



↑ Your Yale netid ↑

# Columnar Databases

April 23, 2025

## Instructions

- Do not flip over this page or read the questions on the other side of this page until the start of class.
- Your quiz will be collected after ten minutes.
- Write your netid *clearly* at the top right.
- *Unless otherwise stated*, each question has one correct answer.
- Fill in the appropriate bubble below. I will grade nothing but these bubbles. If you need to change an answer please indicate your final answer clearly.
- If a question stinks, Kyle will fix it later. I will not answer questions during the quiz.
- This quiz is closed book, closed device. You can only use your own meat computer.
- When you're done, *raise your hand* and an instructor will collect it.

## Multiple Choice Answers



1. (A) (B) (C) (D) (E)
2. (A) (B) (C) (D) (E)
3. (A) (B) (C) (D) (E)
4. (A) (B) (C) (D) (E)
5. (A) (B) (C) (D) (E)
6. (A) (B) (C) (D) (E)
7. (A) (B) (C) (D) (E)
8. (A) (B) (C) (D) (E)
9. (A) (B) (C) (D) (E)
10. (A) (B) (C) (D) (E)

-

-

-

-

-

-

-

## Handwritten SQL Query



1. What is the primary benefit of using a column-oriented DBMS for analytics?
  - A. Faster write speeds if columns are numeric and updated frequently
  - B. Improved data normalization for columns that depend on each other
  - C. Faster read speeds for queries involving a subset of columns
  - D. Enhanced transaction concurrency, regardless of column types
2. Why might someone choose DuckDB for their project?
  - A. It is primarily designed for OLTP systems
  - B. It supports real-time data updating
  - C. It operates as an embedded SQL database system with zero configuration
  - D. It uses a row-oriented storage approach that helps with write-heavy workloads
3. What is a significant advantage of the Parquet file format over CSV?
  - A. Better support for text data
  - B. Faster write speeds for numeric data
  - C. More efficient storage and faster data retrieval
  - D. Higher data integrity using cryptographic hashing
4. Which of the following pairs describes MongoDB and DuckDB accurately?
  - A. columnar database, document-oriented database
  - B. document-oriented database, columnar database
  - C. both are columnar databases
  - D. both are document-oriented databases
5. What does Parquet use to reduce the amount of data scanned during queries?
  - A. Partitioning by hash of primary key
  - B. Metadata to skip irrelevant data chunks
  - C. Foreign key constraints, particularly for joins
  - D. Implicit WHERE clauses that filter out unnecessary data
6. What technique does Parquet use to handle columns with many repeating values?
  - A. B-tree indexing
  - B. Bitmap indexing
  - C. Run-length encoding
  - D. Hash mapping
7. Which of the following is a common use case for DuckDB?
  - A. OLTP workloads
  - B. OLAP workloads
  - C. Session storage for web applications
  - D. Social network/graph storage
8. Which of the following JOIN operations would likely produce the most rows?
  - A. INNER JOIN
  - B. LEFT JOIN
  - C. CROSS JOIN
  - D. FULL OUTER JOIN
9. The Bitcoin blockchain is most similar to which of the following systems?
  - A. DuckDB
  - B. MongoDB
  - C. PostgreSQL
  - D. Git
10. You export your data to a file for analysis. Which format is likely smallest on disk?
  - A. CSV
  - B. Parquet
  - C. JSON

### Answer Key

1	C
2	C
3	C
4	B
5	B
6	C
7	B
8	C
9	D
10	B