

Midterm 2 (Final) Study Guide

These are just representative questions of the kinds of topics you should know. You should study all the reading material and the supplemental books and site links in them.

1. What is an **application database** and what is an **integration database**? What are some concerns that one type has that the other does not?
2. What are examples of large data sets that came about with the increase in scale of modern large web properties?
3. What is the **Impedance Mismatch**?
4. What are the main two ways of dealing with an increase in application traffic? What are the tradeoffs?
5. What does the term NoSQL database refer to? What led to the emergence of these databases?
6. What are the different models of NoSQL databases, and what characterizes each?
7. What does the term **polyglot persistence** mean?
8. What does the term **aggregate** refer to in the context of NoSQL databases? Provide some examples.
9. Give examples of aggregate-ignorant data models.
10. How do key-value stores and document databases differ in their approach to aggregate structures?
11. What is schemaless design? What are its advantages and drawbacks?
12. What is **sharding** and what is **replication**? How do they differ?
13. How does sharding behave with regards to performance and reliability?
14. Describe the two main replication models, their main advantages and disadvantages, and main use cases for each.
15. Peer-to-peer replication suffers from the possibility of inconsistent writes. Explain what that means and how it can be avoided.
16. Describe examples where inconsistent writes can be tolerated by the system.
17. Describe the **pessimistic** and **optimistic** approaches to conflict resolution.
18. Read-write conflicts can suffer from both *logical consistency* and *replication consistency*. Explain the difference between the two.
19. What does the **CAP theorem** say? What is its importance? Describe the meaning of the terms involved.
20. Describe different **quorum systems** with $N=3$ and with $N=5$, and their tradeoffs in terms of the CAP theorem (different values of W and R provide different features).
21. How can we avoid write-write conflicts in a quorum system? How can we avoid read-write conflicts?
22. What data model does MongoDB employ?
23. Describe the key possible stages in MongoDB's **aggregation** framework, and what each of them does. List at least 5 different stages.
24. What are the key stages in a map-reduce setting? Describe what happens in each stage.
25. Suppose we have a database of gpa scores for men and women. We want to

- compute the average gpa for the men and for the women. What would happen in each stage of a map-reduce solution to this problem?
26. What does **web scraping** mean? How does it differ from accessing a **web service**?
 27. What do the C-I-A initials stand for in the context of security? Also provide concrete examples.
 28. What are the four groups that attack methods are usually classified in? Also provide concrete examples.
 29. What is **identification** and what is **authentication**? What is the difference between the two?
 30. What do we mean when we refer to **authentication factors**? What are the different factors? What are examples of each factor?
 31. What is **multi-factor authentication**? Provide concrete examples.
 32. What is **authorization**?
 33. What is the **principle of least privilege**? Also provide concrete examples.
 34. What are the four different kinds of **access control**? Provide concrete examples.
 35. What are **access control lists**? What are **capabilities**? How do they differ?
 36. Describe in broad terms what **symmetric key ciphers** and **asymmetric key ciphers** are and what are the advantages and limitations of each.
 37. What is a cryptographic hash function? What are some of its key properties?
 38. Describe how **digital signing** happens, and what other cryptographic tools it uses.
 39. Describe what a **digital certificate** is, how it works, and what problem it solves.
 40. What are some important precautions to take when implementing a password-protected website?
 41. What is a **salt**? How is it used? Why is it used?
 42. The server digitally signs the sessionID when sending it to the user. Why?