DATA 13600 Autumn 2025
Final Exam - rejected questions

University of Chicago - W. Trimble

December 5, 2025

Fill in the circles, for goodness sake. Some questions have one best answer, some more than one.

- No additional materials, electronic or otherwise, may be used in this midterm. Your work is your own, and any attempt to copy or share exam materials is not allowed
- Partial credit will be given for responses that more than one answer when only one answer is sought. Additional answers "just in case" will be penalized, even if one of them is right.
- Instructional staff will not answer any questions during the exam. If you find yourself stuck write down a reasonable assumption to move on. If you don't like an answer, you can argue for your answer after the exam is marked. (Note, changes in the accepted answers will take effect for everyone.)
- Write your initials on each physical page of your exam. (The reason for this is left as an exercise for the student.)

Name: _____

Student ID number: _____

1. Which of the following is a key characteristic of an OLTP system?
   - ◯ Designed for complex queries and multidimensional analysis
   - ◯ Optimized for large-scale data aggregation
   - ◯ Handles frequent and small-scale transactions
   - ◯ Built for historical data storage and retrieval
   - ◯ Built to allow analysts to perform large queries
2. What type of database operations are most common in OLAP (Online Analysis Processing) systems?
   - ◯ Insert and update operations
   - ◯ Read-intensive operations like aggregations and filtering
   - ◯ Real-time data entry and deletion
   - ◯ Concurrent transactions with low latency
   - ◯ Frequent, simple transactions
3. In 2017, a security team at Adobe accidentally published the team's private key, likely by hitting a button labeled "Attach all?". The public key was promptly revoked. What bad things can happen now?
   - ◯ Anyone with a copy of the private key can decrypt historical messages sent by the Adobe team.
   - ◯ Anyone with a copy of the private key can verify historical messages signed by the Adobe team.
   - ◯ Anyone with a copy of the private key can decrypt historical messages sent to the Adobe team.
   - ◯ Anyone with a copy of the private key can verify historical messages to the Adobe team.
   - ◯ Anyone with a copy of the private key can verify historical messages signed by the (now invalid) public key.

Initials: _____

4. What is the primary purpose of proof-of-work in a blockchain system?

   ◯ To increase transaction speed

   ◯ To prevent bad actors from easily tampering with the blockchain

   ◯ To allow miners to earn rewards without expending resources

   ◯ To centralize control of the network

   ◯ To allow pre-computation of potentially useful hashes in the future

5. Cloud orchestration systems (amazon, Microsoft, google, github) tend to have a public key management interface. Users can upload public keys. What do these companies do with user's keys?

   ◯ The platforms also collect users' private keys so that the cloud providers have a backdoor into their own cloud resources.

   ◯ The platforms prefer public key / private key encryption because it is less computationally intensive than symmetric encryption.

   ◯ The platforms publish the users public keys to allow verification of identity.

   ◯ The platforms install the public keys on cloud resources, allowing the users to authenticate and control the resources they've rented.

   ◯ The platforms collect a large sample of keys so that they can research the state of cloud security protocols.

6. If you were running a cloud computing platform called Gingko, and you wanted a backdoor into the virtual machines your clients were using. (Let's say everyone's friendly, and it's for support purposes, like your clients sometimes need help fixing their cloud deployment). Which of these would accomplish your goal? (I have seen precisely this technique used on an academic cloud.)

   ◯ Install gingko's public key alongside the user's public key on each new instance when the instances are initialized.

   ◯ Install gingko's private key alongside the user's private key on each new instance when the instances are initialized.

   ◯ Install both gingko's private and public keys on each new instance.

   ◯ Start an SSHD server on each new instance equipped with the user's public key.

   ◯ Configure the virtual machines to auto-update and use the default username + password to get in.

Initials: _____

7. One system I was using once had a blacklist of "bad" ssh private keys. Keys that were stored unencrypted (no passphrase, anyone who reads the key file can use it) made the list, and I presume the IT team tried to crack the passphrases of everyone's SSH keys and blacklisted the keys if they succeeded. Why does IT not let me use SSH keys with dictionary words, addresses, words followed by numbers, or lightly modified dictionary words?

8. What is not a benefit of containers (like docker and singularity)?
   - ◯ Containers facilitate deployment and testing by packaging software stacks together
   - ◯ Containers save on resource utilization, CPU, memory, and disk space, over native apps.
   - ◯ Containers allow static environments to be preserved, useful for regression testing.
   - ◯ Containers facilitate testing in different operating system versions.
   - ◯ Containers have some security benefits, disallowing apps in containers full access to the host system.

