

# ETL Interview Questions Guide



give me the interview questions on ETL - Extract Transform and load



Here are some common interview questions related to **ETL (Extract, Transform, Load)** processes that you may encounter:

## General ETL Questions

1. What is ETL, and why is it important in data processing?
2. Can you explain each phase of ETL: Extraction, Transformation, and Loading?
3. What are the common tools used for ETL? Can you name a few popular ones?
4. How do you handle data extraction from different source systems, such as databases, APIs, and flat files?
5. What are the challenges in the data transformation phase, and how do you overcome them?
6. What is data cleaning, and how do you handle inconsistent data in the ETL process?
7. What is incremental load, and how does it differ from a full load in ETL?
8. How do you ensure data accuracy and integrity during ETL?
9. How would you optimize an ETL process for large volumes of data?
10. Can you explain the difference between ETL and ELT?

## Technical ETL Questions

11. What are slowly changing dimensions (SCDs) in data warehousing, and how do you implement them in ETL?
12. How do you handle errors and exceptions in an ETL process?
13. What is a staging area, and why is it important in ETL?
14. How would you deal with duplicate records during the ETL process?
15. Can you explain the concept of data lineage and its importance in ETL?
16. How do you ensure that the ETL jobs are scheduled and executed on time?
17. What is partitioning, and how can it improve ETL performance?
18. How would you handle a data load failure halfway through a process?
19. Can you explain the use of lookup and merge operations in ETL?
20. What strategies would you implement to handle changes in the source schema?

## Advanced ETL Questions

21. How do you implement parallel processing in ETL to improve performance?
22. Can you explain the concept of ETL pipelines in cloud platforms like AWS Glue, Azure Data Factory, or Google Dataflow?
23. What is the role of metadata in the ETL process?
24. How would you design an ETL pipeline to handle real-time data processing?
25. What is the difference between batch processing and real-time processing in ETL?

26. How do you handle schema evolution in ETL workflows?
27. Explain how ETL processes support data governance and compliance.
28. What is surrogate key generation, and how do you handle it during the transformation phase?
29. How would you handle unstructured or semi-structured data (like JSON, XML) in an ETL process?
30. Can you walk through the architecture of a typical ETL system?

These questions cover various aspects of ETL, from basic concepts to more advanced, real-world scenarios. Preparing answers with examples from your experience will help demonstrate your practical knowledge of ETL systems.

ChatGPT can make mistakes. Check important info.