Xi'an Jiaotong-Liverpool University
西交利物浦大学

XJTLU Entrepreneur College (Taicang) Cover Sheet

Module code and Title	Database Development and Design (DTS207TC)
School Title	School of AI and Advanced Computing
Assignment Title	R001- CW
Submission Deadline	**3rd Aug 2024 17:00 Beijing Time**
Final Word Count	NA
If you agree to let the university use your work anonymously for teaching and learning purposes, please type **"yes"** here.	**Yes**

I certify that I have read and understood the University's Policy for dealing with Plagiarism, Collusion and the Fabrication of Data (available on Learning Mall Online). With reference to this policy I certify that:

- My work does not contain any instances of plagiarism and/or collusion.
  My work does not contain any fabricated data.

By uploading my assignment onto Learning Mall Online, I formally declare that all of the above information is true to the best of my knowledge and belief.

Scoring – For Tutor Use	
Student ID	

Stage of Marking	**Marker Code**	**Learning Outcomes Achieved (F/P/M/D) (please modify as appropriate)**			**Final Score**
		A	**B**	**C**	
1st Marker – red pen					
Moderation – green pen	**IM Initials**	The original mark has been accepted by the moderator (please circle as appropriate):			Y / N
		Data entry and score calculation have been checked by another tutor (please circle):			Y
2nd Marker if needed – green pen					

For Academic Office Use			**Possible Academic Infringement (please tick as appropriate)**	
Date Received	**Days late**	**Late Penalty**	☐ **Category A**	Total Academic Infringement Penalty (A,B, C, D, E, Please modify where necessary) _____
			☐ **Category B**	
			☐ **Category C**	
			☐ **Category D**	
			☐ **Category E**	

Xi'an Jiaotong-Liverpool University
西交利物浦大學

Students

The assignment must be typed in an MS Word document and submitted via Learning Mall Online to the correct drop box. Only electronic submission is accepted and no hard copy submission.

All students must download their file and check that it is viewable after submission. Documents may become corrupted during the uploading process (e.g. due to slow internet connections). However, students themselves are responsible for submitting a functional and correct file for assessments.

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Database Development and Design (DTS207TC)

Resit Assessment (R001): Individual Coursework

Weight: 100%

Maximum Marks: 100

Submission deadline: 3rd Aug, 2024, 17:00 Beijing Time

Overview

This assignment aims to gain experience in understanding the internal functionality of different database management systems, including RDBMS, XML, Object-relational databases, and gain experience in designing data warehouses to perform OLAP operations. The coursework will be assessed for the following learning outcomes:

A. Identify and apply the principles underpinning transaction management within DBMS.
B. Demonstrate an understanding of advanced SQL topics.
C. Illustrate the issues related to Web technologies and DBMS and XML as a semi-structured data representation formalism.
D. Identify the principles underlying object-relational models.
E. State the main concepts in data warehousing and data mining.

Assessment Tasks

Your task is to answer every question by carefully reading the questions and guidelines for the system setup. Record your thoughts and assumptions, where necessary, while reporting your answers. There are the following five parts for this coursework corresponding to five learning outcomes of this module:

1. Advanced SQL, Triggers, Indexing, and Query Optimization
2. Transaction Management
3. Querying XML Data
4. Object-Relational Database
5. Data Warehousing and OLAP

Marking Criteria

This coursework will be graded out of 100% with a maximum of 100 marks. There are five (5) parts, and twenty (20) marks available for each part. Marks will be awarded as follows:

0 mark: Not attempted/wrong

50% marks: Partially attempted/correct

100% marks: Complete and correct

Xi'an Jiaotong-Liverpool University
西交利物浦大学

1: Advanced SQL, Triggers, and Indexing (20 Marks)

Q1 The dataset to be used in this exercise is the COMPANY database provided along with this assignment. The dataset contains a normalized schema for employee, department, project, work_on, and dependent table. You are required to load the data in Oracle Database and use SQL Developer to answer the below question (Note: you can use any DBMS such as MYSQL, SQLite etc. to answer this question). Write appropriate SQL statements and show the result of each query as it would apply to the database state.

