78 Data Warehousing Interview Questions to Ask Candidates

Questions

1. Can you explain what a data warehouse is and its primary purpose?
2. What are the main differences between OLTP and OLAP systems?
3. Describe the ETL process and its importance in data warehousing.
4. What are some common challenges faced during data warehouse implementation?
5. Can you explain what a star schema is and its components?
6. How do you ensure data quality in a data warehouse?
7. What are some common data warehouse architectures?
8. How do you handle slowly changing dimensions (SCD) in a data warehouse?
9. What role does metadata play in a data warehouse?
10. Can you describe the difference between a data warehouse and a data lake?
11. What is dimensional modeling and why is it important in data warehousing?
12. Can you explain the concept of a fact table and provide an example?
13. How would you approach data cleaning in a warehouse environment?
14. What is the difference between a dimension table and a fact table?
15. Can you describe what a data mart is and how it relates to a data warehouse?
16. What are some common data warehouse performance optimization techniques?
17. How would you handle missing data in a data warehouse?
18. Can you explain what a surrogate key is and why it's used in data warehousing?
19. What is data denormalization and why is it often used in data warehouses?
20. How would you ensure data consistency across different sources in a warehouse?
21. What is the purpose of a staging area in the ETL process?
22. Can you explain what a slowly changing dimension type 2 (SCD2) is?
23. How would you handle real-time data integration in a data warehouse?
24. What is the difference between a conformed dimension and a non-conformed dimension?
25. Can you describe what a fact constellation schema is?
26. How would you approach data archiving in a data warehouse?
27. What is the purpose of indexing in a data warehouse?
28. Can you explain what a junk dimension is and when you might use it?
29. How would you handle data versioning in a data warehouse?
30. What is the difference between a data warehouse and a operational data store (ODS)?
31. What strategies would you use to maintain data security in a data warehouse?
32. Can you explain what a data warehouse appliance is and when it might be used?
33. How do you handle data integration from multiple sources in a data warehouse?
34. What methods would you use to optimize query performance in a data warehouse?
35. Can you describe the differences between a logical and a physical data model?
36. How would you approach capacity planning for a data warehouse?
37. What is the role of a data warehouse in business intelligence (BI)?
38. How do you ensure the scalability of a data warehouse?
39. Can you explain the concept of data warehousing as a service (DWaaS)?
40. How would you approach troubleshooting performance issues in a data warehouse?
41. What strategies would you use to manage large datasets in a data warehouse?
42. Can you describe your experience with data warehouse automation tools?
43. How have you implemented data lineage tracking in previous data warehousing projects?
44. What is your approach to managing schema changes in a data warehouse?
45. Can you explain how you would use machine learning techniques to optimize data warehousing processes?
46. Describe a time when you had to reconcile data discrepancies in a data warehouse. How did you approach it?
47. How do you ensure compliance with data governance policies in a data warehouse?
48. Can you discuss your experience with cloud-based data warehouse solutions?
49. What methods do you use for real-time analytics in a data warehouse environment?
50. Describe how you would optimize the storage and retrieval of semi-structured data in a data warehouse.
51. Can you explain the role of data warehousing in supporting predictive analytics?
52. How do you manage and monitor data warehouse performance over time?
53. Describe your experience with data warehouse migration projects.
54. What are your best practices for designing scalable data warehouses?
55. How do you handle data warehouse documentation and knowledge transfer within a team?
56. Can you explain the difference between a conceptual, logical, and physical data model?
57. What is a snowflake schema, and how does it differ from a star schema?
58. How do you approach designing a data model for a new data warehouse?
59. Can you explain the role of normalization in data modeling?
60. What is a data modeler's role in a data warehousing project?
61. How do you handle data redundancy in a data model?
62. What are some common pitfalls in data modeling, and how do you avoid them?
63. How do you ensure the scalability of a data model in a data warehouse?
64. Can you explain the difference between ETL and ELT?
65. What are some common tools used in ETL processes?
66. How do you handle data validation during the ETL process?
67. What strategies do you employ for error handling in ETL processes?
68. How do you ensure data security during the ETL process?
69. What is your approach to optimizing ETL performance?
70. How do you manage ETL process scheduling and monitoring?
71. Imagine you have to integrate a new data source into an existing data warehouse. How would you approach this task?
72. How would you handle a situation where a critical ETL job fails just before a major report is due?
73. Describe a time when you had to reconcile data discrepancies in a data warehouse. How did you approach it?
74. How would you handle a situation where the data warehouse's performance is degrading due to an increase in data volume?
75. How would you ensure data security in a data warehouse environment?
76. Can you describe a challenging data migration project you worked on and how you managed it?