. It helps in ETL and query optimization.		
4	**Define business metadata. (Apr/May 2019)** Business metadata explains data in business terms. It includes definitions, rules, and policies. It helps non-technical users understand data. It improves decision-making accuracy.	CO3	BTL1
5	**What is operational metadata? (Nov/Dec 2018)** Operational metadata stores information about data loading and processing. It includes load time, data refresh frequency, and error logs. It supports monitoring and maintenance. It ensures system reliability.	CO3	BTL2
6	**State the role of metadata in a Data Warehouse. (Apr/May 2017)** Metadata acts as a guide for users and administrators. It helps in data integration and transformation. It supports efficient query processing. Metadata ensures consistency and clarity.	CO3	BTL2
7	**What is a Metadata Repository? (Nov/Dec 2016)** A Metadata Repository is a centralized storage for metadata. It maintains definitions, mappings, and rules. It supports data governance. It improves system management and integration.	CO3	BTL1
8	**Mention any two challenges of metadata management. (Apr/May 2016)** Maintaining metadata consistency is a major challenge. Handling frequent changes in data sources is another challenge. Metadata integration across tools is complex. High maintenance effort is required.	CO3	BTL2
9	**Define Data Mart. (Nov/Dec 2020)** A Data Mart is a subset of a data warehouse. It focuses on a specific business function such as sales or finance. It stores summarized data. It supports faster decision-making.	CO3	BTL1
10	**What is the need for a Data Mart? (Apr/May 2021)** Data marts provide department-specific data access. They reduce query complexity. They improve performance and response time. They support focused analysis.	CO3	BTL2
11	**List the types of Data Marts. (Nov/Dec 2019)**	CO3	BTL1

S.No	PART A	CO	BTL
	The types of data marts are dependent data marts, independent data marts, and hybrid data marts. Each type differs in data source and integration level.		
12	**What is a cost-effective Data Mart? (Apr/May 2020)** A cost-effective data mart minimizes storage and maintenance cost. It uses focused data. It requires less infrastructure. It provides quick analytical results.	CO3	BTL2
13	**State the steps involved in designing a Data Mart. (Nov/Dec 2018)** Steps include identifying business requirements, selecting data sources, designing schema, defining ETL process, and implementing data access tools. These steps ensure efficient design.	CO3	BTL2
14	**Mention any two advantages of Data Marts. (Apr/May 2019)** Data marts provide faster query performance. They are easier to manage. They reduce system complexity. They support departmental decision-making.	CO3	BTL1
15	**What is meant by partitioning strategy? (Nov/Dec 2017)** Partitioning strategy divides large tables into smaller parts. It improves query performance. It enhances manageability. It supports efficient data access.	CO3	BTL1
16	**Define vertical partitioning. (Apr/May 2022)** Vertical partitioning divides a table by columns. Frequently accessed columns are grouped together. It improves performance. It reduces I/O cost.	CO3	BTL2
17	**What is normalization in partitioning? (Nov/Dec 2022)** Normalization organizes data to reduce redundancy. It improves data integrity. It divides tables logically. It supports efficient storage management.	CO3	BTL1
18	**Define row splitting. (Apr/May 2023)** Row splitting divides a table into multiple tables based on row access patterns. Frequently used rows are separated. It improves query efficiency.	CO3	BTL2
19	**What is horizontal partitioning? (Nov/Dec 2021)** Horizontal partitioning divides a table by rows. Each partition contains a subset of rows. It improves performance for large datasets. It supports parallel processing.	CO3	BTL1
20	**Mention any two advantages of horizontal partitioning. (Apr/May 2018)** It improves query performance. It enables efficient data management. It supports scalability. It reduces response time.	CO3	BTL1

S.No	PART A	CO	BTL
21	**Differentiate vertical and horizontal partitioning. (Nov/Dec 2020)** Vertical partitioning divides tables by columns. Horizontal partitioning divides tables by rows. Both improve performance but differ in data division method.	CO3	BTL2
22	**What is the impact of partitioning on query performance? (Apr/May 2021)** Partitioning reduces data scanning. It improves response time. Queries access only relevant partitions. System efficiency is enhanced.	CO3	BTL2
23	**State one limitation of partitioning strategy. (Nov/Dec 2022)** Partitioning increases design complexity. Maintenance effort is high. Improper partitioning affects performance. Careful planning is required.	CO3	BTL1
24	**What is hybrid partitioning? (Apr/May 2019)** Hybrid partitioning combines vertical and horizontal partitioning. It improves flexibility. It enhances performance. It is suitable for large data warehouses.	CO3	BTL2
25	**Give one advantage of Data Mart over Data Warehouse. (Nov/Dec 2023)** Data marts provide faster access to specific data. They are cost-effective. They are easier to manage. They support focused analysis.	CO3	BTL3

