ITC 510 - Software and Data Modeling
Spring 2024 Semester
Midterm Exam #1

This is a closed book, closed notes exam. You are allowed a one-sided single sheet (8 ½" x 11") of handwritten notes which must be turned in with this exam. This exam must be done individually. Corresponding with someone else during the exam is a violation of the academic integrity policy of the University and grounds for expulsion.

This exam is worth a maximum of 150 points and contains 24 true/false and multiple-choice questions (5 pts each) as well as three short answer (10 pts each) questions. There is also an extra credit question worth an additional 15 points that you can answer after you turn in this exam. The extra credit can only be applied to your score on this exam and cannot be applied to exceed the maximum attainable points on this exam.

Clearly mark your answers to the true/false questions. Answers to all short answer questions should be concise and must be kept to 50 words or less. Only information provided at the beginning of any answers exceeding that limit will be considered.

You have until 4:30pm to complete this exam. The remaining 15 minutes of the class period can be used on the extra credit portion of the exam.

1. TRUE/FALSE: Visualizations (plots) are an important task in data analysis because they are part of exploratory process.

2. TRUE/FALSE: A real number (with decimal) in an Excel file always gets loaded as a float into a pandas DataFrame.

3. TRUE/FALSE: Backward compatibility means that older code can read data that was written by newer code.

4. TRUE/FALSE: Python strings are immutable.

5. TRUE/FALSE: In Python, an int is an object that has attached methods that have access to the object's internal contents.

6. TRUE/FALSE: A Python function return statement can return more than one value provided the values being returned are placed in a Python list.

7. TRUE/FALSE: The Document Model is closely related to the structure of a Python dictionary.

8. TRUE/FALSE: Spaces around subplots can be adjusted usings the Figure objects add_subplots method.

9. TRUE/FALSE: The pandas file read operation can be used to load an Excel file into a pandas Series.

10. TRUE/FALSE: SSTables can only be created from LSM-trees..

11. TRUE/FALSE: In Python, a colon (:) denotes the start of an indented code block after which all of the code within the block must be identically indented.

12. TRUE/FALSE: A pandas Series is like a traditional array of similar type values, but it also includes a corresponding array of labels for each value.

13. MULTIPLE CHOICE: Which of the following is considered a feature of reliability?
    A. Capacity planning
    B. Scaling out
    C. Simplicity and the ability to understand and ramp up
    D. Quick roll-back

14. MULTIPLE CHOICE: What does the following line of code do?

    arr = np.arange(28).reshape((7, 4))

    A. Creates a four row and seven column NumPy matrix where every cell contains the same value.
    B. Creates a four row and seven column NumPy matrix where every cell contains a different value.
    C. Creates a seven row and four column NumPy matrix where every cell contains the same value.
    D. Creates a seven row and four column NumPy matrix where every cell contains a different value.

15. MULTIPLE CHOICE: If the NumPy array constructor is given a list with a combination integer and float (real) numbers, it will:
    A. Generate an error
    B. Return an empty array
    C. Automatically convert all elements to float
    D. Automatically convert all elements to int

16. MULTIPLE CHOICE: Which of the following is supported by Python lists and NumPy arrays?
    A. List elements and NumPy array elements are always stored in a sequential location in memory.
    B. List elements and NumPy array elements can be referenced using indexing and slicing.
    C. List elements and NumPy arrays elements all must be of the same type.
    D. List elements and NumPy array elements are both considered static data structures.

17. What does the following method do?

    info.dtypes()

    A. Returns a Python list containing the type of each element in a list at the corresponding index.
    B. Returns Python dictionary of key/value pairs corresponding with the key representing the index of an element in a Python list and the value the species the type of the value at that index.
    C. Returns a pandas Series with the data type of each column of a pandas DataFrame called info.
    D. None of the above

18. MULTIPLE CHOICE: What does the names parameter in the code below do for the creation of the DataFrame below:

pd.read_csv("ex1.csv", names=["one", "two", "three", "four"])

A. Only include the first four columns in the input file
B. Implicitly convert the first four columns datatype to string
C. Place the indicated values into the first row of the DataFrame.
D. None of the above

19. MULTIPLE CHOICE: A Python dictionary is an example of which of the following models?
A. Relational Model
B. Document Model
C. Graph Model
D. All of the above

20. MULTIPLE CHOICE: Which of the following is a feature of OLTP?
A. Processing several transactions with high availability
B. Vectorized processing
C. Column-based store
D. Optimized for data crunching

21. MULTIPLE CHOICE: Which of the following use a simple schema to describe a binary encoding format?
A. Protocol Buffers
B. Thrift
C. Avro
D. All of the above

