

**DASC 5333 Database Systems for Data Science**  
**CSCI 4333 Design of Database Systems**  
**Spring 2023**  
**Section 1 Final Examination**

**Family Name:** \_\_\_\_\_ **First Name:** \_\_\_\_\_ **Student Id:** \_\_\_\_\_

**Number:** \_\_\_\_\_ **Circle One:** CSCI 4333 or DASC 5333

Time allowed: 2 hours. Total score: 100 points. *Closed* book examination. An information sheet prepared by yourself is allowed. Answer all questions. Turn in both question and answer sheets (if needed).

**Academic honesty policy will be followed strictly. Cheating will be pursued vigorously and will result in a failing grade of D or below, a permanent academic record, and possibly other more serious penalty!**

**Use the toyu database in the supplementary sheet for questions including SQL and Python.**

(1) [24 points] Construct SQL statements for the following queries. Make sure that your answers generate the exact results, including column names and orders (if ordered).

(a) Show the information about every member and number of its major students who have enrolled in at least one course in the following manner.

```
+-----+-----+-----+
| deptCode | department | number of enrolled majors |
+-----+-----+-----+
| ACCT     | Accounting | 0 |
| ARTS     | Arts       | 1 |
| CINF     | Computer Information Systems | 2 |
| CSCI     | Computer Science | 3 |
| ENGL     | English    | 0 |
| ITEC     | Information Technology | 1 |
| MATH     | Mathematics | 0 |
+-----+-----+-----+
```

7 rows in set (0.001 sec)

(b) Show the ids and names of faculty of the school "CSE" who has advised two or more students in the following manner.

```
+-----+-----+-----+
| facId | faculty   | Number of advisees |
+-----+-----+-----+
| 1011  | Paul Smith | 2 |
+-----+-----+-----+
1 row in set (0.001 sec)
```

(c) Show all classId, course rubric and number, and the number of enrolled students of all classes in the following manner and order. Show only classes with an enrollment within 2 students of the maximum student enrollment of any class. In the toyu instance attached with the examination, the classes 10003 and 10004 both have four students enrolled. Thus, all classes with 2 to 4 students enrolled are shown.

```
+-----+-----+-----+
| classId | course    | enrollment |
+-----+-----+-----+
| 10003  | CINF 3321 | 4 |
| 10004  | CINF 4320 | 4 |
| 10000  | CSCI 2315 | 3 |
| 10005  | ITEC 3335 | 2 |
| 10001  | CSCI 4333 | 2 |
| 10007  | ENGL 1410 | 2 |
| 10002  | CSCI 5333 | 2 |
+-----+-----+-----+
7 rows in set (0.001 sec)
```

(2) [18 points + 2 bonus points] True or False. Circle one choice, or clearly write 'T' or 'F'.

(a) [T or F] MySQL transactions using the default InnoDB storage engine do not support the ACID property.

(b) [T or F] It is possible that "SELECT \* FROM R NATURAL JOIN S" returns an empty set for the relations R and S.

(c) [T or F] MongoDB is an example of a graphical NoSQL DB.

(d) [T or F] In Python, a function is not an object.

(e) [T or F] The function LIKE is a binary operator in SQL.

(f) [T or F] In SQL injection, syntactically correct SQL code is entered into a text field of a Web page by the attackers.

(g) [T or F] For R(A,B,C), the following SQL statement contains an error.

```
SELECT A, COUNT(B)
FROM R
WHERE COUNT(B) >= 10
GROUP BY A;
```

(h) [T or F] In relational theory, for a relation R, if  $X^+ = Z$ , then  $XY \rightarrow Z$ ,

(i) [T or F] It is possible that R(A,B) is not in BCNF.

(j) [T or F] (Bonus) Today is 5/3/2022.

(3) [9 points] Short Questions. State the candidate keys and the highest normal forms of the following relations. Assume the relations are at least in 1NF.

(a)  $R(A,B,C,D)$  with  $\{B \rightarrow D, C \rightarrow D, D \rightarrow A\}$

(b)  $R(A,B,C,D)$  with  $\{B \rightarrow AC, A \rightarrow BD\}$

(c)  $R(A,B,C,D)$  with  $\{B \rightarrow AC, A \rightarrow BD, C \rightarrow D\}$

(4) [9 points] Consider the following relation

$R(A,B,C,D,E)$   $\{A \rightarrow B, AB \rightarrow CD, D \rightarrow AC, C \rightarrow E\}$

(a) Show all candidate keys?

(b) What is the highest normal form (up to BCNF)? Why?