S. No	PART B	CO	BTL
1	Explain Metadata and its categories in detail.	CO3	BTL4
2	Describe the role of Metadata Repository in a Data Warehouse.	CO3	BTL4
3	Discuss the challenges involved in Metadata Management.	CO3	BTL4
4	Explain Data Mart and its types with suitable examples.	CO3	BTL4
5	Explain the need for Data Marts in decision support systems.	CO3	BTL4
6	Describe the steps involved in designing a Data Mart.	CO3	BTL4
7	Explain partitioning strategy in Data Warehousing.	CO3	BTL4
8	Differentiate Vertical and Horizontal Partitioning with examples.	CO3	BTL4

S. No	PART B	CO	BTL
9	Explain normalization and row splitting in partitioning strategy.	CO3	BTL4
10	Discuss the advantages and limitations of partitioning strategies.	CO3	BTL5

UNIT IV DIMENSIONAL MODELING AND SCHEMA

Dimensional Modeling- Multi-Dimensional Data Modeling – Data Cube- Star Schema- Snowflake schema- Star Vs Snowflake schema- Fact constellation Schema- Schema Definition - Process Architecture- Types of Data Base Parallelism – Data warehouse Tools.

S.No	PART A	CO	BTL
1	**Define Dimensional Modeling. (Nov/Dec 2019)** Dimensional modeling is a data design technique used in data warehouses to support analytical processing. It organizes data into facts and dimensions. Facts represent measurable business data, while dimensions provide context. This model improves query performance. It is easy to understand for end users. Dimensional modeling is widely used in OLAP systems.	CO4	BTL1
2	**What is multi-dimensional data modeling? (Apr/May 2018)** Multi-dimensional data modeling represents data in multiple dimensions such as time, product, and location. It allows users to analyze data from different perspectives. Data is stored in cubes. It supports fast aggregation. This model improves decision-making. It is commonly used in OLAP applications.	CO4	BTL1
3	**Define a Data Cube. (Nov/Dec 2017)** A data cube is a multi-dimensional structure used to store and analyze data. It consists of dimensions and measures. Each cell contains aggregated data	CO4	BTL1

S.No	PART A	CO	BTL
	values. Data cubes support OLAP operations. They improve query efficiency. Data cubes help in trend analysis.		
4	**What are the components of a data cube? (Apr/May 2019)** The components of a data cube are dimensions, measures, and attributes. Dimensions describe perspectives of analysis. Measures represent numerical values. Attributes describe dimension details. These components enable efficient analysis. They support business intelligence.	CO4	BTL1
5	**Define Star Schema. (Nov/Dec 2018)** Star schema is a dimensional database design. It consists of a central fact table connected to dimension tables. The structure resembles a star. It is simple to design and understand. Query performance is high. It is widely used in data warehouses.	CO4	BTL1
6	**What is a Snowflake Schema? (Apr/May 2017)** Snowflake schema is a variation of star schema. Dimension tables are normalized. It reduces data redundancy. The structure is complex. Query performance is slightly lower. It saves storage space.	CO4	BTL1
7	**Differentiate Star and Snowflake Schema. (Nov/Dec 2016)** Star schema uses denormalized dimension tables. Snowflake schema uses normalized dimension tables. Star schema is simple and faster. Snowflake schema is complex. Star schema uses more storage. Snowflake schema saves space.	CO4	BTL2
8	**What is a Fact Table? (Apr/May 2016)** A fact table stores quantitative business data. It contains foreign keys to dimension tables. Measures such as sales and revenue are stored. It supports aggregation. Fact tables are central in dimensional models. They enable analytical queries.	CO4	BTL1
9	**Define Dimension Table. (Nov/Dec 2020)** A dimension table stores descriptive attributes. It provides context to facts. Examples include customer and product tables. It supports filtering and grouping. Dimension tables improve readability. They enhance analysis.	CO4	BTL1
10	**What is a Fact Constellation Schema? (Apr/May 2021)**	CO4	BTL2

