Set (B) (Odd Roll Number)

Assignment 1

CAP446 Data Warehousing and Datamining

Name : Jayshri lal Pandit.
Roll No. : RD2112A103
Reg No. : 12111670
Section : D2112

Set (B)

(1)

(2) Why is an OLTP system considered economic as compared to OLAP?

Ans:- OLTP provides an immediate record of current business activity, while OLAP generates and validates insights from that data as it's compiled over time.

OLTP is operational, while OLAP is informational. The following comparison gives why OLTP system considered economic as compared to OLAP :-

Comparision on the basis	OLTP	OLAP
(1) Characteristic	Handles a large amount of small transactions	Handles large volume of data with complex queries
(2) Query types	Simple standardized queries	Complex queries
(3) Operations	Based on INSERT, UPDATE, DELETE commands	Based on SELECT commands to aggregate data for reporting
(4) Response Time	Milliseconds	Seconds, minutes, or hours depending on the amount of data to process

(5)	Design	Industry-specific, such as retail, manufacturing, or banking	Subject-specific, such as Sales, inventory, or marketing
(6)	Source	Transactions	Aggregated data from transactions
(7)	Purpose	Control and run essential business operations in real time	plan, solve problems, support decisions, discover hidden insights.
(8)	Data updates	Short, fast updates initiated by user	Data periodically refreshed with scheduled, long-running batch jobs
(9)	Space requirements	Generally small if historical data is archived	Generally large due to aggregating large datasets
(10)	Backup and recovery	Regular backups required to ensure business continuity and meet legal and governance requirements	Lost data can be reloaded from OLTP database as needed in lieu of regular backups
(11)	Productivity	Increase productivity	Increase productivity

	of end users	of bussiness managers, data analysts, and executives
(12) Data view	Lists day-to-day business transactions	Multi-dimensional view of enterprise data
(13) User examples	Customer-facing personnel, clerks, online shopers	Knowledge workers such as data analysts, bussiness analysts, and executive.
(14) Data Base Design	Normalised databases for efficiency	Denormalized databases for analysis.

(4) Data mining provides various functionalities? Explain each with example.

Ans:- Data mining provides following various functionalities each explained with examples:—

(1) Class/concept Description: Characterization and Discrimination

Data is associated with classes or concepts so they can be correlated with results. For example, new iphone model is released on three

varients to attend to the targeted customess based on their requirements like pro, Promax, and Plus.

(2) Classification :- It uses data models to predict the trends in data. For example, the spending chart ~~for~~ our internet banking or mobile application shows based on our spend patterns. This is sometimes used to define our risk of getting a new loan.

(3) Prediction :- Prediction finds the missing numeric values in the data. It uses regression analysis to find the unavailable data. Prediction is popular because of its importance in business intelligence. There are two ways one can predict data :-

(i) Predicting the unavailable or missing data using prediction analysis.
(ii) Predicting the class label using the previously built class model.

(4) Association Analysis :- It relates two or more attributes of the data. it discovers the relationship between the data and ~~use~~ rules

that are binding them. for examples if mobile phones are bought with headphones : support is 2%. and confidence is 40%. This means that 2% of the time that customers bought mobile phones with headphones. 40% of confidence is the probability of the same association happening again.

(5) Cluster Analysis :- ~~Unsupervised~~ Unsupervised classification is called cluster analysis. It is similar to the classification where the data are grouped. Unlike classification, in cluster analysis, the class label is unknown. Data are grouped based on clustering algorithms.

(6) Outlier Analysis :- When data that cannot be grouped in any of the class appears, we use outlier analysis. There will be occurrences of data that will have different attributes to any of the other classes or general models. These outstanding data are ~~the~~ called outlier.

(7) Evolution & Deviation Analysis :- With evolution analysis, we get time-related clustering of data. We can find trends and changes in

behaviour over a period. We can find trends and changes in the features like time-series data, periodicity, and similarity in trends with such distinct analysis.

(5) Explain any 3 types of data repositories stating relevant examples.

Ans:- Any 3 type of data repositories are explained below with examples :-

(1) Data warehouse :-

A data warehouse is large central data repository that brings together data from several sources or business segments. The stored data generally used for reporting and analysis to help users make critical business decisions.

The main objective of a data warehousing is to establish a connection between data from current systems.

For example, Product catalog data stored in one system and procurement orders for a client

stored in another one.

(2) Data Mart:-

A data mart is a subject-oriented data respository that's often a segregated section of a Data warehouse. It holds a subset of data usually aligned with a specific business department, such as Marketing, finance, or support.

Due to its similar size, a data mart can fast-track business procedures as you can easily access relevant data within days instead of months. As it only includes the data relevant to a specific area, a data mart is an economical way to acquire actionable insights swiftly.

(3) Metadata Repositories :-

Metadata incorporates information about the structures that include the actual data. Metadata repositories contain information about the data model that store and share this data. They describe where the source of data is, how it was

collected, and what it signified. It may define the arrangement of any data or subject deposited in any format.

For businesses, metadata repositories are essential in helping people understand administrative changes, as they contain detailed information about the data.

(1) Justify the statement "frequent patterns" are used to increase the sales of Organization ?

Ans:-

(3) What is KDD process ? State each step with an example.

Ans:→ KDD stands for Knowledge discovery of Databases. It refers to the broad procedure of discovering knowledge in data. KDD process is to extract information from data in the context of large databases. Data mining is the root of KDD process. it extracting the ~~data~~ knowledge from data, analyze the data and predict the data.

It involves the list of steps in the knowledge discovery process :-

- Data cleaning :- In this step, the noise and inconsistent data is removed.
- Data Integration :- In this step, multiple data sources are combined.
- Data Selection :- In this step, data relevant to the analysis task are retrived from the database.
- Data Transformation :- In this step, data is transformed or consolidated into forms appropriate

for mining by performing summary or aggregation operations.

- **Data Mining :-** In this step, intelligent methods are applied in order to extract data patterns.
- **Pattern Evaluation :-** In this step data patterns are evaluated
- **Knowledge Presentation :-** In this step, knowledge is represented.