(c) If it is not in BCNF, can you losslessly decompose  $R$  into component relations in BCNF while preserving functional dependencies? How?

(5) [18 points] Write a Python CGI program, t2a.py, to accept two HTTP Get parameters *sid1* and *sid2* (both student ids) and display comparison information of the two students, including their ids, names, majors (department names, not department code), advisors' *facId*, and numbers of classes enrolled.

For example, for <http://.../t2a.py?sid1=100000&sid2=100003>, the following result specifies the required output:

major department name, not code

Ms. Lim enrolled in no class. and has no advisor assigned.

Id	Student	Major department	advisor facId	# classes enrolled
100000	Tony Hawk	Computer Science	1011	6
100003	Catherine Lim	Information Technology	Not assigned	0

There is no need for error checking of the user input parameters. A skeleton for t2a.py is provided for you.

```

from dbconfig import *
import pymysql
import cgi
import cgitb
cgitb.enable()

print("Content-Type: text/html;charset=utf-8")
print()
print("<html>\n<head></head>\n<body>")

# Get HTTP parameters: the ids of two students to be compared.
form = cgi.FieldStorage()
sid1 = form.getfirst('sid1')
sid2 = form.getfirst('sid2')

db = get_mysql_param()
cnx = pymysql.connect(user=db['user'], password=db['password'],
                      host=db['host'], database=db['database'])
cursor = cnx.cursor()

# your code here. Write in the back of the previous page if needed.

print('</body></html>')
cursor.close()
cnx.close()
quit()

```

(6) [10 points] Consider the collection 'student' in the db 'toyu' as stored in MongoDB:

```
toyu> db.student.find()
[
  {
    _id: ObjectId("63c19f66c1fb90601512c759"),
    stuId: 100000,
    fname: 'Tony',
    lname: 'Hawk',
    major: 'CSCI',
    minor: 'CINF',
    ach: 40,
    advisor: 1011
  },
  {
    _id: ObjectId("63c19f66c1fb90601512c75a"),
    stuId: 100001,
    fname: 'Mary',
    lname: 'Hawk',
    major: 'CSCI',
    minor: 'CINF',
    ach: 35,
    advisor: 1011
  },
  {
    _id: ObjectId("63c19f66c1fb90601512c75b"),
    stuId: 100002,
    fname: 'David',
    lname: 'Johnson',
    major: 'CINF',
    minor: 'ITEC',
    ach: 18,
    advisor: 1011
  }
]
```

Construct **Mongosh** query in JS to show the information of all students majoring or minoring in 'CINF' and with 15 or more ach credits in the following format. Answer in **the back of the previous page** if needed.

```
[
  {
    stuId: 100000,
    major: 'CSCI',
    minor: 'CINF',
    student: 'Tony Hawk',
    'ach credits': 40
  },
  {
    stuId: 100001,
    major: 'CSCI',
    minor: 'CINF',
    student: 'Mary Hawk',
    'ach credits': 35
  },
  {
    stuId: 100003,
    major: 'ITEC',
    minor: 'CINF',
    student: 'Catherine Lim',
    'ach credits': 20
  },
  {
    stuId: 100006,
    major: 'CINF',
    minor: 'ITEC',
    student: 'Lillian Johnson',
    'ach credits': 18
  }
]
```

(7) [12 points] (a) [3 points] Three facts are known for R(A,B,C,D,E):

1. There are two candidate keys. One of them is A.
2. D and E are non-prime attributes.
3. There are 20 superkeys.

What is the second candidate key?

(b) [9 points] Consider the relation Tutoring(TutorId, TutorEMail, StudentId, StudentEMail, SubjectId, SubjectName), which stores tutoring information about a tutor tutoring a student in a subject. For example ('T11', 'karl404@gmail.com', 'S21', 'paul503@gmail.com', 'CSCI', 'Computer Science') stores the information that the tutor 'T11', (with an email address of 'karl404@gmail.com'), tutors the student 'S21' (with an email address of 'paul503@gmail.com') on the subject CSCI (with code 'CSCI' and name 'Computer Science').

It is known that TutorId, StudentId and SubjectId are unique identifiers of tutors, students, and subjects respectively. Information about the tutors, students, and subjects are stored in other relations. Only one email is stored for a tutor or a student. Both the code and name of a subject are unique. If necessary, make reasonable assumptions.

(i) List the functional dependencies representing the specification above.

(ii) What are the candidate keys?

(iii) What is the highest normal form for the relation Tutoring? Why?