

MSDS 7330

File Organization and Database Management

Homework Databases

Name:

This is a homework assignment for MSDS 7330, File Organization and Database Management. Print out this document, and unless directed otherwise, hand write your answer to a question in the white space under the question. Scan your completed solutions and submit them in a single pdf file. Be sure to put your last name at the beginning of the filename. For example, for Daniel Engels, the submitted file should be named EngelsMSDS7330HW2.pdf.

Collaboration is expected and encouraged; however, each student must hand in their own homework assignment. To the greatest extent possible, answers should not be copied but, instead, should be written in your own words. Copying answers from anywhere is plagiarism, this includes copying text directly from the textbook. Do not copy answers. Always use your own words. Directly under each question list all persons with whom you collaborated and list all resources used in arriving at your answer. Resources include but are not limited to the textbook used for this course, papers read on the topic, and Google search results. Don't forget to place your name on every page.

Introduction to Databases

Question 1 : A database contains a collection of data relevant to an enterprise. A database-management system (DBMS) provides a set of programs to manage the data and to access the data stored in the enterprise's databases. The database system and the DBMS were designed to overcome the limitations of the typical file processing systems used prior to the invention of the modern database. Identify three broad problems experienced with file processing systems, such as data integrity problems, and discuss how a database system might be able to overcome the problems.

Collaborators:

Resources:

Question 2 : Data may be represented at three different layers of abstraction: *physical*, *logical* and *view*. At which level should a database administrator design a database and why?

Collaborators:

Resources:

Question 3 : Provide an answer to each of the following questions.

Collaborators:

Resources:

Name three (3) approaches to storing data.

Name three (3) applications in which data is used extensively.

Question 4 : Provide a concise definition for each of the following:

Collaborators:

Resources:

Database Schema

Data Manipulation Language (DML)

Data Definition Language (DDL)

Question 5 : Database applications are typically partitioned into either two or three tiers. For each of the following statements regarding database applications identify whether the statement is either *True* or *False* and state why you believe the statement to be thus.

Collaborators:

Resources:

In a three-tier architecture the client machine acts as a front end interface to the user and does not contain any direct database calls.

In a two-tier architecture the user accesses the application located on the database server. The application accesses the database system directly.

In a three-tier architecture the application functionality is split between the client machine and the database server machine.

The Relational Model

Question 6 : The relational model utilizes its own unique vocabulary. It is important to learn this vocabulary in order to quickly and easily understand any discussion on relational databases. To this end, define each of the following.

Collaborators:

Resources:

Relation

Relation Instance

Tuple

Attribute

Domain

Atomic Domain

Question 7 : Keys are used to distinguish tuples within a relation. In the fundamental relational model, no two tuples are allowed to have the same attribute values for all attributes. This ensures that a subset of the attributes may be used to uniquely identify each tuple. These subsets are referred to as *keys*. Define each of the following types of keys.

Collaborators:

Resources:

Superkey

Candidate Key

Primary Key

Foreign Key

Question 8 : The relational model allows for a set of operations to be performed on relations in accordance to relational algebra. A relational operation is the Natural Join. Describe what the Natural Join does. Identify its inputs and its output. Give a simple example. What is the benefit of using a Natural Join?

Collaborators:

Resources:

Consider the following relational database for the next two questions.

employee (*personName*, *street*, *city*)

works (*personName*, *companyName*, *salary*)

company (*companyName*, *city*)

Question 9 : Consider the relational database above. What are the appropriate primary keys?

Collaborators:

Resources:

Question 10 : Consider the relational database above. Give an expression in the relational algebra to express each of the following queries:

- 1) Find the names of all employees who live in the city of Miami.
- 2) Find the names of all employees whose salary is greater than \$100,000.
- 3) Find the names of all employees who live in Miami and whose salary is greater than \$100,000.

Collaborators:

Resources:

Consider the following Python code for the following questions.

```
1  print "Think of a number between 1 and 100."
2  print "Now allow me to guess your number."
3  minNumber = 1
4  maxNumber = 100
5  correctAnswer = False
6  while not correctAnswer:
7      midNumber = (maxNumber + minNumber + 1) / 2
8      answer = raw_input('Is your chosen number ' + str(midNumber) + '?')
9      if answer[0] == 'y':
10         correctAnswer = True
11     elif answer.startswith('lower'):
12         maxNumber = midNumber - 1
13     elif answer.startswith('higher'):
14         minNumber = midNumber + 1
15     else:
16         print "I don't understand your answer."
17
18 print 'Yeah! I got it!'
```

Question 11 : The while loop exits when the variable `correctAnswer` is `True`. What will cause `correctAnswer` to be `True`?

Collaborators:

Resources:

Question 12 : How many times will the program print out 'Yeah! I got it!?'?

Collaborators:

Resources:

Question 13 : What is purpose of the variable `answer`?

Collaborators:

Resources:

Question 14 : The program prints out its guess at line 8. What user response will the program understand in response to its guess?

Collaborators:

Resources:

Question 15 : If the program gets the response 'higher', what does that tell the program about its guess?

Collaborators:

Resources:

Question 16 : What are the variables `minNumber`, `maxNumber` and `midNumber` used for?

Collaborators:

Resources:

Question 17 : A useful data type built into Python is the dictionary. Dictionaries are indexed by *keys*, which can be any immutable type. Keys are commonly of type string or a number. A dictionary is easily considered to be an unordered set of *key:value* pairs. Keys must be unique within a dictionary. The main operations on a dictionary are storing a value with some key and extracting the value given the key.

Create a Python dictionary that catalogs the courses you are currently taking for the SMU MSDS program. The keys should be the course number, and the values should be the course titles.

Write a function `add_course` that takes 2 arguments (a course number a title) and adds a course to your dictionary. Use this function to add the courses you're taking next term to the dictionary.

Write a function `print_courses` that takes one argument (a search string that is a portion of a course number) and prints out all the courses in your dictionary that match the search string in any part of the course number. Be sure to have an appropriate response for a search string that has no matches.

Attach to the homework pdf a pdf document of your source code and a pdf of the screen capture of an execution of your database and functions. Be sure to include a screen capture that captures all possible functionality for your program.

Collaborators:

Resources: