Exam

Module: Business Intelligence and Information Technology
Course: Databases and Business Intelligence

Lecturers: dr. C. Amrit
University of Twente

Name:_____

Student nr:_____

Open Book Exam
This is an open book exam: All course materials (slides handouts, books, and papers) may be used for reference.

Grading

This exam contains 25 multiple-choice questions (from the lectures, book chapters and from the papers that have been included in the course) and 1 optional question (for 4 points).
Each of the following questions is for 2 points.

1. In a relational model a tuple is equivalent to a:

    A.   record
    B.   field
    C.   file
    D.   data base
    E.   data item

Answer: A

2. Which command is used to remove an index from the database in SQL?

    A.   DELETE INDEX
    B.   DROP INDEX
    C.   REMOVE INDEX
    D.   ROLL BACK INDEX
    E.   None of the above

Answer: B

3. What key distinctly identifies a record in a database table?

    A)   Main key
    B)   Primary key
    C)   Principal key

D) Super key

E) All of the above

Answer: B.

StudId	Name	Address	Dept	CrsCode	CrsName	Descr	Sem
s123	Alex	7522KL	CS	GB	Databases	SQL…	F2004
s123	Alex	7522KL	AM	DW	DiscrMath	1+1=2	S2003
s456	Chris	7513BE	CS	GB	Databases	SQL…	F2005

Table 1: Student, Course Dataset

4. In Table 1, the data is in:

A. 2nd Normal form

B. 1st Normal form

C. 3rd Normal form

D. 4th Normal form

E. None of the above

Answer B

5. Upon normalization to the next normal form, Table 1 will be split into:

A. 2 Tables: Student's personal information and Student course information

B. 2 Tables: Student's personal information and Course Information

C. 3 Tables : Student's personal information, Course Information and Student department information

D. 3 Tables : Student's personal information, Course Information and Student department information

E. None of the above

Answer A

6. Solving the problem of having data replicated in different software applications spread over the different departments of an organization, is the main advantage of a

A. Workflow system

B. Database system

C. ERP system

D. Datawarehouse
E. None of the above

Answer: C

7. An example of a supervised learning algorithm is

   A. Statistical analysis
   B. Neural network
   C. Clustering techniques
   D. Naïve Bayesian algorithm
   E. None of the above

Answer: D

8. When the causal relation between the input and the output variables in complex, one would use:

   A. Statistical modeling
   B. Supervised learning
   C. Unsupervised learning
   D. All the above
   E. None of the above

Answer: C

9. The variable marital status can be categorized using the codes (1) single, (2) married, and (3) divorced. This is an example of a:

   A. Ordinal variable
   B. Nominal variable
   C. Interval variable
   D. Ratio variable
   E. None of the above

Answer B

A supermarket stores all the transactions in a large database. These transactions database can be used for "basket analysis". For the sake of simplicity and time we focus only on the following small part of the database and the items:

Transaction	Items
t1	{bread, butter, cheese, jelly, milk}
t2	{bread, cheese, jelly, milk}

t3	{butter, cheese, jelly}
t4	{bread, cheese, jelly, milk}
t5	{butter, milk, pizza}
t6	{bread, cheese, jelly, milk}
t7	{bread, jelly, milk}
t8	{butter, jelly, milk, pizza}
t9	{bread, pizza}
t10	{jelly, milk, pizza}
t11	{bread, cheese}
t12	{bread, butter, jelly, pizza}
t13	{bread, butter, cheese}
t14	{bread, cheese, jelly, milk}
t15	{bread, butter, cheese, jelly, milk}

Table 2 :Part of the transaction data base.

10. In Table 2, the support and confidence of the following association rules:

    {bread} => {milk}

    A. 7/11, 7/14
    B. 7/11,7/15
    C. 7/15,7/11
    D. 7, 7
    E. None of the above

    Answer C

11. The support and confidence for {milk} => {bread} is:
    A. 7, 7
    B. 7/10,7/15
    C. 7/15,7/10
    D. 6/16,6/10
    E. None of the above

Answer C

12. The support and confidence of {cheese}=>{milk, bread} is:
    A. 6,6
    B. 6/9,6/15

C. 6/15,6/9
D. 7/15,7/9
E. None of the above

Answer C

13. The frequent 1 item sets from Table 1 are:
    A. {Bread}, {Cheese}, {Jelly}, {Milk}
    B. {Bread}, {Cheese}, {Jelly}
    C. {Bread}, {Cheese}, {Jelly},{Milk}, {Butter}
    D. Bread}, {Cheese}, {Jelly},{Milk}, {Butter}, {Pizza}
    E. None of the above

Answer A.

