

COSC6340: Database Systems

Final Exam-Take Home

PS id:

Test score:

Instructions

There are two kinds of questions. For theory questions you are asked to rewrite the definition, general explanation or theoretical result, given in the textbook. You should *rewrite* it in your own words, but using the same notation. For example questions you should change the examples in the book as much as possible. At the minimum different data, but also different table, column names to make your answer more unique. You should avoid taking examples from other textbooks, especially if math notation is different. You can add and explain all required assumptions to support your example.

Important requirements: (1) every answer must be different for students in different teams. You can and are encouraged to solve the exam with your team partner. Therefore, some similarity of answers (but not equal answers!) from students in the same team is acceptable. However, you can work individually. (2) you cannot copy/paste information you find on any textbook, Wikipedia or Internet in general. If we detect plagiarism I will apply a significant penalty (50% or more). (3) Submit as a DOC file named final-lastName.doc (example "final-Chen.doc"). Use the old .DOC format to avoid compatibility problems. (4) E-mail to both TA and myself. Indicate if you worked alone or with your team partner.

Theory Questions

1. When is a table in 3NF, but not BCNF? Give a general definition based on FDs.
2. What is blocking factor and how does it related to I/O efficiency in a table scan and updating one record in transaction processing?
3. Which relational operators are more likely to be evaluated earlier and which are not? Discuss what happens if columns are indexed or not. Consider aggregations (SPJA queries).
4. Distinguish scanning, parsing, validation, plan generation and optimization phases in query evaluation.
5. Explain serial, serializability in terms of read/write operations.
6. Under concurrent processing explain when we can get: deadlocks, starvation, lost updates and inconsistency.
7. Compare advantages and disadvantages of 2PL and timestamping.
8. Explain External Merge Sort on n records using 2 buffers holding k records each.
9. Explain COMMIT and ROLLBACK in SQL. Explain implications for UNDO/REDO for the recovery manager.
10. Explain hash join and sort-merge join between two tables R, S with n rows each.

Example Questions

1. Give an example of a SQL query that produces a cartesian product.
2. Give an example of two lost updates among three transactions.
3. Show a RAID 10 example with $M = 8$ disks, chunk size four times block size.
4. Show C++ code to seek and write the i th record on a large binary file having records with two integer columns on table $T(i, h)$.
5. Show an example of a query where pushing σ is good. Show an example of a query where pushing π is bad.
6. Show an example of the wound-wait protocol and explain how deadlock is eliminated.
7. Show a deadlock example with 4 transactions, where the wait graph G has a 4-edge cycle.
8. Give an SQL query example where table-level lock is preferable. Give an SQL query example where row-level lock is preferable.
9. Show 3 different equivalent expressions for an SPJ query involving σ, π, \bowtie . Use relational algebra.
10. Show an example of an insertion into a B+-tree with 3 levels that produces 4 levels with $p = 2$ (max 2 keys per node, min 1 key per node).