CIA MOCK QUESTIONS

Data Warehouse & Data Mining

Question 1:Differentiate between OLAP and OLTP.

Feature	OLAP	OLTP
Purpose	Used for data analysis and decision-making	Used for data processing and day-to-day operations
Data Type	Processes historical data	Processes real-time operational data
Database Structure	Multidimensional and aggregated	Relational and normalized
Query Types	Complex queries with aggregations and drill-down capabilities	Simple queries for data retrieval or modification
Performance	Designed for fast response times for complex queries	Designed for fast transaction processing
User Interface	Graphical user interface with drill-down capabilities	Simple and straightforward user interface
Data Volume	Handles large amounts of data and supports analytical operations	Handles small to moderate amounts of data and supports transactional operations

☐ Stop generating

Question 2:What are the planning strategies involved in creating a data warehouse?

Creating a data warehouse involves careful planning to ensure that the data is organized, integrated, and structured in a way that supports effective decision-making. Here are some planning strategies involved in creating a data warehouse:

- **Define the goals and objectives:** The first step in creating a data warehouse is to define the goals and objectives of the project. This involves identifying the business requirements and the key performance indicators (KPIs) that the data warehouse will support.

- **Identify the data sources:** Once the goals and objectives are defined, the next step is to identify the data sources that will be used to populate the data warehouse. This includes identifying the internal and external sources of data, such as transactional systems, spreadsheets, and external data feeds.

- **Design the data model**: The data model is the conceptual representation of the data warehouse. It defines the structure of the data warehouse, including the tables, columns, and relationships

between them. The data model should be designed to support the business requirements and KPIs identified in the first step.

- **Plan the data integration:** Data integration involves the process of combining data from different sources and transforming it into a format that can be loaded into the data warehouse. This process can be complex and time-consuming, so it is important to plan it carefully. This includes identifying the data cleansing and transformation requirements and selecting the appropriate tools and technologies.

- **Define the data loading strategy**: The data loading strategy defines how the data will be loaded into the data warehouse. This includes selecting the appropriate ETL (Extract, Transform, Load) tools and defining the frequency and timing of the data loads.

- **Plan for data quality and governance:** Data quality and governance are critical to the success of a data warehouse project. This involves establishing data quality standards and policies, defining data ownership and stewardship, and implementing data governance processes.

- **Define the security and access controls**: Data warehouses contain sensitive and valuable data, so it is important to define the security and access controls to ensure that the data is protected from unauthorized access or misuse.

Question 3:Define data mining and   data warehousing

Data mining is the process of extracting useful and valuable insights and knowledge from large sets of data. It involves using statistical and machine learning techniques to analyze and identify patterns, trends, and relationships in the data. The goal of data mining is to uncover hidden patterns and insights that can be used to make better decisions, improve business operations, or gain a competitive advantage.

Data warehousing, on the other hand, is the process of collecting, storing, and organizing large volumes of data in a central repository for easy access and analysis. The data warehouse is typically designed to support business intelligence and decision-making processes, and is optimized for fast querying and reporting. The data in a data warehouse is structured and organized in a way that supports easy analysis and reporting.

Question 4:What is the difference between a data warehouse and data mining?

Feature	Data Warehouse	Data Mining
Purpose	Stores and organizes large volumes of structured data for easy access and analysis	Analyzes data to discover insights, patterns, and relationships
Data type	Historical data, generally structured and integrated	Historical and real-time data, structured and unstructured
Data volume	Large volumes of data	Large volumes of data
Data model	Designed to support easy querying and reporting	Uses statistical and machine learning techniques to analyze data and discover insights
Techniques	ETL (Extract, Transform, Load), OLAP (Online Analytical Processing)	Clustering, classification, regression, association rule mining, neural networks
Business use	Supports business intelligence, decision making and reporting	Helps identify trends, patterns, and relationships in data that can be used to inform business decisions
Query speed	Fast querying and reporting capabilities	Data mining can be time-consuming, but results can be used to inform future decisions
Output	Typically generates reports, dashboards, and visualizations	Generates insights and predictions that can inform business decisions
Data quality	Relies on high-quality, integrated data	Requires high-quality, reliable data for accurate analysis

Data warehouse is a centralized repository of structured data that is designed to support querying and reporting, while data mining is the process of using statistical and machine learning techniques to analyze data and extract insights, patterns, and relationships. The two concepts are complementary and are often used together to inform business decisions and improve operations.