# CS 544 Exam 1 (15%) - Fall 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 8 and fill in bubble 8**. 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. A client is writing 5 MB of data to a 4x replicated HDFS file. About how much data does the client send over the network?

(A) 1.25 MB     (B) 4 MB     (C) 5 MB     (D) 9 MB     (E) 20 MB

## Q2. What is most likely to DECREASE how much memory a NameNode needs in an HDFS cluster?

(A) adding more DataNodes
(B) increasing the block size for the files
(C) increasing the replication factor for the files

## Q3. Which one is a storage device?

(A) CPU     (B) HDD     (B) NIC     (D) RAM

## Q4. What does "-u" do in the following example?

```
python3 -u server.py &> log.txt
```

(A) unzip the server's output prior to writing it to log.txt
(B) run the command with root (sudo) privileges
(C) force the output to be unbuffered
(D) display the usage information

## Q5. The old version of a protocol buffer for a gRPC call looks like this:

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

A newer version adds a field `int32 z = 3;` If a client using the new version makes a call that includes `z` to a server using the old version, what happens?
(A) the server crashes
(B) the server ignores `z`
(C) the server returns an error to the client
(D) the server replaces `z` with a default value

## Q6. You have a dataset of all emails sent by everybody in your company (the table has one row per email), and you query it like this:

```
SELECT sender, COUNT(*) FROM emails GROUP BY sender;
```

You want to filter the results so that you only have senders with >10K emails. What SQL clause can you add to do this? Do not consider answers that require multiple queries or subqueries.
(A) HAVING     (B) LIMIT     (C) SELECT     (D) WHERE

**Q7. In a Dockerfile, how could you specify that "apt update" should execute during build?**

(A) `EXEC apt update`     (B) `DO apt update`     (C) `CMD apt update`     (D) `RUN apt update`

---

**Q8. For a FIFO cache of size 3, how many hits are there for the following workload?**

1, 2, 3, 3, 2, 1, 4, 1

(A) 1     (B) 3     (C) 4     (D) 5     (E) 0.5

---

**Q9. The prefix of an address identifies the NIC's manufacturer. What kind of address is it?**

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

---

**Q10. You want to connect from a browser on your laptop to Jupyter running in a container on your VM. You take the following steps:**

1. use `-L localhost:3000:localhost:1000` when establishing the SSH tunnel
2. write a command in the Dockerfile to launch Jupyter on port ????
3. use `-p 1000:8888` in the `docker run ...` command
4. enter `http://localhost:3000/` in the browser

What should `????` be in step 2?

(A) 1000     (B) 2000     (C) 3000     (D) 5000     (E) 8888

---

**Q11. After many epochs of optimization, your loss decreased steadily, but too slowly. What should you try changing next time you restart the training loop?**

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

---

**Q12. OLTP databases typically use a _____ data layout.**

(A) row-oriented     (B) column-oriented

---

**Q13. The following code is about to run in a container on a VM. The VM has 4 GB of RAM currently available, the container was launched with `-m 2g`, and `big.file` is 3 GB.**

```
import mmap
with open("big.file", "rb") as f:
    mm = mmap.mmap(f.fileno(), 0, access=mmap.ACCESS_READ)
```

Will memory constraints PREVENT the code from running?
(A) yes     (B) no

**Q14. Two threads are running the following two code snippets concurrently on a machine with one CPU core:**

```
# thread 1
lock.acquire() # A
v1 += 1         # B
lock.release() # C

# thread 2
v1 += 10          # W
lock.acquire()    # X
v2 += 20          # Y
lock.release()    # Z
```

Could some combination of context switches produce the following interleaving?

A, B, W, C, X, Y, Z

(A) yes     (B) no

---

**Q15. In MapReduce, what functions do users write with the following arguments?**

- one key
- one or more values

(A) map functions     (B) shuffle functions     (C) reduce functions

---

**Q16. What does `dis.dis(x)` print?**

(A) bytecode for Python code in x
(B) Python code for bytecode in x
(C) list of attributes/methods for object x

---

**Q17. Assume `count` starts at 6, and two threads are running concurrently. 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 count > 4:
    with lock:
        count -= 4

# thread 2
if count > 3:
    with lock:
        count -= 3
```

What is the smallest possible final count?

(A) -1    (B) 0    (C) 1    (D) 2    (E) 6

**Q18. What is the technical name for a region of code where certain interleavings with other code would be a problem?**

(A) context switch     (B) thread-safe zone     (C) collision area     (D) critical section

**Q19. For an LRU cache of size 3, how many hits are there for the following workload?**

1, 1, 2, 3, 4, 1, 2, 3, 4

(A) 1     (B) 3     (C) 4     (D) 5     (E) 0.5

**Q20. How do you redirect ONLY the standard output from program X to file Y?**

(A) X > Y     (B) X -> Y     (C) X | Y     (D) X & Y     (E) X &> Y