

**ITC 510 - Software and Data Modeling**  
**Fall 2023 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 12:50pm to complete this exam (including extra credit).

- 1. TRUE/FALSE: The default for the Figure object's plot method is a solid line graph.**
- 2. TRUE/FALSE: SSTables are similar to Python dictionaries except that keys are stored in sorted order.**
- 3. TRUE/FALSE: Python lists are immutable.**
- 4. TRUE/FALSE: The Relational Model is closely related to structure of a Python dictionary.**
- 5. TRUE/FALSE: The pandas file read operation can be used to load an Excel file into a pandas DataFrame.**
- 6. TRUE/FALSE: The following line of Python code creates a two-dimensional NumPy array with three rows and five columns .**

**`arr = np.arange(15).reshape((3, 5))`**

- 7. TRUE/FALSE: In order to incorporate a pandas object in a Jupyter notebook code cell, the pandas package must have been previously imported in the same cell or a previous executed cell.**

8. **TRUE/FALSE:** The `add_subplot` method of the `Figure` object can be used to partition the `Figure` object display into four subplots.
9. **TRUE/FALSE:** `pandas` uses the value `None` to represent nonnumeric data.
10. **TRUE/FALSE:** OLTP typically utilizes a column-based store.
11. **TRUE/FALSE:** Encoding is also known as serialization or marshalling.
12. **TRUE/FALSE:** Forward compatibility means that newer code can read data that was written by older code.
13. **MULTIPLE CHOICE:** Which of the following is not considered a feature of scalability?
- A. Ability to handle higher traffic volume
  - B. Scaling out
  - C. Chaos testing
  - D. Capacity planning
14. **MULTIPLE CHOICE:** noSQL is used to perform queries on which of the following models:
- A. Relational Model
  - B. Document Model
  - C. Graph Model
  - D. All of the above
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 int
  - D. Automatically convert all elements to float
16. **MULTIPLE CHOICE:** Which of the following is not a feature of OLTP?
- A. Processing several transactions with high availability
  - B. Vectorized processing
  - C. Row-based store
  - D. Data warehousing
17. **MULTIPLE CHOICE:** Graph based models are used to represent:
- A. One-to-one relationships
  - B. Many-to-many relationships
  - C. JSON data
  - D. XML data

- 18. MULTIPLE CHOICE: Hash indexing is:**
- A. Not useful on smaller datasets**
  - B. Works well for very large databases**
  - C. Useful as a building block for more complex indexes**
  - D. Used to maintain a sorted structure on disk**
- 19. MULTIPLE CHOICE: Visualizations (plots) are an important task in data analysis because**
- A. they are part of exploratory process**
  - B. help identify outliers or needed data transformations**
  - C. generating ideas for models**
  - D. all of the above**
- 20. MULTIPLE CHOICE: Which of the following is True about Thrift and Protocol Buffers?**
- A. Both Thrift and protocol buffers do not require a schema for any data that is encoded.**
  - B. Only Thrift buffers require a schema for any data that is encoded.**
  - C. Only Protocol buffers require a schema for any data that is encoded.**
  - D. Both Thrift and Protocol buffer require a schema for any data that is encoded.**
- 21. MULTIPLE CHOICE: Which of the following operations must be done first when performing file processing in Python?**
- A. Open**
  - B. Close**
  - C. Read**
  - D. Write**
- 22. MULTIPLE CHOICE: Python functions:**
- A. Always use call-by-reference parameter passing**
  - B. Can only return one value**
  - C. Return a value of None if nothing is specifically returned**
  - D. All of the above**
- 23. MULTIPLE CHOICE: Python lists can:**
- A. Have values of different types**
  - B. Be directly used in the creation of a pandas Series**
  - C. Dynamically grow or shrink**
  - D. All of the above**
- 24. MULTIPLE CHOICE: Which of the following is considered an advantage of NumPy arrays over Python lists?**
- A. Array elements are not stored in a sequential location in memory**
  - B. Array elements can be referenced using indexing and slicing**
  - C. Array elements must all be of the same type**
  - D. Arrays have built-in functions (methods)**

**25.SHORT ANSWER (50 WORDS MAX):** In your own words, explain why it has been widely accepted that Python is an excellent language for data analytics.

**26.SHORT ANSWER (50 WORDS MAX):** Explain how you would test the code below. In addition, show at least one test case by providing a syntactically correct function call that successfully calls and returns from the function and another test case that calls the function incorrectly and prints an error message.

```
def Stringfun(lst,substr,idx=0):
    if not (isinstance(lst, list) and type(substr) == str and type(idx) == int):
        print("Invalid input: must provide list, string, integer (optional)")
    else:
        flag = True
        for val in lst: # loop to make sure all list elements are strings
            if not type(val) == str:
                flag = False
                break
        if flag == True:
            for i in range(len(lst)):
                tempstr = lst[i]
                newtemp = tempstr[0:idx] + substr + tempstr[idx+len(substr):]
                lst[i] = newtemp[0:len(tempstr)]
```

**27.SHORT ANSWER (50 WORDS MAX):** Using the figure below, in your own words, explain the different operations the cache may have to perform for write client requests.

**Components of data intensive app**