Q1(a) For each department whose average employee salary is more than $30,000, retrieve the department name and the number of employees working for that department. **(2 marks)**

Q1(b) Suppose we want the number of male employees in each department rather than all employees (as in Exercise q1(b)). Can we specify this query in SQL? Why or why not? **(2 marks)**

Q1(c) Use the company database to create a PL/SQL program to Identify the one employee at a time who has been the least active, based upon the number of hours they have been working on projects. This will be the first employee we want to delete from the existing COMPANY database. The program should be able to remove this employee as a manager of any department, as a supervisor of other employees, delete any dependent and all WORKS_ON rows. Upon running, the program should display a message ('Least active employee has been transferred:' inactive employe last name). Write appropriate PL/SQL program and show the database state before and after running the program to verify that inactive employee has been successfully removed. **(6 marks)**

Q1(d) Consider a disk with block size B=512 bytes and a record pointer is PR =7 bytes long. A file has r=50,000 EMPLOYEE records of fixed-length. Each record has the following fields: NAME (30 bytes), SSN (9bytes), DEPARTMENTCODE (9 bytes), ADDRESS (40 bytes), PHONE (9 bytes),BIRTHDATE (8 bytes), SEX (1 byte), JOBCODE (4 bytes), SALARY (4 bytes, real number). An additional byte is used as a deletion marker.

By considering the above specifications, answer the following questions. (5*2=10 marks)

i. What would be the number of blocks required to store this data file assuming an unspanned file organization?

ii. How many blocks are required to search for a record if the data file is ordered on SSN as a key field? Compare with the search performance of unordered file organization.

iii. Suppose the file is ordered by the key field ID, and we have a primary index on SSN. How many blocks are required to search and retrieve a record from the data file given its ISSN value using the primary index?

iv. Suppose we want to construct a multi-level index. Calculate:

(i) How many index entries and the number of blocks are required for each level of the multi-level index? For example, the number of first-level index entries and the number

of first-level index blocks, the number of second-level index entries, and the number of second-level index blocks, and so on.

(ii) What would be the maximum number of blocks required to search for and retrieve a record from the data file with the multi-level index on the SSN field?

v. Suppose we create a B+ index on SSN, and the tree order is 100. What is the maximum capacity of the B+ tree of height 4, and how many blocks accesses are needed to search for and retrieve a record from the file--given its SSN value—using the 4 level B+ tree index? (2 marks)

2: Transaction Management (20 Marks)

Q2: Consider a database consisting of a single relation R (X, Y):

X	Y
1	10
2	0

Q2(a) The following two transactions run concurrently on this database: **(10 marks)**

T1	T2
1. Begin Transaction; 2. update R set Y = Y-10 where X=1; 3. update R set Y = Y+10 where X=2; 4. commit;	1. begin transaction; 2. select sum(Y) from R; 3. commit;

Is it possible for T2 to see a value of zero in its output? Explain why or why not by considering different isolation levels(Read Uncommitted, Read Committed, Repeatable Read).

Q2(b) The following two transactions run concurrently on this database: **(10 marks)**

T1	T2
1. Begin Transaction; 2. insert into R values (3,150) 3. commit;	1. begin transaction; 2. select sum(Y) from R; 3. select sum(Y) from R; 4. commit;

Is it ever possible for T2 to see a different value as the output of the select sum(B) from R statements? Discuss where a schedule sees the different values of sum in T2 by considering different isolation levels (Read Uncommitted, Read Committed, Repeatable Read).

3: XML Data Modeling (20 Marks)

Q3 Assume that the XML data conforms to the following DTD, In this problem, you are asked to design a relational schema for this DTD, and retrieve information from the XML document using Xpth and Xquery language.

```
<!DOCTYPE Bookstore [
   <!ELEMENT Bookstore (Book | Magazine)*>
   <!ELEMENT Book (Title, Authors, Remark?)>
   <!ATTLIST Book ISBN CDATA #REQUIRED
                  Price CDATA #REQUIRED
                  Edition CDATA #IMPLIED>
   <!ELEMENT Magazine (Title)>
   <!ATTLIST Magazine Month CDATA #REQUIRED Year CDATA #REQUIRED>
   <!ELEMENT Title (#PCDATA)>
   <!ELEMENT Authors (Author+)>
   <!ELEMENT Remark (#PCDATA)>
   <!ELEMENT Author (First_Name, Last_Name)>
   <!ELEMENT First_Name (#PCDATA)>
   <!ELEMENT Last_Name (#PCDATA)>
]>
```

Q3(a) Design a relational schema for the XML data. Your schema should have relations corresponding to entity sets such as Book, Magzine, Author, etc., as well as relationships between these entity sets using primary and foreign key relationships. Write only the relation names and their columns; for example, Book(ISBN, Price, Edition); do not write the field types nor the key/foreign key constraints. **(10 marks)**

