Dept. of Computer Science & Engineering
Indian Institute of Technology Kharagpur

Roll no.: _____ Name: _____

CS43002 Database Management Systems Class Test 16.03.2018 Duration: 30 Minutes Full Marks: 20

Answer all questions. Write your answers in the box provided.

1. a-j are multiple choice type questions. For each question, zero or more options may be correct. 2 marks will be awarded for correct answer, 1 mark will be deducted for wrong answer. A question will be considered to be correctly answered if all the correct and only the correct options are chosen. If you think no option is correct for a question, write NONE. If you leave any answer blank, 5 marks will be deducted.

a. In a star schema, usually:
   (i) **the fact table is normalized**
   (ii) the dimension tables are normalized
   (iii) the fact table is de-normalized
   (iv) the dimension tables are de-normalized

b. In OLAP operations, roll-up can be achieved by:
   (i) moving up a dimension hierarchy like month → quarter, etc.
   (ii) moving down a dimension hierarchy like quarter → month, etc.
   (iii) adding a new dimension
   (iv) removing one or more dimensions

c. It is beneficial and practical to materialize all the views in a data cube when
   (i) the number of levels in dimensional hierarchies are very large and there are too many dimensions
   (ii) the speed of retrieval is the primary objective
   (iii) disk space saving is the primary objective
   (iv) we can implement a greedy algorithm for selecting the views to be materialized

d. Data warehouses
   (i) are built to support long term decision making of top-level managers
   (ii) are built to support short term decision making of junior managers
   (iii) usually store data from OLTP as well as from other sources like market data and competitors' data
   (iv) are usually built as a database separate from OLTP databases

e. The Slicing OLAP operation causes
   (i) a reduction in the number of dimensions from an OLAP cube
   (ii) an increase in the number of dimensions from an OLAP cube
   (iii) no change in the number of dimensions from an OLAP cube
   (iv) new facts to be added to an OLAP cube

f. In the context of Association Rule Mining (considering that the same minimum value of support is maintained at each level):
   (i) **If an itemset is frequent, then all its subsets are also frequent**
   (ii) If an itemset is not frequent, then none of its subsets can be frequent
   (iii) If an itemset is frequent, then all its supersets are also frequent
   (iv) If an itemset is not frequent, then none of its supersets can be frequent

g. In the context of Association Rule Mining: **NONE**
   (i) the support of an association rule is always greater than its confidence
   (ii) the support of an association rule is always less than its confidence
   (iii) the support of an association rule is always equal to its confidence
   (iv) the support of an association rule can never be equal to its confidence

h. The *a priori* approach to Association Rule Mining (i.e., the approach proposed by Agrawal and Srikant)
   (i) starts with 1-itemsets and progressively searches for larger itemsets that are frequent
   (ii) starts with the largest itemsets and progressively searches for smaller itemsets that are frequent
   (iii) uses pruning to prevent unnecessary trips to the database by removing itemsets from the candidate set that are surely not frequent
   (iv) will need as many passes through the database as the size of the largest frequent itemset

i. In the context of Association Rule (AR) Mining,
   (i) if there is a meaningful AR (i.e., a rule having support > min_support and confidence > min_confidence) of the form A→B, where A and B are itemsets, then B→A is also a meaningful AR.
   (ii) if an itemset of size L is found to be frequent, then the maximum number of possible meaningful association rules derived from it is also L
   (iii) if an itemset of size L is found to be not-frequent, then the minimum number of possible meaningful association rules derived from it is also L
   (iv) the upward and downward closure properties may not hold if we change the values of min-support at various levels (i.e., we vary it with the length of the itemset)

j. Results of Association Rule Mining can be used to determine
   (i) the layout of a store
   (ii) which items to be put on offer jointly
   (iii) which items are purchased jointly more often
   (iv) purchase of which item(s) causes purchase of which other item(s)

Dept. of Computer Science & Engineering
Indian Institute of Technology Kharagpur

Roll no.: _____ Name: _____

CS43002 Database Management Systems Class Test 16.03.2018 Duration: 30 Minutes Full Marks: 20

Answer all questions. Write your answers in the box provided.

1. a-j are multiple choice type questions. For each question, zero or more options may be correct. 2 marks will be awarded for correct answer, 1 mark will be deducted for wrong answer. A question will be considered to be correctly answered if all the correct and only the correct options are chosen. If you think no option is correct for a question, write NONE. If you leave any answer blank, 5 marks will be deducted.

a. In a star schema, usually:
   (i) **the fact table is normalized**
   (ii) the fact table is de-normalized
   (iii) **the dimension tables are de-normalized**
   (iv) the dimension tables are normalized

b. In OLAP operations, roll-up can be achieved by:
   (i) moving down a dimension hierarchy like quarter → month, etc.
   (ii) **moving up a dimension hierarchy like month → quarter, etc.**
   (iii) **removing one or more dimensions**
   (iv) adding a new dimension

c. It is not beneficial and practical to materialize all the views in a data cube when
   (i) the number of levels in dimensional hierarchies are very large
   (ii) **the speed of retrieval is not the primary objective**
   (iii) disk space saving is not the primary objective
   (iv) **there are too many dimensions**

d. Data warehouses
   (i) are built to support short term decision making of junior managers
   (ii) are built to support long term decision making of top-level managers
   (iii) are usually built as a database separate from OLTP databases
   (iv) usually store data from OLTP as well as from other sources like market data and competitors' data

e. The Slicing OLAP operation causes
   (i) a reduction in the number of dimensions from an OLAP cube
   (ii) no change in the number of dimensions from an OLAP cube
   (iii) new facts to be added to an OLAP cube
   (iv) an increase in the number of dimensions from an OLAP cube

f. In the context of Association Rule Mining (considering that the same minimum value of support is maintained at each level):
   (i) If an itemset is not frequent, then none of its supersets can be frequent
   (ii) If an itemset is frequent, then all its supersets are also frequent
   (iii) If an itemset is not frequent, then none of its subsets can be frequent
   (iv) If an itemset is frequent, then all its subsets are also frequent

g. In the context of Association Rule Mining: **NONE**
   (i) the support of an association rule can never be equal to its confidence
   (ii) the support of an association rule is always equal to its confidence
   (iii) the support of an association rule is always less than its confidence
   (iv) the support of an association rule is always greater than its confidence

h. The *a priori* approach to Association Rule Mining (i.e., the approach proposed by Agrawal and Srikant)
   (i) starts with the largest itemsets and progressively searches for smaller itemsets that are frequent
   (ii) starts with 1-itemsets and progressively searches for larger itemsets that are frequent
   (iii) will need as many passes through the database as the size of the largest frequent itemset
   (iv) uses pruning to prevent unnecessary trips to the database by removing itemsets from the candidate set that are surely not frequent

i. In the context of Association Rule (AR) Mining,
   (i) if there is a meaningful AR (i.e., a rule having support > min_support and confidence > min_confidence) of the form A→B, where A and B are itemsets, then B→A can never be a meaningful AR.
   (ii) if an itemset of size L is found to be frequent, then the maximum number of possible meaningful association rules derived from it must be greater than L
   (iii) the upward and downward closure properties may not hold if we change the values of min-support at various levels (i.e., we vary it with the length of the itemset)
   (iv) if an itemset of size L is found to be not-frequent, then the minimum number of possible meaningful association rules derived from it is also L

j. Results of Association Rule Mining cannot be used to determine **NONE**
   (i) the layout of a store
   (ii) which items to be put on offer jointly
   (iii) which items are purchased jointly more often
   (iv) purchase of which item(s) causes purchase of which other item(s)