

Total No. of Questions : 5]

SEAT No. :

P6989

[5865] - 302

[Total No. of Pages : 5

M.C.A. (Management)

IT 32: DATA WAREHOUSING AND DATA MINING

(2020 Pattern) (Semester - III)

Time : 2½ Hours]

[Max. Marks : 50

Instructions to the candidates:

- 1) All questions are compulsory.
- 2) Draw neat & labelled diagram wherever necessary.

Q1) Answer the following multiple choice questions. [20×½=10]

- i) Find the wrong statement of the K-means clustering
 - a) K-means clustering is a method of vector quantization.
 - b) K-means neighbour is same as K-nearest.
 - c) K-means clustering aims to partition 'n' observations into K clusters.
 - d) K-means clustering produces the final estimate of cluster centroids.
- ii) How many tier data warehouse architecture?
 - a) 2
 - b) 1
 - c) 3
 - d) 4
- iii) _____ is an intermediate storage area used for data processing in ETL process of data warehousing.
 - a) Buffer
 - b) Virtual memory
 - c) Staging area
 - d) Inter storage area
- iv) _____ is a good alternative to the star schema.
 - a) Star schema
 - b) Snowflake schema
 - c) Fact constellation
 - d) Star-snowflake schema
- v) In a snowflake schema which of the following types of tables are considered?
 - a) Fact
 - b) Dimension
 - c) Both fact and dimension
 - d) None of the mentioned

P.T.O.

- vi) The role of ETL is to _____.
 a) Find erroneous data
 b) Fix erroneous data
 c) Both finding and fixing erroneous data
 d) Filtering of the data source
- vii) _____ is a data transformation process.
 a) Comparison b) Projection
 c) Selection d) Filtering
- viii) OLTP stands for _____.
 a) Online Transaction Protocol
 b) Online Transaction Processing
 c) Online Terminal Protocol
 d) Online Terminal Processing
- ix) An approach in which the aggregated totals are stored in a multidimensional database while the detailed data is stored in the relational database is a _____.
 a) MOLAP b) ROLAP
 c) HOLAP d) OLAP
- x) Summary of data from an OLAP can be presented in _____.
 a) Normalization b) Primary keys
 c) Pivot Table d) Foreign keys
- xi) Efficiency and scalability of data mining algorithm is related to _____.
 a) Mining methodology b) User interaction
 c) Diverse data types d) None of the mentioned
- xii) Strategic value of data mining is _____.
 a) Cost sensitive b) Work sensitive
 c) Time-sensitive d) Technical-sensitive
- xiii) If the ETL process fetches the data separately from the host server during the automatic load to the data warehouse, one of the challenge involved is _____.
 a) the associated network may be down
 b) it may end up pulling the incomplete / incorrect file.
 c) it may end up connecting to an incorrect host server.
 d) None

- xiv) Webmining helps to improve the power of web search engine by identifying _____
- a) Web pages and classifying the web documents
 - b) XML documents
 - c) Text documents
 - d) Database
- xv) _____ is achieved by splitting the text into white spaces.
- a) Text cleanup
 - b) Tokenization
 - c) Speech tagging
 - d) Text transformation
- xvi) Assigning data to one of the clusters and recompute the centroid are the steps in which algorithm.
- a) Apriori algorithm
 - b) Bayesian classification
 - c) FP tree algorithm
 - d) K-mean
- xvii) A disadvantage of KNN algorithm is, it takes _____
- a) More time for training
 - b) More time for testing
 - c) Equal time for training
 - d) Equal time for testing
- xviii) Which is not a characteristics of Data warehouse?
- a) Volatile
 - b) Subject oriented
 - c) Non volatile
 - d) Time variant
- xix) What is the first stage of Kimball Life Cycle diagram.
- a) Requirement Definition
 - b) Dimensional Modelling
 - c) ETL Design Development
 - d) Maintenance
- xx) A connected region of a multidimensional space with a comparatively high density objects.
- a) Clustering
 - b) Association
 - c) Classification
 - d) Subset

Q2) a) What is a Data warehouse. Explain the need and characteristics of Data warehouse. [5]

b) Explain the schemas of Data warehouse. [5]

OR

a) Explain Kimball Life Cycle diagram in detail. [5]

b) What is a Data warehouse? Explain the properties of Data warehouse architecture. [5]

Q3) a) What is ETL? Explain data preprocessing techniques in detail. [6]

b) What is OLAP? Describe the characteristics of OLAP. [4]

OR

a) Describe ETL. What are the tasks to be performed during data transformation. [6]

b) What are the basic operations of OLAP? [4]

Q4) a) What is Data mining? Explain the architecture of Data mining. [3]

b) Apply FP Tree Algorithm to construct FP Tree and find frequent itemset for the following dataset given below (minimum support = 30%) [7]

Transaction ID	List of Products
1	Apple, Berries, Coconut
2	Berries, Coconut, Dates
3	Coconut, Dates
4	Berries, Dates
5	Apple, Coconut
6	Apple, Coconut, Dates

OR

- a) Explain data mining techniques in brief. [3]
- b) How does the KNN algorithm works? [7]

Apply KNN classification algorithm for the given dataset and predict the class for $X(P_1 = 3, P_2 = 7)$ ($K = 3$)

P_1	P_2	Class
7	7	False
7	4	False
3	4	True
1	4	True

- Q5) a) What is text mining? Explain the process of text mining. [4]
- b) Explain K-means algorithm. Apply K-means algorithm for group of visitors to a website into two groups using their age as follows:

15, 16, 19, 20, 21, 28, 35, 40, 42, 44, 60, 65

(Consider initial centroid 16 and 28 of two groups) [6]

OR

- a) Apply K-means algorithm for the given data set where K is the cluster number $D = \{2, 3, 4, 10, 11, 12, 20, 25, 30\}$, $K = 2$. [6]
- b) What are the different types of web mining? [4]

