Multiple Choice

1. In a $B^+$-tree, the last pointer in a non-leaf node points to (a) the next node (b) the subtree with values greater than or equal to the last key value (c) nothing; it is always null (d) the physical location of the row with the last key value

2. The SQL command to remove all changes made by a transaction is (a) undo (b) commit (c) rollback (d) exit

3. If item $i$ has not been rated, which of the following could a recommender system use to make a prediction for user $u$ on $i$? (a) similar items to $i$ (b) similar users $u$ (c) average ratings by $u$ (d) average ratings on $i$

4. ORM stands for (a) Object Relational Mapping (b) Object Retrieval Method (c) Original Relational Model (d) Ordered Retrieval Model

5. The relational algebra operator that corresponds to the SELECT clause in SQL is (a) $\Pi$ (b) $\sigma$ (c) $\rho$ (d) $\times$

6. If the three nearest neighbors had values 1,1,4, under constant regression, the prediction would be (a) 1 (b) 2 (c) 4 (d) 6.

7. Within a DBMS, deadlocks are used to (a) prevent disk failures (b) prevent atomicity failures (c) prevent concurrency errors (d) prevent CPU overloading

8. The computing paradigm used by Google in its cloud computing system is (a) BigTable (b) Google File System (c) Hibernate (d) MapReduce overloading
Short Answers

9. Describe a scenario where linear search would be better than indexed search. (4)

10. If auto-commit is turned on, describe an unfortunate situation that can occur during a transfer of funds between two accounts. (4)

11. Expand the acronym ACID. (4)
12. What are the four transaction isolation levels allowed under the SQL standard? (4)

13. Provide the formulas for support and confidence for association rules. (4)

14. Provide the formulas for precision and accuracy. (4)
15. What is the basic IR equation for determining the relevance of a document to a set of keywords?

16. Describe a LOOCV process to compute RMSE for recommender systems.
17. A data structure description for Big Table was provided in class. What was it? (4)

18. Provide the methods within a Java program needed to store a persistent object in a database using Hibernate. Pseudocode is acceptable. (4)
Problems

19. Assume a $B^+$-tree with four pointers in each node. Assume the following values are inserted (one at a time) into the tree – 2, 3, 5, 7, 11, 13, 17, 19, 23, 29. Show the final result of the tree.
20. Prove that $\sigma_{p_1 \lor p_2}(R) \equiv \sigma_{p_1}(R) \cup \sigma_{p_2}(R)$ (14)
21. Show the results of applying the k-means algorithm to the data 2, 3, 5, 7, 11, 13 with k = 3.