# 72 ETL Testing Interview Questions to Ask Your Candidates

## Questions

1. Can you explain the difference between ETL and ELT? When would you choose one over the other?

2. How would you approach testing data transformation rules in an ETL process?

3. What strategies do you use to validate data integrity during the extraction phase?

4. Describe a situation where you encountered data quality issues during ETL testing. How did you address it?

5. How do you ensure the completeness of data loaded into the target system?

6. What tools have you used for ETL testing, and what are their strengths and limitations?

7. How would you test the performance of an ETL process?

8. Can you explain the importance of data reconciliation in ETL testing?

9. How do you handle testing of incremental loads versus full loads in an ETL process?

10. What approach would you take to test error handling and logging in an ETL job?

11. What is the primary goal of ETL testing?

12. How do you validate that data transformation rules are correctly implemented?

13. What steps would you take to verify data consistency between source and target systems?

14. How would you manage and document test cases for an ETL process?

15. What is your approach to handling large data volumes during ETL testing?

16. Can you describe how you would test ETL processes for various data formats?

17. How do you ensure the accuracy of aggregated data in an ETL process?

18. What measures do you take to handle and test ETL processes under different data quality conditions?

19. How would you design a test plan for an ETL process that involves multiple data sources and complex transformations?

20. Can you explain the concept of slowly changing dimensions in data warehousing and how it impacts ETL testing?

21. What techniques do you use to test data quality in real-time streaming ETL processes?

22. How would you approach testing an ETL process that involves sensitive data subject to compliance regulations?

23. Describe your strategy for testing ETL jobs that involve data type conversions and handling of null values.

24. What methods do you employ to test the idempotency of ETL processes?

25. How do you test for data lineage in complex ETL workflows?

26. Can you explain how you would test ETL processes that involve data deduplication?

27. What approach do you take when testing ETL processes that involve time-based or event-driven data?

28. How would you test the scalability of an ETL process?

29. Describe your approach to testing ETL processes that involve data masking or encryption.

30. How do you ensure referential integrity is maintained during the ETL process?

31. What strategies do you use to test ETL processes that involve data archiving and purging?

32. How would you test an ETL process that involves merging data from multiple source systems with different update frequencies?

33. Can you explain your approach to testing ETL processes that involve complex business logic and calculations?

34. How would you approach testing an ETL process that involves data from multiple legacy systems with inconsistent data formats?

35. Can you explain how you would test for data lineage in a complex ETL workflow with multiple intermediate stages?

36. How would you design a test strategy for an ETL process that involves sensitive data subject to compliance regulations like GDPR or HIPAA?

37. How would you test the fault tolerance and recovery mechanisms of a distributed ETL system?

38. How would you approach testing an ETL process that involves real-time data streaming and complex event processing?

39. How would you test the impact of schema changes in source systems on an existing ETL process?

40. How would you verify the accuracy of calculated fields in the transformed data?

41. What methods do you use to detect and handle duplicate records during the ETL process?

42. How do you ensure that date formats are consistent across different source systems and the target database?

43. Can you explain your approach to validating complex business rules during data transformation?

44. How would you test for data truncation issues when moving data between systems with different field length constraints?

45. What techniques do you employ to validate the referential integrity of data across multiple tables after the load process?

46. How do you handle and validate currency conversions in an ETL process involving international data?

47. Can you describe your method for verifying the correct application of data cleansing rules during transformation?

48. How would you approach validating aggregated data against source data in an ETL process?

49. What strategies do you use to ensure that all required fields are populated in the target system?

50. How do you validate the correct handling of null values and default values during the ETL process?

51. Can you explain your approach to testing data type conversions, especially for complex data types like JSON or XML?

52. How would you approach testing a complex data transformation involving multiple lookups and joins?

53. Can you explain how you would validate the correct application of business rules during data transformation?

54. What methods do you use to test for data loss or corruption during transformation processes?

55. How would you verify the accuracy of calculated fields or derived data in the transformed dataset?

56. Can you describe your approach to testing data transformations that involve time-based or historical data?

57. How do you ensure that data type conversions are handled correctly during transformation, especially for complex data types?

58. What techniques do you use to validate the correct handling of null values and default values in data transformations?

59. How would you test transformations that involve data normalization or denormalization?

60. Can you explain your approach to testing transformations that require currency or unit conversions?

61. How do you verify that data transformations maintain referential integrity across related tables?

62. What methods do you employ to test transformations involving complex string manipulations or regular expressions?

63. How would you approach testing data transformations that involve pivoting or unpivoting data structures?

64. Describe a time when you had to troubleshoot a failed ETL job. What steps did you take to identify and resolve the issue?

65. How would you handle a scenario where the ETL process is causing performance bottlenecks in the production environment?

66. Can you explain how you would test for and manage data skew in an ETL process?

67. Imagine you are given a new ETL pipeline to test that has no documentation. How would you approach this task?

68. How would you test an ETL process that needs to handle data from an API that has rate limits?