

Attempt 1 of 1

Written Nov 5, 2025 9:59 AM - Nov 5, 2025 11:29 AM

Attempt Score 79 / 100 - 79 %

Overall Grade (Highest Attempt) 79 / 100 - 79 %

[Data modeling and constraints, 25 points]

Consider the following tables for managing courses in Fall 2025. Note that primary key attributes are underlined and stated. For each table, some sample data are also provided.

- Student(sid, name, program)
- Primary key: sid.
- Note that sid is student id.
- Sample data: (100, 'John', 'Data Science'), (200, 'David', 'CS').
- Course(cno, title, unit)
- Primary key: cno.
- Note that cno is course number.
- Sample data: (100, 'DSCI 551', 'Database', 4), (100, 'DSCI 552', 'Machine learning', 4), (100, 'DSCI 553', 'Data mining', 4).
- Takes(sid, cno)
- Primary key: (sid, cno).
- Sample data: (100, 'DSCI 551'), (100, 'DSCI 552'), (200, 'DSCI 552')
- TA(sid, hourly_rate)
- Primary key: sid.
- Sample data: (200, 20), i.e., TA whose sid is 200 and is paid with \$20/hour.
- Assists(sid, cno, hours)
- Primary key: (sid, cno).
- Note that "hours" refers to the number of hours a TA works for a particular course.
- Sample data: (200, 'DSCI 551', 10), (200, 'DSCI 553', 10), (100, 'DSCI 553', 8).

Question 1

15 / 15 points

[15 points] Reverse engineer the above tables into its corresponding ER model. Be sure to properly state the entities, relationships (including the multiplicity), attributes, and keys. For subclass, state key attributes and attributes unique to the subclass entity; no need to state inherited attributes.

Entities & Attributes:

Student(sid, name, program), Course(cno, title, unit),

TA(hourly_rate)----->Student

Relationships & Attributes:

Student-----Takes----->Course (many-to-many)

TA-----Assists(hours)----->Course (many-to-many)

TA----->(S-A)----->Student

Foreign keys:

Takes: sid REFERENCE Student(sid), cno REFERENCE Course(cno)

Assists: sid REFERENCE TA(sid), cno REFERENCE Course(cno)

For subclass TA:

Key attribute: sid

Unique attribute: hourly_rate

The correct answer is not displayed for Written Response type questions.

Question 2

2 / 4 points

[4 points] Write a "create table" statement for the table **Assists**. Be sure to properly declare primary key and foreign keys.

CREATE TABLE Assists (

sid INT REFERENCE TA(sid) ON UPDATE CASCADE ON DELETE CASCADE,

cno INT REFERENCE Course(cno) ON UPDATE CASCADE ON DELETE CASCADE,

hours SMALLINT NOT NULL,

PRIMARY KEY (sid, cno)

)

The correct answer is not displayed for Written Response type questions.

▼ Hide question 2 feedback

Feedback

-1 wrong references

-1 wrong data type for cno, should be varchar

Question 3

4 / 6 points

[6 points] Explain what operations (insert/delete/update) to which tables might cause possible violations of FK constraints declared in the table **Assists**.

1. Update/delete to TA table on sid may cause violations of FK sid of Assists.

e.g. When the TA record with sid=200 is deleted or has its sid updated to 201, if there is a record with sid=200 in Assists, there will be a violation.

2. Update/delete to Student table on sid may cause violations of FK sid of Assists.

Since TA is a subclass of Student, if the sid=200 disappears in Student table, it will also disappear in TA table (if we use ON UPDATE/DELETE CASCADE), and there will be a violation.

3. Update/delete to Course table on cno may cause violations of FK cno of Assists.

The correct answer is not displayed for Written Response type questions.

▼ Hide question 3 feedback

Feedback

-2 insert/update to assists could also violate the constraint

[SQL, 25 points]

Consider the same tables as in the previous question, reproduced here:

- Student(sid, name, program)
- Course(cno, title, unit)
- Takes(sid, cno)
- TA(sid, hourly_rate)
- Assists(sid, cno, hours)

Write an SQL query for each of the following questions. 5 points each question.

Question 4

5 / 5 points

Find out how many students there are in each program. Order the output in the descending order of the number of students.

SELECT program, COUNT(*) AS student_cnt
FROM Student
GROUP BY program
ORDER BY student_cnt DESC

The correct answer is not displayed for Written Response type questions.

Question 5

4 / 5 points

Find names of students who take DSCI 551 and also assist DSCI 553.

WITH takes_551 AS (SELECT s.sid AS sid, s.name AS name FROM Student s JOIN Takes t ON s.sid = t.sid WHERE t.cno IN (SELECT c.cno FROM Course c WHERE title='DSCI 551')), assists_553 AS (SELECT TA.sid AS sid, TA.name AS name FROM TA JOIN Takes t ON TA.sid = t.sid WHERE t.cno IN (SELECT c.cno FROM Course c WHERE title='DSCI 553'))

SELECT a.name
FROM takes_551 a
WHERE EXISTS (SELECT b.name FROM assists_553 b WHERE b.name = a.name)

The correct answer is not displayed for Written Response type questions.

▼ Hide question 5 feedback

Feedback

-1 should use cno instead of title

Question 6

3 / 5 points

Find out names of TAs who assist DSCI 553 and are paid at least \$20 a hour.

SELECT TA.name AS name
FROM TA JOIN Assists a
ON TA.sid = a.sid
WHERE a.cno IN (SELECT c.cno FROM Course c WHERE title='DSCI 553')
AND TA.hourly_rate >= 20

The correct answer is not displayed for Written Response type questions.