In the next task, use the following XML document to answer Q3(b) and Q3(c).

```
<Bookstore>
   <Book ISBN="ISBN-0-13-713526-2" Price="85" Edition="3rd">
      <Title>A First Course in Database Systems</Title>
      <Authors>
         <Author>
            <First_Name>Jeffrey</First_Name>
            <Last_Name>Ullman</Last_Name>
         </Author>
         <Author>
            <First_Name>Jennifer</First_Name>
            <Last_Name>Widom</Last_Name>
         </Author>
      </Authors>
   </Book>
   <Book ISBN="ISBN-0-13-815504-6" Price="100">
      <Title>Database Systems: The Complete Book</Title>
      <Authors>
         <Author>
            <First_Name>Hector</First_Name>
            <Last_Name>Garcia-Molina</Last_Name>
         </Author>
         <Author>
            <First_Name>Jeffrey</First_Name>
            <Last_Name>Ullman</Last_Name>
         </Author>
         <Author>
            <First_Name>Jennifer</First_Name>
            <Last_Name>Widom</Last_Name>
         </Author>
      </Authors>
      <Remark>
         Buy this book bundled with "A First Course" - a great deal!
      </Remark>
   </Book>
</Bookstore>
```

Q3(b): Write down the Xpath expression for the following queries. For each query, provide query expression and results: (3*2=6 marks)

i. Retrieve all books costing less than $90.
ii. Retrieve titles of books containing a remark.
iii. Retrieve all magazines where there is a book with the same title.

Q3(c): Write down the XQuery expression for the following queries. For each query provide query expression and results: (2*2=4 marks)

i. Retrieve books where every author's first name includes "J".
ii. Retrieve all pairs of titles containing a shared author. The query should return results in XML format, as shown below.

```
<BookPair>
      <Title1> Title </Title1>
      <Title2> Title </Title2>
   </BookPair>
```

4: Object-Relational Data Model (20 Marks)

Consider the following set of requirements for a UNIVERSITY database that is used to keep track of students' transcripts.

- The university keeps track of each student's name, student number, Social Security number, current address and phone number, permanent address and phone number, birth date, sex, class (freshman, sophomore, ..., graduate), major department, minor department (if any), and degree program (B.A., B.S., ..., Ph.D.). Some user applications need to refer to the city, state, and ZIP Code of the student's permanent address and to the student's last name. Both Social Security number and student number have unique values for each student.
- Each department is described by a name, department code, office number, office phone number, and college. Both name and code have unique values for each department.
- Each course has a course name, description, course number, number of semester hours, level, and offering department. The value of the course number is unique for each course.
- Each section has semester, year, course, and section number. The section number distinguishes sections of the same course taught during the same semester/year; its values are 1, 2, 3, ..., up to the number of sections taught during each semester.

Q4(a) Design UML schema for this application and draw UML class diagram for the schema. Specify key attributes of each entity type and structural constraints on each relationship type. Note any unspecified requirements, and make appropriate assumptions to make the specification complete. **(10 marks)**

Q4(b): Translate the UML class diagram created in the above question **(Q4(a)** into a relational schema. In the relational schema, only use relation name and attribute, e.g., student (name, number, SSN,..etc.) and specify foreign key as fk. **(10 marks)**

5: Data Warehousing and OLAP (20 Marks)

Suppose a data warehouse for game management consists of five dimensions (Date, game, location, players, and spectators) and two measures (count and amount_paid). The count is the total tickets sold, and amount_paid is what the spectators paid to watch a game at the given date and location. The spectators can be adults and children with different charges to buy a ticket. The player information includes name, age, and type of game played. The date consists of day, month, and year. Location information includes street, city, and stadium.

Q5(a) state which schema (star or snowflake) is the most appropriate to model the above data warehouse. Give your opinion of which might be more empirically useful and state the reasons behind your answer. **(5 marks)**

Q5(b) Draw schema diagram for the above data warehouse using one of the most appropriate dimensions modeling techniques such as star, snowflake, or fact constellation diagram. Dimension and fact tables must capture the primary keys by introducing the attribute that identifies each entity. Furthermore, a complete schema diagram must capture associations among all the entities to perform OLAP operations. **(10 marks)**

Q5(c) If we want to summarize results to list the total charge paid by all adults at a given Stadium in the fall of 2018, and Chris Gale was one of the players. What OLAP operation, such as roll-up, slicing, dicing, etc., should be performed starting with the base cuboid, e.g. [date, game, location, player, spectator]? Outline your query step by step to retrieve the total charges, E.g., roll-up on the date to get quarte (Fall), and so on to satisfy other conditions. **(5 marks)**

---------- **End of the Assessment**----------
