CS2640 Midterm Exam (Practice)

Modern (Computer) Storage Systems
Spring 2026

Duration	Format	Total	Date
75 minutes	Closed-book	100 points	March 23, 2026

Instructions

- Answer all questions. Show your reasoning for short-answer and design questions.
- Write clearly and justify performance, reliability, and design tradeoffs with storage-specific terminology.
- For multiple-choice questions, circle exactly one answer.
- This exam covers material taught before the March 23 midterm: HDDs, SSDs, block storage, file systems, distributed storage, storage performance, caching, and storage reliability/data protection.

Part I — Multiple Choice (30 points, 3 points each)

1. Which statement best explains why HDD random reads are much slower than sequential reads?

A. HDDs compress sequential blocks more aggressively

B. Random reads incur seek and rotational latency

C. Sequential reads require more ECC checks

D. Random reads are limited primarily by DRAM bandwidth

2. In NAND flash SSDs, an overwrite to an existing logical page usually cannot be done in place because:

A. flash pages cannot be read after being written

B. writes must be aligned to cache-line boundaries

C. erasures happen at block granularity, not page granularity

D. the FTL only supports sequential workloads

3. The Flash Translation Layer (FTL) primarily exists to:

A. compress file contents before persistence

B. map logical addresses to physical flash locations

C. replace RAID functionality inside the SSD

D. provide application-level durability semantics

4. Which workload benefits most from append-oriented (log-structured) storage layout?

A. tiny random overwrites across the full address space

B. workloads that can write sequentially and avoid internal write amplification

C. read-only archival data with no metadata

D. CPU-bound in-memory analytics

5. A key goal of a filesystem journal is to improve:

A. sequential bandwidth only

B. crash consistency of metadata updates

C. user-level encryption throughput

D. deduplication ratio

6. Compared with a classic update-in-place filesystem, a log-structured filesystem most directly tries to improve write performance by:

A. converting many small writes into large sequential writes

B. disabling metadata

C. avoiding all garbage collection

D. storing files only in fixed-size extents

7. In a typical storage stack, which layer sits closest to the application?

A. device firmware

B. block layer

C. filesystem

D. flash controller

8. In distributed storage, replication is commonly used primarily to improve:

A. only throughput, never availability

B. availability and fault tolerance

C. only metadata lookup speed

D. only compression efficiency

9. A cache replacement policy is evaluated mainly by its effect on:

A. packet loss rate

B. miss ratio under a workload

C. RAID rebuild bandwidth only

D. inode size

10. RAID-5 protects against:

A. no failures

B. one disk failure

C. two disk failures

D. any number of correlated failures

Part II — Short Answer (50 points, 10 points each)

11. HDD vs SSD. Compare HDDs and SSDs along these four axes: access latency, sequential bandwidth, behavior under random I/O, and cost/capacity trends.

12. SSD internal behavior. Explain the relationship among out-of-place writes, garbage collection, and write amplification in SSDs. Why can performance become unstable as the device fills up?

13. Filesystem design tradeoff. Compare update-in-place filesystems and log-structured filesystems. Give one advantage and one drawback of each.

14. Distributed storage semantics. What are two major design challenges in distributed storage systems that do not arise, or arise less severely, in a single-node local filesystem? Briefly explain each.

15. Reliability metrics. Why can a storage system with redundancy still have poor real-world reliability? Give at least two reasons, such as correlated failures, rebuild windows, latent sector errors, or operational mistakes.

Part III — Quantitative / Reasoning (20 points)

16. RAID reasoning. You have 8 disks, each with annualized failure rate 2%.

(a) Compare RAID-0, RAID-1 (4 mirrored pairs), and RAID-5 at a high level in terms of usable capacity, performance tradeoffs, and failure tolerance.

(b) Which configuration is riskiest for data loss and why?

(c) During rebuild after a disk failure, why does system risk often increase rather than decrease?