14. The frequent 2 item sets from Table 1 are:
    A. {bread, cheese}, {bread, jelly}, {bread, milk}, {cheese, jelly}, {cheese, milk}, {jelly, milk}
    B. {bread, cheese}, {bread, jelly}, {bread, milk},
    C. {cheese, jelly}, {cheese, milk}, {jelly, milk}

    D. {bread, cheese}, {bread, jelly},{jelly, milk}

    E. {bread, milk}, {cheese, jelly}, {cheese, milk},
    Answer D.

15. The frequent 3 item sets from Table 1 are:
            A. {bread, cheese, jelly}, {bread, jelly, milk}, {bread, cheese, milk}
            B. {bread, cheese, jelly}, {bread, jelly, milk},
            C. {bread, jelly, milk}, {bread, cheese, milk}
            D. {bread, jelly, milk},
            E. None of the above

        Answer E.

16. _____ is essentially a hierarchy of if-then statements, mostly used to classify categorical or numerical data
                A. Neural network
                B. Decision Tree
                C. Clustering algorithm
                D. Apriori algorithm
                E. Association Rule Mining

        Answer B

17. The simple split methodology splits the data into:

    A. a training and a test set which have some overlap
    B. a training and a test set which are mutually exclusive

C. a training and a trial and a test set which have some overlap
D. a training and a trial and a test set which are mutually exclusive
E. None of the above

Answer B.

18) Data warehouse is a(n) _____, integrated, time-variant, nonvolatile collection of data in support of management's decision making process.
    A) analysis-oriented
    B) object-oriented
    C) subject-oriented
    D) model-oriented
    E) None of the above
Answer: C

19. A star schema contains a central _____ surrounded by several dimension tables.
    A) database
    B) fact table
    C) data tree
    D) data table
    E) None of the above

20. Which of the following is needed to determine how data are to be retrieved from a data warehouse, and will assist in the physical definition of the warehouse by helping to define which data require indexing?
    A) Indexing modeling
    B) Retrieval modeling
    C) Access modeling
    D) Tactic modeling
    E) None of the above
Answer: C

21. BPM encompasses a core set of processes, including financial and _____ planning, consolidation and reporting, modeling, analysis, and monitoring of KPIs, linked to organizational strategy.
A) strategic
B) tactical
C) operational
D) daily
E) None of the above

Answer: C

22. Fundamentally, an operational plan is like a project plan that is designed to ensure that an organization's _____ is realized.
    A) business policy
    B) plan
    C) strategy
    D) business rules

E) None of the above

Answer: C

23 A star schema contains a central _____ surrounded by several dimension tables.
   A) database
   B) fact table
   C) data tree
   D) data table
   E) None of the above

Answer: B

24. ETL process consists of extract, transform, and load. Transformation occurs by using _____ or lookup tables or by combining the data with other data.
   A) rules
   B) policies
   C) strategies
   D) procedures
   E) None of the above

Answer: A

25. Organizations are being compelled to capture, understand, and harness their data to support decision making in order to improve:
   A) business operations
   B) competition
   C) compliance
   D) risk
   E) None of the above

Answer: A

BONUS Question for 4 points

26.. *Data mining exercise.*
A car salesman is trying to optimize the effort he puts in with prospective customers. To help with this he has decided a decision tree based on data he has collected about when enquiries have resulted in a car purchase

Enquiry number	Car Type	Paintwork	Fuel	Car Purchased
E1	Sports	Matt	Diesel	No
E2	Sports	Metallic	Petrol	Yes
E3	Saloon	Matt	Diesel	No
E4	Saloon	Metallic	Hybrid	No
E5	Estate	Metallic	Petrol	Yes
E6	Estate	Metallic	Diesel	Yes

The best attribute for the root node of the decision tree is:

A. Paintwork

B. Fuel
C. Car Purchased
D. Enquiry Number
E. Car Type