

CS 544 Exam 1 (16%) - Spring 2023

Instructor: Tyler Caraza-Harter

First/Given Name: _____. Last/Surname: _____

Net ID: _____@wisc.edu

Fill in these fields (left to right) on the scantron form (use pencil):

1. LAST NAME (surname) and FIRST NAME (given name), fill in bubbles
2. IDENTIFICATION NUMBER is your Campus ID number, fill in bubbles
3. Under A of SPECIAL CODES, tell us about the nearest person (if any) to your left. 0=no person to the left in your row, 1=somebody you do not know is there, 2=somebody you do know is there.
4. Under B of SPECIAL CODES, do the same as B, but for the person to your right
5. **Under C of SPECIAL CODES, write 6 and fill in bubble 6.** This is very important!

Make sure you fill all the special codes above accurately in order to get graded.

You have 40 minutes to take the exam. Use a #2 pencil to mark all answers. When you're done, please hand in these sheets in addition to your filled-in scantron. You may not sit adjacent to your friends or other people you know in the class (having only one empty seat is considered "adjacent"). You may only reference your notesheet. You may not use books, your neighbors, calculators, or other electronic devices on this exam. Please turn off and put away portable electronics now.

If multiple answers are correct, choose the best answer.

(Blank Page for You to Do Scratch Work)

Q1. True/False: when a thread is holding a lock during a critical section, the scheduler will never context switch to another thread in the same process.

(A) True (B) False

Q2. What is the following?

```
message {  
    int32 x = 1;  
    int32 y = 2;  
}
```

(A) python class (B) bytecode (C) C struct (D) protocol buffer (E) pytorch array

Q3. Which SQL clause is responsible for projection?

(A) WHERE (B) HAVING (C) GROUP BY (D) SELECT (E) PROJECT

Q4. The prefix of an address identifies the network. What kind of address is it?

(A) IP address (B) MAC address (C) port number (D) host name

Q5. How do you pipe the output from program X to program Y?

(A) $X > Y$ (B) $X \rightarrow Y$ (C) $X | Y$ (D) $X \& Y$ (E) $X \&> Y$

Q6. Parquet is a _____ format, making it most suitable for _____.

(A) column-oriented, analytics
(B) row-oriented, analytics
(C) column-oriented, transactions
(D) row-oriented, transactions

Q7. What is the primary reason RDDs record data lineage?

(A) to compute gradients
(B) to make data scientists accountable
(C) to provide troubleshooting information
(D) to describe how to compute results on demand

Q8. What does the following query produce? Do not make any assumptions about the order of rows in the prizes table.

```
SELECT category, prize
FROM prizes
ORDER BY prize DESC
LIMIT 3
```

- (A) the three biggest prizes
- (B) the first three prizes in the table
- (C) prizes in category 3, from large to small
- (D) prizes in category 3, from small to large

Q9. What can you say about the mutability of DataFrames in pandas and Spark?

- (A) neither are mutable
- (B) only pandas DataFrames are mutable
- (C) only Spark DataFrames are mutable
- (D) both are mutable

Q10. Sequential I/O is generally faster than random I/O for block devices. The difference is larger for what kind of device?

- (A) HDD
- (B) SSD

Q11. Assume x starts at "A" and y starts at 0. After running the following threads together, what is the biggest possible value for y? For simplicity, assume: there is a single CPU core, context switches only occur between lines of Python code, and code/instructions within a single thread are not re-ordered by any system (such as the compiler or CPU).

```
# thread 1
if x == "B":
    y += 1
if x == "C":
    y += 2

# thread 2
x = "C"
x = "B"
```

- (A) 0
- (B) 1
- (C) 2
- (D) 3
- (E) 4

Q12. What type of object is X?

```
Y = X.to("cuda")
```

- (A) numpy array
- (B) PyTorch array
- (C) pandas DataFrame
- (D) Spark DataFrame
- (E) gRPC message

Q13. About how much memory does A use to store elements?

```
A = torch.rand(1024, 256, dtype=torch.float64) # 1024 rows by 256 cols
```

- (A) 8 bytes (B) 64 bytes (C) 256 KB (D) 2 MB (E) 8 MB

Q14. During the first few epochs of optimization, your loss increases, before becoming Inf. What should you do?

- (A) use a smaller learning rate (B) use a bigger learning rate

Q15. A cache contains [1,3,2,4] (assume evictions happen from the front of the list). There is an access, and now the cache contains [3,2,4,1]. We can infer the cache is using a(n) _____ policy and the access was a _____.

- (A) FIFO, miss (B) FIFO, hit (C) LRU, miss (D) LRU, hit

Q16. You have 2 billion floating point operations to do on a device capable of 500 MFLOPS. How many seconds will it take to do the operations?

- (A) 1 (B) 2 (C) 4 (D) 250 (E) 1000

Q17. If you want to estimate how much disk I/O is caused by an HDFS operation, knowing the replication count is most helpful for _____.

- (A) reads (B) writes

Q18. If you want to get a bash shell for debugging purposes inside a container that is already started, what docker command can you use?

- (A) exec (B) logs (C) ps (D) run (E) shell

Q19. For what storage device must a whole block be erased (internally) before a single page can be re-written in the same location?

- (A) HDD (B) SSD

Q20. What does the JVM execute?

- (A) bytecode (B) machine code (C) Java source code (D) Python source code