9. One-time pad. Alice has a binary message A. Alice generates a random sequence of bits, the key, K, of the same length as A. Which of the following describes the encryption step of a symmetric stream cipher in terms of bitwise operations?
   - ◯ A AND K
   - ◯ A OR K
   - ◯ A AND NOT K
   - ◯ A XOR K
   - ◯ NOT(NOT A OR NOT K)

10. Given the cyphertext A' and the key K, how does Bob decrypt the message?
   - ◯ A′ AND K
   - ◯ A′ OR K
   - ◯ A′ AND NOT K
   - ◯ A ′ XOR K
   - ◯ NOT(NOT A′ AND NOT K)

Initials: _____

11. If you have sufficient computing power to guess all possible values of the key K (2^K of them) and you have access to the encrypted message A', can you recover the message? Briefly (in perhaps one sentence) explain why or why not.

12. What is the primary benefit of using Dask for parallel execution?

    ◯ It automatically converts Python code to machine code.

    ◯ It enables the execution of tasks in parallel across multiple CPUs or clusters.

    ◯ It improves single-threaded performance by optimizing for GPU usage.

    ◯ It provides data-informed vector representations of linguistic tokens whose proximity is related to their usage and meaning.

    ◯ It replaces the need for all other Python libraries.

13. What is the average time complexity of searching for an element in a well-implemented hash table in terms of the number of elements stored?

    ◯ $O(1)$

    ◯ $O(\log n)$

    ◯ $O(n)$

    ◯ $O(n^2)$

    ◯ $O(n \log n)$

14. Using hash tables to store data trades off what resources?

    ◯ space - time : faster lookups but more storage

    ◯ time - space : slower lookups but less storage

    ◯ money - time : renting more resources to get results faster

    ◯ time - money : renting out spare computing power to fund, say, education

    ◯ initial time - lookup time : fast lookup in exchange for slow indexing

15. Why is lazy computing employed in large-scale data processing systems?

    ◯ Lazy computing is faster than eager computing.

    ◯ Lazy computing is more ethically defensible than greedy computing.

    ◯ Lazy computing is more parallelizable than eager computing.

    ◯ Lazy computing allows the system to better optimize the commutation by "planning" before executing it on a potentially large data volume.

    ◯ Lazy computing is a theoretical concept that is never actually used in practice.

 Initials: _____

Here is a schema for a twitter-like social media system: Users create messages, called tweets, follow other users, and have the option to attach media to their messages.

Users Table

Column Name	Data Type	Constraints
UserID	INT	Primary Key, Auto Increment
Username	VARCHAR(50)	Unique, Not Null
Email	VARCHAR(100)	Unique, Not Null
PasswordHash	VARCHAR(255)	Not Null
CreatedAt	DATETIME	Default: CURRENT_TIMESTAMP
Bio	TEXT	Nullable
ProfilePicture	VARCHAR(255)	Nullable

Tweets Table

Column Name	Data Type	Constraints
TweetID	INT	Primary Key, Auto Increment
UserID	INT	
Content	TEXT	
ParentTweetID	INT	
CreatedAt	DATETIME	
LikesCount	INT	
RetweetsCount	INT	

Follows Table

Column Name	Data Type	Constraints
FollowID	INT	Primary Key, Auto Increment
FollowerID	INT	Foreign Key → Users(UserID), Not Null
FollowedID	INT	Foreign Key → Users(UserID), Not Null
FollowedAt	DATETIME	Default: CURRENT_TIMESTAMP

16.

Initials: _____

17. In a twitter-like social media management system, we have three tables with partial schemas described above. What foreign-key constraints are appropriate and necessary in the Tweets table?

    ◯ UserID references Users, TweetID references Follows

    ◯ UserID references Users, TweetID references Tweets

    ◯ UserID references Users, TweetID references Follows, ParentTweetID references Follows (NULLABLE)

    ◯ UserID references Users, ParentTweetID references Tweets (NULLABLE)

    ◯ UserID references Users, ParentTweetID references Follows (NULLABLE), LikesCount references INT

18. In a twitter-like social media management system, we have a table with two very similar columns, Follows.FollowerID and Follows.FollowedID. Will this cause a problem?

19.
```
#!/bin/bash
echo "a" > a
while [ 1 ]   # continue if there is no error
do
    cat a  a  > a2
    cat a2 a2 > a
done
```

    On a digital archeology dig, you find the above script without a lot of clues as to its purpose.

    What does this script do? Why would someone what to do this?

Initials: _____

20. Suppose you are asked to protect the comments section of some digital asset from abuse, so you are tasked with implementing a content moderator. One approach is to fine-tune an LLM that is pre-trained on vast corpora of text from the internet. What additional training data do you need to get started? In other words, what kind of data are you going to search for or commission?

End practice exam 3.

Initials: _____