S.No	PART A	CO	BTL
	Fact constellation schema contains multiple fact tables. These share commor dimension tables. It supports complex business processes. It is also calle galaxy schema. It improves flexibility. It handles multiple subjects.		
11	**Define Schema Definition. (Nov/Dec 2019)** Schema definition describes the logical structure of a database. It includes tables, relationships, and constraints. It ensures data consistency. Schema definition supports query execution. It helps database design. It is essential in data warehousing.	CO4	BTL1
12	**What is process architecture in data warehousing? (Apr/May 2020)** Process architecture defines data flow in a warehouse. It includes ETL, storage, and access processes. It ensures smooth operation. It supports scalability. Proper architecture improves performance. It enhances reliability.	CO4	BTL2
13	**Define database parallelism. (Nov/Dec 2018)** Database parallelism executes multiple operations simultaneously. It improves performance. Large queries are divided into tasks. Tasks run in parallel. It reduces response time. It supports large datasets.	CO4	BTL1
14	**List the types of database parallelism. (Apr/May 2019)** The types include inter-query, intra-query, and inter-operation parallelism. Inter-query runs multiple queries. Intra-query divides a single query. Inter-operation executes operations simultaneously. These types improve efficiency. They support scalability.	CO4	BTL1
15	**What is inter-query parallelism? (Nov/Dec 2017)** Inter-query parallelism executes multiple queries at the same time. Each query uses separate resources. It improves throughput. It is useful in multi-user environments. It enhances performance. It supports concurrency.	CO4	BTL2
16	**What is intra-query parallelism? (Apr/May 2022)** Intra-query parallelism divides a single query into subtasks. Subtasks run in parallel. It reduces query execution time. It is useful for complex queries. It improves system efficiency. It supports large data processing.	CO4	BTL2
17	**Define inter-operation parallelism. (Nov/Dec 2022)** Inter-operation parallelism executes different operations simultaneously. Operations like join and sort run in parallel. It improves performance. It	CO4	BTL2

S.No	PART A	CO	BTL
	utilizes resources efficiently. It supports complex queries. It reduces execution time.		
18	**What are Data Warehouse Tools? (Apr/May 2023)** Data warehouse tools support data integration and analysis. They include ETL, OLAP, and reporting tools. These tools automate processes. They improve efficiency. They support decision making. They enhance business intelligence.	CO4	BTL1
19	**Mention any two ETL tools. (Nov/Dec 2021)** Informatica and Talend are popular ETL tools. They extract data from sources. They transform data formats. They load data into warehouses. They support automation. They improve data quality.	CO4	BTL1
20	**What is the role of OLAP tools? (Apr/May 2018)** OLAP tools support analytical queries. They provide multi-dimensional views. They enable slice and dice operations. They improve query performance. They support decision support systems. They enhance reporting.	CO4	BTL2
21	**Define reporting tools in data warehousing. (Nov/Dec 2020)** Reporting tools generate reports and dashboards. They present data visually. They support ad-hoc queries. They help managers analyze trends. They improve decision making. They enhance usability.	CO4	BTL1
22	**What is the advantage of Star Schema? (Apr/May 2021)** Star schema is easy to understand. Query performance is high. It requires fewer joins. It supports fast aggregation. It is user-friendly. It is widely adopted.	CO4	BTL2
23	**Mention one limitation of Snowflake Schema. (Nov/Dec 2022)** Snowflake schema is complex. Query performance is slower. It requires multiple joins. Design is difficult. Maintenance is high. User understanding is reduced.	CO4	BTL1
24	**What is scalability in dimensional modeling? (Apr/May 2019)**	CO4	BTL2

S.No	PART A	CO	BTL
	Scalability refers to handling data growth. Dimensional models support expansion. New dimensions can be added. Performance remains stable. It supports business growth. It ensures long-term usability.		
25	**Give one advantage of fact constellation schema. (Nov/Dec 2023)** Fact constellation supports multiple business processes. It shares dimensions across facts. It improves flexibility. It handles complex analytics. It reduces redundancy. It supports enterprise-level analysis.	CO4	BTL3

