

90

Data Engineer

Interview Questions

Seen in interviews at FAANGs, startups and consulting firms



Coding and Algorithms

1. Implement a function to find the largest sum of contiguous subarray.
2. Write a function to determine if a string is a palindrome.
3. How would you find the intersection of two arrays?
4. Implement a binary search algorithm.
5. Write a function to check if two strings are anagrams.
6. Implement the quicksort algorithm.
7. How would you detect a cycle in a linked list?
8. Write a function to merge two sorted linked lists.
9. Implement a function to find the nth Fibonacci number.
10. How would you reverse a linked list?
11. Write a function to find the longest common prefix among an array of strings.
12. How would you implement a stack using two queues?
13. Write a function to perform a breadth-first search (BFS) on a graph.
14. Implement the depth-first search (DFS) algorithm on a graph.
15. How would you find the lowest common ancestor in a binary search tree?
16. Write a function to sort an array of integers using the merge sort algorithm.
17. How would you implement a hash table from scratch?
18. Write a function to find the minimum number of steps to reduce a number to zero by subtracting its largest digit.
19. Implement an LRU (Least Recently Used) cache.
20. How would you find the longest increasing subsequence in an array?

System Design

1. Design a URL shortening service like Bitly.
2. How would you design a scalable web crawler?
3. Design a distributed file storage system.
4. How would you design a real-time chat application?
5. Design a recommendation system for an e-commerce platform.
6. How would you design a logging system that can handle millions of log entries per second?
7. Design a system for real-time analytics on streaming data.
8. How would you design a scalable API rate limiter?
9. Design a ride-sharing service like Uber.
10. How would you design a social media news feed?
11. Design a search engine that indexes and searches web pages.
12. How would you design a scalable notification system?
13. Design a system for managing and processing large volumes of image uploads.
14. How would you design a distributed cache system?
15. Design a video streaming service like YouTube.
16. How would you design a high-frequency trading platform?
17. Design a real-time collaborative document editing system.
18. How would you design a system to handle financial transactions and detect fraud?
19. Design an online multiplayer gaming platform.
20. How would you design a scalable email delivery system?

Data Modeling and SQL

1. Write a SQL query to find the second highest salary in a table.
2. Design a schema for a blogging platform.
3. How would you model a many-to-many relationship in a relational database?
4. Write a SQL query to find the top 5 customers by total purchase amount.
5. Design a schema for an online retail store.
6. How would you optimize a slow-running SQL query?
7. Write a SQL query to find employees who have not made any sales in the last month.
8. Design a database schema for a social networking site.
9. How would you handle data versioning in a database?
10. Write a SQL query to find the average salary of employees by department.
11. Design a schema for a task management application.
12. How would you implement a full-text search in a database?
13. Write a SQL query to find duplicate records in a table.
14. Design a schema for a library management system.
15. How would you handle database migrations in a production environment?
16. Write a SQL query to find the median value of a numeric column.
17. Design a schema for a movie recommendation system.
18. How would you ensure data integrity in a relational database?
19. Write a SQL query to pivot rows into columns.
20. Design a schema for a ride-sharing application.

Data Warehousing & ETL

1. What is a data warehouse, and how does it differ from a traditional database?
2. Explain the concept of OLAP vs. OLTP.
3. Describe the differences between star schema and snowflake schema.
4. How do you design a data warehouse for a retail company?
5. What are fact and dimension tables? Provide examples.
6. How do you ensure data consistency in a data warehouse?
7. Explain the concept of slowly changing dimensions (SCD) and their types.
8. What is a data mart, and how does it relate to a data warehouse?
9. How would you implement a surrogate key in a data warehouse?
10. Describe the process of data warehouse partitioning.
11. What are materialized views, and how are they used in a data warehouse?
12. Explain the concept of data aggregation and summarization.
13. How do you handle incremental data loading in a data warehouse?
14. Describe the architecture of a data warehouse you have worked on.
15. How do you optimize query performance in a data warehouse?
16. What are the best practices for data warehouse schema design?
17. Explain the ETL process in the context of data warehousing.

Big Data Technologies

1. Explain the architecture of Hadoop.
2. How does Apache Spark differ from Hadoop MapReduce?
3. What is a data lake, and how is it used?
4. How would you design a data pipeline using Apache Kafka?
5. Describe the process of deploying a model on AWS SageMaker.
6. How do you use AWS S3 for data storage in ML Ops?
7. Explain the concept of serverless computing and its advantages.
8. How would you use Apache Airflow to manage workflow orchestration?
9. What are the key features of Google Cloud AI Platform?
10. How do you implement continuous integration and deployment (CI/CD) for data pipelines?
11. Explain the concept of Infrastructure as Code (IaC).
12. How would you use Docker and Kubernetes for container orchestration?
13. How does Kafka work?

Join DataInterview.com

Access courses, mock interviewing and private community with coaches and candidates like you



Created by interviewers at top companies like Google and Meta