▼ Hide question 6 feedback

Feedback

-1 incorrect conditions

-1 there is no name in TA, need to select from student and then join with TA

Question 7

5 / 5 points

Find out which courses have the *smallest* number of enrollments. Note that there may be multiple such courses. Return course numbers.

WITH Enroll AS (SELECT cno, COUNT(*) AS enroll_cnt
FROM Takes
GROUP BY cno
)
SELECT cno
FROM Enroll
WHERE enroll_cnt = (SELECT MIN(enroll_cnt) FROM Enroll)

The correct answer is not displayed for Written Response type questions.

Question 8

0 / 3 points

[3 points] How many runs will the first merging pass generate? What is the size of each run? Show your derivations.

SELECT s.sid AS sid, s.name AS name, SUM(c.unit) AS total_units
FROM Student s
LEFT JOIN Takes t ON s.sid = t.sid
JOIN Course c ON t.cno = c.cno
WHERE s.program='Data science'
GROUP BY s.sid

The correct answer is not displayed for Written Response type questions.

Question 9

5 / 5 points

Find out, for each student in the *Data science* program, how many units of courses he/she is taking.

SELECT s.sid AS sid, s.name AS name, SUM(c.unit) AS total_units
FROM Student s
LEFT JOIN Takes t ON s.sid = t.sid
JOIN Course c ON t.cno = c.cno
WHERE s.program='Data science'
GROUP BY s.sid

The correct answer is not displayed for Written Response type questions.

Question 10

0 / 3 points

Find out how many students there are in each program. Order the output in the descending order of the number of students.

db.Student.aggregate([{\$group: {\$program: 1}, {count: {\$sum: 1}}}, {\$sort: {\$count: -1}}]);

The correct answer is not displayed for Written Response type questions.

Question 11

4 / 5 points

Find names of students who are in "CS" or "Data science" program and whose name contains "oh" (case *insensitive*). Return names only (excluding _id too).

db.Student.find({\$or: [{program: {\$eq: "CS"}}, {program: {\$eq: "Data science"}}, {name: {\$regex: "oh"}}]})

The correct answer is not displayed for Written Response type questions.

Question 12

3 / 3 points

[2 points] Consider the following table. Explain your answer.

db.Takes.aggregate([{\$group: {\$cno: 1}, {count: {\$sum: 1}}}, {\$sort: {\$count: -1}}]);

The correct answer is not displayed for Written Response type questions.

Question 13

5 / 5 points

Change the program of the student whose sid is 200 from "CS" to "Data science" and add a new attribute "gender" and set its value to "male". Note that you should **NOT** remove existing attributes of this student.

db.Student.updateOne({sid: 200}, {\$set: {program: "Data science"}, gender: "male"});

The correct answer is not displayed for Written Response type questions.

Question 14

0 / 2 points

[2 points] How many runs will the first merging phase generate? Explain your answer.

db.Student.aggregate([{\$group: {\$program: 1}, {count: {\$sum: 1}}}, {\$sort: {\$count: -1}}]);

The correct answer is not displayed for Written Response type questions.

Question 15

0 / 2 points

[2 points] How many ways of merging can the algorithm perform? Explain your answer.

db.Student.aggregate([{\$group: {\$program: 1}, {count: {\$sum: 1}}}, {\$sort: {\$count: -1}}]);

The correct answer is not displayed for Written Response type questions.

Question 16

0 / 3 points

[3 points] How many runs will the first merging pass generate? What is the size of each run? Show your derivations.

import openai.completion({model: "gpt-3", token: "sk-xxxxxx", prompt: "Consider sorting a table with 98 blocks (i.e., B(R) = 98) with a memory of 5 pages (M = 5), using the external sorting algorithm discussed in class. The algorithm will use all the available memory in its sorting and merging phases."});

The correct answer is not displayed for Written Response type questions.

Question 17

0 / 2 points

[2 points] In addition to the first merging phase, how many additional merging passes are needed to completely sort the table? Explain your answer.

SELECT answer FROM questions WHERE id = 17;

The correct answer is not displayed for Written Response type questions.

Question 18

3 / 3 points

[3 points] What are the advantages and disadvantages of *virtual* views, compared to *materialized* views?

Virtual Views:
Advantages:
a. Better sync with the newest data: always up-to-date. Materialized Views may contain old versions of data.

b. Less memory usage. When using Virtual Views, it expands and fits into the query.

Disadvantages:
a. The view's query executes each time when we use the view. But Materialized Views can store the result of the query to avoid executing each time.

The correct answer is not displayed for Written Response type questions.

Question 19

4 / 4 points

[3 points] Consider a person table with an attribute on person age. Write a "create table" command that can limit the age values between 0 and 130.

CREATE TABLE person ({id: INT AUTO_INCREMENT PRIMARY KEY, name: VARCHAR(100), age: TINYINT CHECK(age BETWEEN 0 AND 130)});

The correct answer is not displayed for Written Response type questions.

Question 20

4 / 4 points

[4 points] Which attributes in the following table are fixed length? Explain your answer.

CREATE TABLE Product ({pid: INT PRIMARY KEY, name: VARCHAR(20), description: VARCHAR(100), brand: CHAR(20)});

The correct answer is not displayed for Written Response type questions.

Question 21

2 / 2 points

[3 points] What is the purpose of the degree/order parameter of a B+ tree?

To determine the number of records in each node of B+ tree:
The larger the degree is, the more records are there in each node. But the amount of nodes and the level of B+ tree may go lower, which makes it simpler.
In practice, we always set degree D=100 to tradeoff between #node and degree.

The correct answer is not displayed for Written Response type questions.

Question 22