S.No	PART B	CO	BTL
1	Explain the need for Dimensional Modeling in Data Warehousing. (Nov/Dec 2019)	CO4	BTL2
2	Illustrate Multi-Dimensional Data Modeling using a Data Cube with example. (Apr/May 2018)	CO4	BTL3
3	Apply Data Cube operations to analyze sales data across dimensions. (Nov/Dec 2020)	CO4	BTL3
4	Analyze the structure of Star Schema and explain how it improves query performance. (Apr/May 2019)	CO4	BTL4
5	Compare Star Schema and Snowflake Schema with respect to performance and storage. (Nov/Dec 2018)	CO4	BTL4
6	Examine the suitability of Fact Constellation Schema for enterprise data warehouses. (Apr/May 2021)	CO4	BTL4
7	Evaluate different types of Database Parallelism used in data warehouses. (Nov/Dec 2017)	CO4	BTL5
8	Assess the role of Process Architecture in achieving scalability in data warehouses. (Apr/May 2022)	CO4	BTL5

S.No	PART B	CO	BTL
9	Design an appropriate dimensional model for a retail business scenario. (Nov/Dec 2022)	CO4	BTL6
10	Justify the selection of suitable Data Warehouse tools for a large-scale analytical system. (Nov/Dec 2023)	CO4	BTL6

UNIT V SYSTEM & PROCESS MANAGERS

Data Warehousing System Managers: System Configuration Manager- System Scheduling Manager System Event Manager - System Database Manager - System Backup Recovery Manager - Data Warehousing Process Managers: Load Manager – Warehouse Manager- Query Manager – Tuning – Testing

S.No	PART A	CO	BTL
1	**Define System Configuration Manager. (Nov/Dec 2019)** The System Configuration Manager manages the configuration of hardware, software, and network resources in a data warehouse. It ensures all components work together efficiently. Proper configuration improves system stability and performance. It supports scalability and smooth operation. It plays a vital role in warehouse reliability.	CO5	BTL1
2	**What is the function of System Scheduling Manager? (Apr/May 2018)** The System Scheduling Manager schedules warehouse tasks such as ETL	CO5	BTL1

S.No	PART A	CO	BTL
	jobs and backups. It ensures tasks are executed at predefined times. Resource conflicts are avoided. It improves system utilization. It supports automation of warehouse processes.		
3	**Explain the role of System Event Manager. (Nov/Dec 2017)** The System Event Manager monitors events in the data warehouse. It detects failures and abnormal activities. Alerts are generated for administrators. It supports quick corrective action. This improves system availability and reliability.	CO5	BTL2
4	**Define System Database Manager. (Apr/May 2019)** The System Database Manager controls database storage and access paths. It manages indexes and data placement. It ensures data integrity and consistency. It improves query execution. It supports performance optimization.	CO5	BTL1
5	**What is System Backup and Recovery Manager? (Nov/Dec 2018)** The Backup and Recovery Manager protects warehouse data. It performs regular backups. It restores data during failures. It minimizes data loss. It ensures business continuity.	CO5	BTL2
6	**Define Load Manager. (Apr/May 2020)** The Load Manager controls data extraction and loading. It applies transformation rules. It filters incorrect data. It ensures accurate data loading. It supports periodic data refresh.	CO5	BTL1
7	**What is Warehouse Manager? (Nov/Dec 2020)** The Warehouse Manager manages data after loading. It performs aggregation and indexing. It ensures data consistency. It monitors storage usage. It improves query performance.	CO5	BTL1
8	**Define Query Manager. (Apr/May 2021)** The Query Manager manages user queries. It converts queries into optimized execution plans. It schedules queries efficiently. It balances system resources. It improves response time.	CO5	BTL1
9	**What is tuning in data warehousing? (Nov/Dec 2021)**	CO5	BTL2

