1. [1] Which of the following is not true about databases?
   a) A random assortment of data cannot correctly be referred to as a database.
   b) Among other things, a database is an organized collection of data.
   c) Among other things, a database is a collection of programs that allow quick and easy retrieval of data.
   d) A database represents a certain aspect of reality.

2. [1] Which of the following is not true about database management systems?
   a) A DBMS is a general purpose program and cannot generally be used off the shelf for a specific application without a certain amount of customization and/or programming.
   b) A DBMS allows users to create (define and populate) data, maintain (insert, delete, and update) data, and manipulate (query) a database.
   c) A DBMS is a collection of related data.
   d) DBMS have one form or another of a DDL and a DML.

   a) ____________________________________________  (Any combination of the following):
      DBMS
      Database
      Application Programs
      Users
   b) ____________________________________________
   c) ____________________________________________

4. [1] Which of the following is not true regarding DBMS and traditional file processing systems (FPS)?
   a) There is a greater amount of data abstraction in a DBMS.
   b) DBMS allow for greater program-data independence because meta data are not stored together with the database.
   c) Adding a new attribute in a FPS typically has more impact on programs than adding a new attribute in a DBMS environment.
   d) None of the above (all are true).

5. [1] As we discussed, data models can be grouped into three general categories: conceptual, such as the ER model; implementation (or representational), such as the Relational model; and physical. Which of the following is not true?
   a) Conceptual data models involve high levels of abstraction.
   b) Both conceptual and implementation data models can be used to design a database schema without concern for physical storage requirements.
   c) Although the ER model is a popular conceptual model, it is of no use when the DBMS to be implemented is a relational DBMS.
   d) Conceptual models provide a convenient way to understand the data requirements of a database and communicate with users.

6. [1] What does a relational database schema consist of?
   Relation schemas and integrity constraints
7. [1] What is the advantage of the Three-Schema Architecture?

In general, changes to one schema have minimal impact on changes to another schema.

8. [1] Explain what is meant by an attribute’s domain.

The set of allowable values for that attribute.

9. [2] Which of the following is not true about representing a N:M relationship between two tables, A and B, using the Relational Model?
   a) Two foreign keys are stored in the same table.
   b) Table A will have a primary key and a foreign key.
   c) Three tables are involved.
   d) N:M relationships are created via two 1:N relationships.
10. Aurora Skies Airlines tracks flights they make to and from several locations around the world. Aurora's database has a table for flights (FLIGHT), where the primary key is Flight Number, and one for locations (LOCATION), where the primary key is Location Number. Every flight has exactly one departure location and exactly one destination location. At any given time, there may be departure and/or destination locations that are not represented in any flights. Aurora also has a table of flight attendants (ATTENDANT), in which the primary key is Attendant Number. Every flight can have several attendants but must have at least three, and attendants can work on many flights, but an attendant may not be working on any flights at a given time.

a) [3] Explain the existence of a DEPARTURE entity-type and a DESTINATION entity-type.

If every flight has a departure location and a destination location, then every flight has a departure location attribute and a destination location attribute, each of which must include valid locations, i.e., locations that are found in the primary key of the LOCATION table. While the domains are the same for both (the primary key in the LOCATION table), the roles are different (departure location and destination location), which makes them distinct. These facts describe distinct referential integrity constraints between the FLIGHT table and the LOCATION table, and therefore these attributes are distinct foreign keys. If they are distinct foreign keys, they must point to distinct entities in one or more entity-type. This establishes the existence of a DEPARTURE entity-type and a DESTINATION entity-type, both of which co-exist in the LOCATION table.

Alternate explanation (simpler):
The LOCATION table plays two roles with respect to FLIGHT; any location entity in the LOCATION table can be a departure entity or a destination entity, and so the LOCATION table includes a DEPARTURE entity-type and a DESTINATION entity-type.

b) [5] Draw a Formal ER diagram that represents the facts described. Use the min/max notation.
c) [3] In the table below, specify the cardinality ratio and the participation constraint for each of the relationships listed.

<table>
<thead>
<tr>
<th>For the relationship between these two relations:</th>
<th>What is the cardinality ratio?</th>
<th>What is the participation constraint for each of these relations?</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLIGHT &amp; DESTINATION</td>
<td>1:N</td>
<td>FLIGHT Total</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DESTINATION Partial</td>
</tr>
<tr>
<td>FLIGHT &amp; ATTENDANT</td>
<td>N:M</td>
<td>FLIGHT Total</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTENDANT Partial</td>
</tr>
</tbody>
</table>


d) [3] Which constraint, if any, is violated in the following scenarios? In all scenarios, only one maintenance operation is performed on only one table. (Write “None” if no constraints are violated.)

i) An attempt to delete an attendant succeeds. At the time of the deletion, the attendant was scheduled to work on a flight.
   CONSTRAINT VIOLATED IF ANY: Referential

ii) An attempt to add a new flight with a null Flight Number succeeds.
    CONSTRAINT VIOLATED IF ANY: Entity

iii) An attempt to assign an existing attendant to a flight succeeds. At the time of the assignment, the flight did not exist in the FLIGHT table.
    CONSTRAINT VIOLATED IF ANY: Referential

iv) An attempt to insert a new attendant succeeds. The values of all the attributes of the new attendant are identical to the values of all the attributes of one other attendant already in the table.
    CONSTRAINT VIOLATED IF ANY: Key
CSCI S-T 31206 DATABASE MANAGEMENT SYSTEMS