22. MULTIPLE CHOICE: Which of the following modes of Dataflow does REST fall under?
A. Dataflow Through Databases
B. Dataflow Through Services
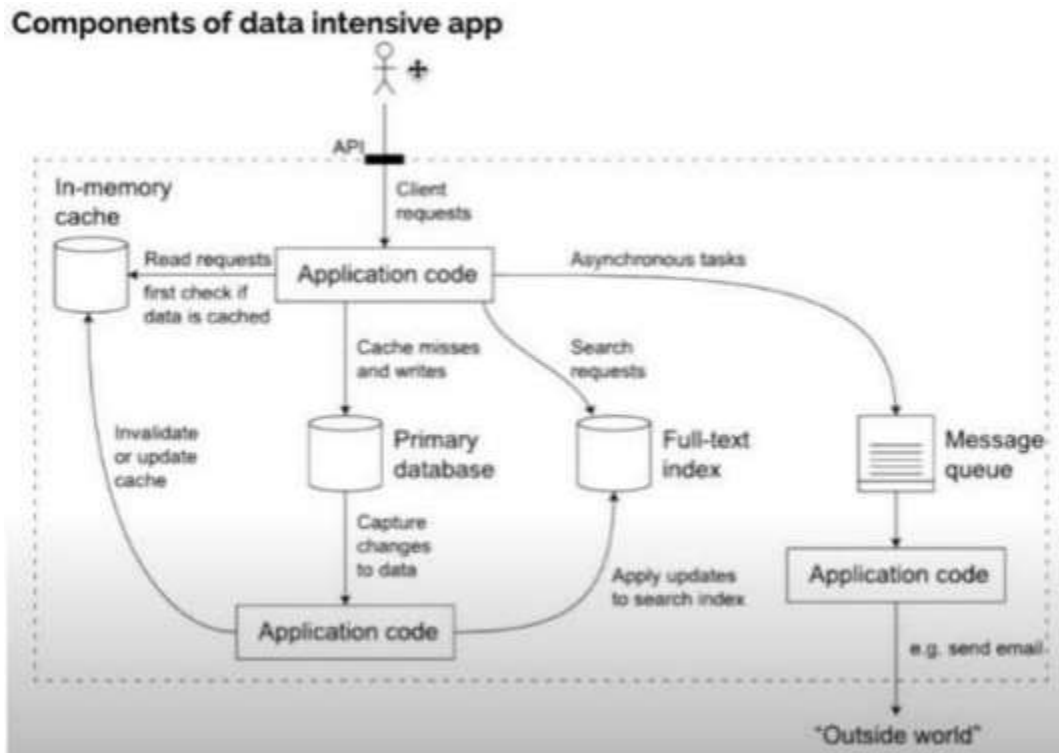C. Message-Passing Dataflow
D. None of the above

23. MULTIPLE CHOICE: Batch processing
A. stores data so that another application can find it again later.
B. remembers the result of an expensive operation to speed up read operations.
C. uses asynchronous message passing between processes.
D. involves crunching a large amount of accumulated data.

24. SHORT ANSWER (50 WORDS MAX): Many programming languages come with built-in support for quick and easy encoding in memory objects into byte sequences. (i.e. Java - java.io.Serializable, Ruby - Marshal, Python pickle). What are some of the _advantages_ to utilizing these built-in capabilities? (do not discuss disadvantages)

25. SHORT ANSWER (50 WORDS MAX): The pandas Python package provides that Series object and the DataFrame object. These two objects have some similarities (including methods) as well as several difference. Discuss a few of their _differences_. (do not discuss similiaries)

26. SHORT ANSWER (50 WORDS MAX): Using the figure below, it doesn't show the cache directly connected to the search index. However, assuming part of the cache is dedicated to the search index, in your own words, explain the different operations the cache may have to perform for client search index requests.

Components of data intensive app

ITC 510 - Software and Data Modeling
Spring 2024 Semester
Midterm Exam #1 - Extra Credit

Using the code below:

```python
def RefineList(largelst,num):
    flag = True
    uniqlst = None
    if type(largelst) == list:
        for i in range(len(largelst)):
            if type(largelst[i]) != str:
                flag = False
        if flag == False:
            print("Error #1")
    else:
        flag = False
        print("Error #2")
    if type(num) != int:
        flag = False
        print("Error #3")
    if flag:
        uniqlst = []
        worddict = {}
        for i in range(len(largelst)):
            if largelst[i] in worddict:
                worddict[largelst[i]] = worddict[largelst[i]]+1
            else:
                worddict[largelst[i]] = 1
        for key in worddict:
            if worddict[key] == num:
                uniqlst.append(key)
    return uniqlst
```

A) Provide at least one syntactically correct function call that successfully calls the function, and it returns a non-empty result.

B) Provide at least one syntactically correct function call that successfully calls the function, but the function prints either error message #1 or error message #2 and returns an empty result.

C) Provide a header comment that identifies the input parameters and their type as well as the return value and type (if any).   Include a thorough high-level description of what the function does.