

复旦大学软件学院

2022-2023 学年第一学期期末考试试卷

☒ A 卷 ☐ B 卷 ☐ C 卷

课程名称: 数据库设计 课程代码: SOFT130015

开课院系: 软件学院 考试形式: 开卷

姓名: _____ 学号: _____ 专业: _____

提示: 请同学们秉持诚实守信宗旨, 谨守考试纪律, 摒弃考试作弊。学生如有违反学校考试纪律的行为, 学校将按《复旦大学学生纪律处分条例》规定予以严肃处理。

题号	1	2	3	4	5	6	7	8	总分
得分									

Answer all questions on the Answer Sheets.

Questions marked with "[H]" are only for the students enrolled in the Honor Class.

Time allowed: 2 hours

I.(9 Points) Explain the following concepts in your own words.

(i)(3 Points) Clustering Index

(ii)(3 Points) Sequential File Organization

(iii)(3 Points) Recoverable Schedule (of transactions)

II.(18 Points) Answer the following questions.

(i)(6 Points) Please explain why B⁺-tree is an efficient storage structure for sequential file organization? And then tell the difference between B⁺-tree index and B⁺-tree file organization.

(ii)(6 Points) Tom is a database application developer and works on logical schemas. One night, Jerry, the database administrator, changed the physical organization of data storage when Tom was not working. The next morning when Tom was back to work, he was not aware of what Jerry had changed but continued working with his logical schemas. Please explain what is the supporting mechanism provided by the database management system.

(iii)(6 Points) Someone says that the data consistency requirement should always be satisfied in database transaction management. Weaker consistency levels produce inaccurate data so they are not practically useful in real applications. Please give at least one example of non-serializable transactions and show the usefulness of weaker consistency levels.

III.(10+4 Points) Query writing.

(i)(3+4 Points) Consider the following schemas: $R = (A, B, C)$ and $S = (D, E)$. $r(R)$ and $s(S)$ are relations.

Please write:

- (a) a statement in SQL,
- (b) [H] an expression in tuple relational calculus, and
- (c) [H] an expression in domain relational calculus,

so that all of them are equivalent to the following query:

$$\prod_{A,D,E} (\sigma_{C=D \wedge B=0 \wedge E>1000} (r \times s))$$

(ii)(3 Points) Please rewrite the above relational algebra expression with a theta join operator.

(iii)(4 Points) Given schemas $R = (\underline{A}, \underline{B}, \underline{C}, D)$ and $S = (\underline{C}, \underline{E})$, where the primary keys are underlined. $r(R)$ and $s(S)$ are relations. Please write a query in SQL, which is equivalent to the following expression.

$$\prod_{A,B,C} (r) \div \prod_C (\sigma_{E=0} (s))$$

IV. (22 Points) Answer all questions based on the given database.

It is a rental management system for a public library. The library typically possesses multiple copies of a book so that more than one customer may rent the book at the same time. The system manages all rental records. There are mainly four schemas, named *book*,

customer, *rent_record*, and *rent_record_detail*.
book(*book_id*, *book_type*, *title*, *date_publish*, *total_count*, *on_shelf_count*);
customer(*customer_id*, *customer_name*, *addr_country*, *addr_province*, *addr_city*,
addr_street, *addr_detail*, *phone*, *email*);
rent_record(*record_id*, *date_rent*, *status*, *customer_id*);
rent_record_detail(*record_id*, *book_id*, *date_return*, *book_status*)

The field *total_count* means the number of copies of a specific book possessed by the library, while the field *on_shelf_count* means the number of copies of the book that are left on the shelf (i.e., not being rented).

The primary keys are underlined. Foreign keys can be literally inferred.

- (i)(6 Points) Please write a **query in SQL** to calculate the correct value of *on_shelf_count* based on the *rent_record_detail*. (Hint: You may assume that the *book_status* is "IN_RENT" if a copy of the specific book is rented by a customer.)
- (ii)(10 Points) Please recover an appropriate **E-R diagram** based on the given relation schemas. Should you make any assumptions, please write them down as well.
- (iii)(6 Points) If the *on_shelf_count* is maintained by triggers, please tell **which triggers** should be deployed and then discuss the **benefits** and **drawbacks** of the use of these triggers. You do not need to write the implementation details of the triggers but only to describe the functions and triggering conditions of them.

V.(22 Points)

Given the schema $R = (A, B, C, D, E)$ and the following set F of functional dependencies for R .

$ABC \rightarrow DE$

$B \rightarrow C$

$AB \rightarrow CE$

$D \rightarrow A$

- (i)(2 Points) Is BD a candidate key of R ? How do you know that?
- (ii)(3 Points) What data anomalies might occur when *deleting a tuple* from a relation of R ? Please give one example.
- (iii)(4 Points) Please compute a canonical cover F_c of F and show your steps.
- (iv)(4 Points) Please give a lossless decomposition into 3NF of schema R , preserving functional dependencies.
- (v)(4 Points) Please give a lossless decomposition into BCNF of schema R .
- (vi)(5 Points) If we decompose R into $R_1(B, C)$, $R_2(A, D)$, $R_3(A, B, E)$, please show whether the decomposition (1) is lossless and (2) preserves functional dependencies.

VI.(19 Points) Answer the following questions.

Consider the following description of a sales management system and answer the questions that followed.

We are developing a basketball matching statistics system for a professional basketball team. The team manager wishes to use a database to store players' personal information and match performance. Players' personal information includes name, date of birth, age, height, weight, phone number, home address (including province, city, and detailed address), etc. Match performance information includes the players' technical statistics in each match, such as goals(投篮命中), assists(助攻), faults(犯规), time played(上场时间), etc. Match information should also be stored in the system. Match information includes match type (e.g., league season, club cup, continental cup), location, time, team total score, and opponent team's name and score in the match.

Please answer the following questions based on the above descriptions.

- (i)(3 Points) What do you think could be the most frequently used queries (data requirements) in terms of team management and player training? Please list at least three of them in natural language. [Hint: such as "list the players that have scored the most in a given season". Your answers should be different from this hint.]
- (ii)(5 Points) Please perform a conceptual design by constructing an E-R diagram based on the above description. Please include all attributes that are mentioned in the description. Add any attribute if you think you have to.
- (iii)(4 Points) Please derive appropriate relation schemas based on ERD designed in (i). Each schema should be in the form *relation(attribute1, attribute2, attribute3)*, where the primary key is underlined. Should you make any assumptions, document them as well.
- (iv)(4 Points) Write a query in SQL to find how many matches each player has played in year 2020 and how many goals he has made in each match on average.
- (v)(7 Points) Consider that the team now wishes to store all players of all teams in the National Basketball League and to keep track of their competitors' information. Players may be transferred annually between teams and the team wishes to find and buy talented players. Meanwhile, some aged players may get retired but his performance should be recorded permanently. Please explain briefly how you would change your design to fulfill the new requirements (or how your original design has already addressed the new requirements).

VII. [H] (+6 Points) Static Analysis Tools (SATs) are usually used for checking source code quality. However, many developers suffer from numerous false alarms of the analysis results and do not really care about the reported issues. Keeping track of all issues in the development history helps development managers to understand which types of issues are most likely to be "false alarms". Please give a brief discussion on how to identify potential "false alarms" with the help of the lifecycle of detected issues. Your discussion should include which data should be stored and how to analyze the data.