MIDTERM EXAM (40 points)

Harvard Summer School
Peter Avila, Instructor

YOUR NAME: ______________________________________________________________

This is a closed-book test. You have 2 hours in which to complete it. For each of the multiple-choice questions, circle only one answer. For each of the True/False questions, circle one of the words True or False which corresponds to your answer. For each question regarding a written answer, write your answer in the space provided below the answer. If you require more space, please indicate with an arrow, continue on the back of the page, and identify the question you are answering in your continuation. Feel free to draw diagrams, etc., if needed.

e) [3] For each of the five scenarios in question 10d which result in a constraint violation, what is/are the option(s) available to the DBMS to prevent the integrity constraint violation?

i) Reject or cascade the delete

ii) Reject

iii) Reject

iv) Reject

11. Expanding on the scenario given in question 10, suppose the Location relation acquires the following additional attributes: City Name (unique), Country Name, Population, and Weather Code.

a) [1] List fields that would make up a superkey of this relation.

(Location Number, City Name, or any combination of attributes involving these two.)

b) [1] What is/are the key(s) of the Location relation? (If more than one, list each on a different line.)

City Name
Location Number

Because the Location relation has more than one key.

c) [1] Using the definition of candidate key adopted in this course, does the Location relation have any candidate keys? Yes Why or why not?

Because the Location relation has more than one key.

12. [2] In mathematical terms, a relation state is a set of tuples \( r(R) = \{t_1, t_2, \ldots, t_n\}\). What, in mathematical terms, is a relation? (You can answer this in English without mathematical notation if you prefer.)

A relation is the Cartesian product of the domains of its attributes.

(Using mathematical notation):

If \( \{A_1, A_2, \ldots, A_n\} \) is the set of attributes of a relation \( R \), and \( \text{Dom}(A_i) \) is the domain of an attribute \( A_i \) in \( R \), then for a relation \( R \),

\[
R = \text{Dom}(A_1) \times \text{Dom}(A_2) \times \ldots \times \text{Dom}(A_n)
\]

13. [2] If two attributes share the same domain, such as “State of Birth” and “State of Residence,” why would it make sense for them to have different names?

Because the domains play different roles.
14. [1] RoseWind is a company that conducts polls in different neighborhoods. Results of polls are often sorted by many different criteria, including zip codes, street names, and even/odd address numbers.

TRUE  [FALSE]

The following address value is atomic: “2233 Westmore Lane”

15. [1] If a single, atomic attribute is unique in a relation state, does that mean that it is a key of the relation?

>No Why or why not?

It must be single and atomic (minimal superkey, or key) in the relation.

16. For each of the ER diagrams shown below, map the information shown to the relational model using an Informal ER diagram on the back of this page (clearly identified as 16a and 16b) that maps as many of the facts described as possible and any others that may be necessary.

a) [2]

b) [2]
This is a closed-book test. You have 2 hours in which to complete it. For each of the multiple-choice questions, circle only one answer. For each of the True/False questions, circle one of the words True or False which corresponds to your answer. For each question regarding a written answer, write your answer in the space provided below the answer. If you require more space, please indicate with an arrow, continue on the back of the page, and identify the question you are answering in your continuation. Feel free to draw diagrams, etc., if needed.

16a)

(If you included a PK for the Author table, that’s a good idea, but it is not required because it is not indicated in the original diagram.)

16b)

(If you included a PK for the Project table, that’s a good idea, but it is not required because it is not indicated in the original diagram.)
This is a closed-book test. You have 2 hours in which to complete it. For each of the multiple-choice questions, circle only one answer. For each of the True/False questions, circle one of the words True or False which corresponds to your answer. For each question regarding a written answer, write your answer in the space provided below the answer. If you require more space, please indicate with an arrow, continue on the back of the page, and identify the question you are answering in your continuation. Feel free to draw diagrams, etc., if needed.

**Bonus Questions**

Bonus questions can help you recover points you may have lost on this test. You cannot score more than 40 points on this test, so if you missed 3 points and earned 5 bonus points, you simply recovered the three you lost, for a total of 40 points. If you missed 5 and earned 2 bonus points, your score went from 35 to 37, and so on.

**BONUS 1 [3]**: Draw an Enhanced Informal ER diagram for question 16a that shows as many of the facts shown in the diagram of question 16a as possible.

While less accurate because it does not map to the relational model, I will also accept this:

(If you included a PK for the Author table, that's a good idea, but it is not required because it is not indicated in the original diagram. Also, while the participation constraint of Author is not explicitly indicated in the original diagram, it is implied, since an author is only an author if the author has at least one publication! However, I will accept whatever you put for this.)

**BONUS 2 [1.5]**: In terms of the three-schema architecture, explain what is meant by physical and logical data independence and why it is easier to achieve physical data independence than it is to achieve logical data independence.

Physical data independence means that when we make changes to the internal schema, we do not need to make changes to the conceptual schema, and logical data independence means that when we make changes to the conceptual schema, we do not need to make changes to the external schema. When changes are made to the internal schema, they rarely impact the conceptual schema because of the nature of the changes, which usually involve simply reorganizing what's in the conceptual schema, but when we make changes to the conceptual schema, we may easily add or remove the data structures referenced by the external schema, and so such changes are more likely to impact the external schema.

**BONUS 3 [0.5]**: What is another term for a unique identifier that is a single, atomic attribute of a relation?

Minimal superkey, or key.