

1 (10 points) Terms and Definitions

(2 points each) In *at most* two short sentences each, explain the meaning of the following terms as they relate to database management systems.

1. **Entity Set**

2. **Super Key**

3. **Null**

4. **Integrity Constraint**

5. **Natural Join**

2 (10 points) EJBANK Relational Algebra

Evan's bank stores its data in two relations, with the following SQL schema:

```
CREATE TABLE Customers(  
  cid int PRIMARY KEY,  
  name text,  
  state text  
);  
  
CREATE TABLE Accounts(  
  aid int PRIMARY KEY,  
  cid int NOT NULL REFERENCES cid,  
  balance real NOT NULL  
);
```

1. **(2 points)** Write a relational algebra expression to compute the account ids that have balances greater than \$50,000.
2. **(4 points)** Write a relational algebra expression to compute the names of customers with balances greater than \$50,000 in the state of "NY".
3. **(4 points)** Given the following values for the Account relation:

aid	cid	balance
1	102	1000.00
2	102	2000.00
3	107	2000.00
4	108	1000.00

What is the result of the following relational algebra expression?

$$\begin{aligned} & \rho(A, Accounts) \\ & \rho(B, Accounts) \\ & \pi_{A.aid, B.aid}(A \bowtie_{A.balance > B.balance} B) \end{aligned}$$

(continued on next page)

Fill in your answer in this table. Do not fill in the names for the fields. *Note:* you may or may not need all the columns and rows.

3 (20 points) Medical testing Entity-Relationship Modelling

A medical lab testing company has several testing centers all over the country. In this problem, you will design a schema to keep track of the testing centers, tests and order information. Specifically, you will need to keep track of:

1. The address, city, state and manager for each testing center.
2. The equipment at each testing center including the name of the machine, manufacturing year, and status of each machine (either functional or in repair). Note that one center may have multiple machines of the same kind.
3. The types of tests the company can perform. Tests have a name, time to run and price. The price of the test depends on the specific testing center. Some tests are only available at specific centers.
4. When an order for a test is submitted, the lab must record the doctor's name, the patient's name, and the date it was ordered. Each order is performed at a specific testing center. It must be performed at a center that offers that type of test.

Part 1: (10 points) Draw an ER diagram representing your database. Include 1-3 sentences of justification for why you drew it the way you did.

Part 2: (10 points) Write a SQL schema for your database. Include 1-3 sentences of justification for why you chose the tables you did.

(Additional space on the following page)

4 (20 points) Wikipedia in SQL

For this question, we will use a simplified schema based on Wikipedia, shown below. Each page has a unique id, a human readable title, and the length of the page (in bytes). Links between pages are stored in the Link table. The source is the page that contains the link, and dest is the page that the link points to. As an example, a Link tuple with values (source=50, dest=100) means that page id 50 contains a link to page id 100.

```
CREATE TABLE Page(  
    id int PRIMARY KEY,  
    title text NOT NULL,  
    length int NOT NULL  
);  
  
CREATE TABLE Link(  
    source int REFERENCES Page,  
    dest int REFERENCES Page,  
    PRIMARY KEY (source, dest)  
);
```

1. (6 points) Circle true or false for the following statements:

- (a) **True / False** There can be two pages with the title “Venus”.
- (b) **True / False** Broken links may exist (a link where the source or the destination pages do not exist).
- (c) **True / False** A page can only be the source of one link.
- (d) **True / False** If the page “Venus” contains a link to “Mars”, deleting “Venus” will not be permitted.
- (e) **True / False** If the page “Venus” is not the source of any links, deleting it will be permitted.
- (f) **True / False** Renaming pages is not permitted.

Write SQL queries to answer the following questions:

2. (2 points) What is the id and title of all pages that have titles that begin with the string “Database”?

3. (2 points) What are the titles of the 10 longest pages (in bytes)?

4. (2 points) What are the titles of all pages linked from the page with id 42?

5. **(4 points)** What are the titles of all pages linked from pages with the title “Database”?
6. **(4 points)** Popularity of a page is defined as the number of incoming links (links leading to that page). What are the titles of the 5 most popular pages?