S.No	PART A	CO	BTL
	Tuning improves system performance. It optimizes queries and indexes. Bottlenecks are identified. Resource utilization is improved. It reduces response time.		
10	**Why is testing required in data warehousing? (Apr/May 2022)** Testing ensures data accuracy and consistency. It validates ETL processes. Performance testing checks efficiency. Errors are identified early. It improves system reliability.	CO5	BTL2
11	**Differentiate Load Manager and Query Manager. (Nov/Dec 2016)** The Load Manager handles data loading activities. The Query Manager handles user queries. Load Manager works during refresh cycles. Query Manager works during analysis. Both support warehouse operations.	CO5	BTL3
12	**State the importance of System Scheduling Manager. (Apr/May 2017)** The Scheduling Manager automates task execution. It prevents resource conflicts. It ensures timely processing. It improves system efficiency. It supports large-scale warehouses.	CO5	BTL3
13	**Explain how System Event Manager supports fault tolerance. (Nov/Dec 2018)** The Event Manager detects failures. Alerts are generated. Immediate corrective action is taken. Downtime is reduced. System reliability is enhanced.	CO5	BTL3
14	**Analyze the role of Warehouse Manager in performance improvement. (Apr/May 2019)** The Warehouse Manager creates aggregates. It manages indexes. It organizes data efficiently. Query response time improves. Overall system performance increases.	CO5	BTL4
15	**Explain how Query Manager handles multiple users. (Nov/Dec 2019)** The Query Manager schedules multiple queries. It optimizes execution plans. It balances workloads. Resource contention is minimized. User experience improves.	CO5	BTL4
16	**Examine the need for tuning in large data warehouses. (Apr/May 2020)** Tuning handles large data volumes. It removes bottlenecks. Query speed improves. Resource usage is optimized. System scalability increases.	CO5	BTL4

S.No	PART A	CO	BTL
17	**Analyze the importance of backup strategies. (Nov/Dec 2020)** Backup strategies prevent data loss. They support disaster recovery. Downtime is reduced. Business continuity is ensured. Enterprise reliability improves.	CO5	BTL4
18	**Evaluate the effectiveness of System Configuration Manager. (Apr/May 2021)** Proper configuration improves stability. It supports scalability. Performance is optimized. Errors are reduced. System efficiency improves.	CO5	BTL5
19	**Assess the role of System Scheduling Manager in enterprise warehouses. (Nov/Dec 2021)** It automates complex tasks. Resource usage is optimized. Conflicts are avoided. Timely execution is ensured. Enterprise efficiency improves.	CO5	BTL5
20	**Evaluate the importance of testing in warehouse deployment. (Apr/May 2022)** Testing validates data correctness. Performance issues are detected. Scalability is tested. Risk is reduced. System reliability improves.	CO5	BTL5
21	**Design a workflow involving system managers. (Nov/Dec 2022)** Configuration sets environment. Scheduling runs tasks. Event Manager monitors failures. Database Manager manages storage. Backup Manager protects data.	CO5	BTL6
22	**Design a process flow for data warehousing process managers. (Apr/May 2023)** Load Manager loads data. Warehouse Manager aggregates data. Query Manager serves queries. Tuning improves performance. Testing ensures correctness.	CO5	BTL6
23	**Justify the need for Load Manager in ETL. (Nov/Dec 2023)** It ensures clean data loading. Errors are filtered. Consistency is maintained. Periodic refresh is supported. Accurate analysis is enabled.	CO5	BTL6
24	**Propose a tuning strategy for a high-volume warehouse. (Apr/May 2019)** Indexes are optimized. Queries are rewritten. Resources are balanced. Bottlenecks are removed. Performance improves.	CO5	BTL6

S.No	PART A	CO	BTL
25	**Suggest testing techniques for data warehouse validation. (Nov/Dec 2023)** ETL testing ensures accuracy. Performance testing checks speed. Scalability testing validates growth. Errors are minimized. Reliable analytics is achieved.	CO5	BTL6

S. No	PART B	CO	BTL
1	Explain the functions of Data Warehousing System Managers in detail.	CO5	BTL2
2	Describe the roles of System Configuration Manager and System Scheduling Manager.	CO5	BTL2
3	Explain how the System Event Manager ensures fault tolerance in a data warehouse.	CO5	BTL3
4	Discuss the importance of System Database Manager in maintaining warehouse performance.	CO5	BTL3
5	Analyze the role of Backup and Recovery Manager in enterprise data warehouses.	CO5	BTL4
6	Analyze the responsibilities of Load Manager, Warehouse Manager, and Query Manager.	CO5	BTL4
7	Differentiate Load Manager and Query Manager with suitable examples.	CO5	BTL4

S. No	PART B	CO	BTL
8	Examine the need for tuning in large-scale data warehouse environments.	CO5	BTL4
9	Analyze the importance of testing during data warehouse implementation.	CO5	BTL4
10	Evaluate the effectiveness of System Scheduling and Event Management in data warehouses.	CO5	BTL5
11	Assess the importance of backup and recovery strategies in data warehousing systems.	CO5	BTL5
12	Design a workflow showing interaction among system managers in a data warehouse.	CO5	BTL6
13	Propose an efficient management strategy for a high-volume data warehousing system.	CO5	BTL6