

These are sample MCQs to indicate pattern, may or may not appear in examination

University of Mumbai
Online Examination 2020

Program: BE Computer Engineering

Curriculum Scheme: Revised 2012

Examination: Final Year Semester VIII

Course Code: CPC801 and Course Name: Data Warehousing and Mining

Time: 1hour

Max. Marks: 50

Note to the students:- All the Questions are compulsory and carry equal marks .

Q1. Data warehouse is not meant ____

Option A: To provide strategic information

Option B: To provide structured data

Option C: To perform day to day business

Option D: To help in making strategic decisions

Q2. What is data mining?

Option A: Time variant non volatile collection of data

Option B: The actual discovery phase of a knowledge

Option C: The stage of selecting the right data

Option D: Dimensionality reduction

Q3. Changes in data types and length of the field comes under which type of data transformation

Option A: Calculated and derived values

Option B: Date/time conversion

Option C: Format revision

Option D: Merging of information

Q4. Which is not a constituent of information package diagram?

Option A: Dimensions

Option B: Dimension hierarchies

Option C: Facts

Option D: Metadata

Q5. Which of the following is a correct statement?

Option A: Fact table is wide and contains facts.

Option B: Dimension table is deep and contains dimension hierarchies.

Option C: Fact table is deep and contains facts.

Option D: Dimension table is wide and contains facts.

Q6. What are maximal frequent itemsets?

Option A: A frequent itemset whose no super-itemset is frequent

Option B: A frequent itemset whose super-itemset is also frequent

Option C: A non-frequent itemset whose super-itemset is frequent

Option D: A Non- frequent itemset whose no super-itemset is frequent

Q7. K from K means method indicates

Option A: Number of partitions

Option B: Number of elements to be clustered

Option C: Number of labelled data

Option D: Number of unlabelled data

Q8. In decision tree, node represents

Option A: Name of attribute on which data is partitioned

Option B: Attribute value on which data is aggregated

Option C: Class label

Option D: Name of attribute with highest entropy among all

Q9. What is a correct statement regarding ROLAP?

Option A: Data is summarized at predefined levels, so it can not provide detailed Data

Option B: Data is stored at fine granularity level, so it fails in providing summarized data

Option C: Data is summarized and presented dynamically
Option D: It requires less space for storage compared to MOLAP.

Q10. When data warehouse is built in bottom up approach
Option A: It involves high risk of failure
Option B: Needs cross functional skills
Option C: It involves low risk of failure
Option D: Data mart becomes larger in size than data warehouse

Q11. Preserving the periodic nature of the changes in the data warehouse is possible due to ____ data loading type
Option A: Incremental loads
Option B: Initial load
Option C: Full refresh
Option D: Partial load

Q12. Multidimensional databases use _____ formats to store presummarized cube structures.
Option A: SQL
Option B: Proprietary file
Option C: Object oriented
Option D: Non proprietary file

Q13. Which of the following method assigns weight to vote of classifying model?
Option A: Holdout
Option B: Cross validation
Option C: Bagging
Option D: Boosting

Q14. _____ Can find arbitrarily shaped clusters.
Option A: Partitioning methods
Option B: Hierarchical methods
Option C: Density-based methods

Option D: Grid-based methods

Q15. Which of the following is direct application of frequent itemset mining?

Option A: Social Network Analysis

Option B: Market Basket Analysis

Option C: Outlier Detection

Option D: Intrusion Detection

Q16. Smooth by bin mean on data { 18, 19,23} results in

Option A: { 20,20,20}

Option B: {18,18,23}

Option C: {18,23,23}

Option D: {19,19,19}

Q17. Which of the following is not data preprocessing technique?

Option A: Data cleaning

Option B: Data integration

Option C: Data reduction

Option D: Data generation

Q18. What is not true about FP growth algorithms?

Option A: It mines frequent itemsets without number of candidate generation.

Option B: There are chances that FP trees may not fit in the memory

Option C: FP trees are very expensive to build

Option D: It expands the original database to build FP trees.

Q19. This method is also called as aggregate bootstrap method.

Option A: Cross validation

Option B: Bagging

Option C: Boosting

Option D: Holdout

Q20. With which technique, two clusters are merged if the maximum distance of connecting edge is less than equal to threshold distance?

Option A: Single link technique

Option B: Complete link technique

Option C: Average link technique

Option D: Divisive technique

Q21. Which of the below is not used to improve the efficiency of apriori algorithm

Option A: Transaction reduction

Option B: Partitioning

Option C: Hash-Based technique

Option D: Tree pruning

Q22. While loading data to data warehouse, if we do not want to preserve existing data, then which of the below mode we should use while loading data?

Option A: Load

Option B: Append

Option C: Destructive merge

Option D: Constructive Merge

Q23. What is the type of metadata which informs about the processing of data ?

Option A: Extraction transformation and loading metadata

Option B: Opearational metadata

Option C: End user metadata

Option D: Data dictionary

Q24. Which of the following is an incorrect statement??

Option A: If primary key from operational system is used, then it may combine different records with same primary key in case of reuse of primary key.-

Option B: If built in meaning key is used as primary key, then on updation of attribute value contributing in primary key, the primary key also get changed. Two records corresponding to different primary keys are treated separately , even though entity is same.

Option C: Surrogate key can be used to set primary key for dimension table.

Option D:

Fact table refers dimension table by using foreign key of dimension table as primary key.

Q25. Which of the following feature shows that data warehouse stores data permanently

Option A: Data granularity

Option B: Time variant

Option C: Non volatile

Option D: Integrated

