Etl testing interview questions and answers for freshers

1. What is ETL testing?
2. Why is ETL testing important?
3. What are the main types of ETL testing?
4. Explain the ETL process.
5. What are the different stages of ETL testing?
6. What are the key challenges in ETL testing?
7. What tools are commonly used for ETL testing?
8. How do you perform data completeness testing in ETL?
9. What is data transformation testing?
10. Explain the concept of data reconciliation in ETL testing.
11. What is data quality testing in ETL?
12. What is initial load and incremental load testing in ETL?
13. How do you perform null validation in ETL testing?
14. What is source-to-target data mapping?
15. What is surrogate key validation in ETL?
16. How would you handle duplicate records in ETL testing?
17. What is end-to-end ETL testing?
18. Explain data integrity testing in ETL.
19. What are the common ETL testing scenarios?
20. What is a test harness in ETL testing?
21. What is a dimension table in ETL?
22. What is a fact table in ETL?
23. Explain the importance of data validation in ETL testing.
24. What is metadata testing in ETL?
25. How do you validate data transformation logic in ETL testing?
26. What is performance testing in ETL?
27. What is exception handling in ETL testing?
28. How do you handle incremental data load validation in ETL?
29. What are some common data validation SQL queries used in ETL testing?
30. What is schema validation in ETL testing?
31. What is meant by data loading validation?

1. What is ETL testing?

Answer:

ETL testing (Extract, Transform, Load testing) ensures data is accurately extracted from a source, transformed based on business rules, and loaded into the target system without data loss, corruption, or inconsistency.

2. Why is ETL testing important?

Answer:

It ensures data accuracy, completeness, and quality in data warehouses, helping organizations make reliable data-driven decisions.

3. What are the main types of ETL testing?

Answer:

Main types include data transformation testing, data completeness testing, data quality testing, performance testing, and integration testing.

4. Explain the ETL process.

Answer:

ETL involves **Extracting** data from sources, **Transforming** it to fit target data structures and business logic, and **Loading** it into the target system, like a data warehouse.

5. What are the different stages of ETL testing?

Answer:

Common stages include requirement analysis, test planning, test case design, data validation, executing test cases, defect reporting, and summary reporting.

6. What are the key challenges in ETL testing?

Answer:

Key challenges include handling huge volumes of data, data transformation complexities, managing data quality issues, and testing for system performance.

7. What tools are commonly used for ETL testing?

Answer:

Popular ETL testing tools include Informatica, Talend, Apache Nifi, QuerySurge, and custom SQL-based scripts.

8. How do you perform data completeness testing in ETL?

Answer:

Verify that all expected records are transferred from the source to the target by checking row counts or record counts and comparing them.

9. What is data transformation testing?

Answer:

It validates that data transformations specified in business rules are accurately applied during the ETL process, ensuring accurate field mappings and aggregations.

10. Explain the concept of data reconciliation in ETL testing.

Answer:

Data reconciliation checks that data in the source and target are consistent after ETL processes by verifying transformations and mappings.

11. What is data quality testing in ETL?

Answer:

It involves validating data for accuracy, consistency, completeness, and uniqueness, ensuring the data conforms to expected quality standards.

12. What is initial load and incremental load testing in ETL?

Answer:

Initial load tests the first-time data loading from scratch, while incremental load validates data loading with only the changed or new records.

13. How do you perform null validation in ETL testing?

Answer:

Ensure fields that should not have null values are populated and that fields allowed to be null contain valid nulls, according to requirements.

14. What is source-to-target data mapping?

Answer:

It specifies how data in each source field corresponds to fields in the target, used to ensure accurate data flow during ETL.

15. What is surrogate key validation in ETL?

Answer:

Verify that surrogate keys (unique identifiers) are correctly generated in the target system and are unique, especially in dimensional tables.

16. How would you handle duplicate records in ETL testing?

Answer:

Identify duplicates through unique key checks, eliminate duplicates as per business rules, and ensure only unique records are loaded.

17. What is end-to-end ETL testing?

Answer:

It covers the entire ETL process, validating data extraction from source systems, transformation logic, and correct loading into the target system.

18. Explain data integrity testing in ETL.

Answer:

Ensures data relationships (such as primary-foreign key relationships) are preserved from the source to the target, ensuring data consistency.

19. What are the common ETL testing scenarios?

Answer:

Common scenarios include verifying source-to-target counts, transformation logic, data integrity, null value handling, and performance.

20. What is a test harness in ETL testing?

Answer:

A test harness is a framework for automating ETL tests, where test scripts validate ETL processes, often customized using SQL or scripting.

21. What is a dimension table in ETL?

Answer:

Dimension tables store descriptive information used in analysis, such as customer or product information, often linked to fact tables in a star schema.

22. What is a fact table in ETL?

Answer:

Fact tables store measurable data, such as sales or profit values, and typically contain foreign keys to related dimension tables.

23. Explain the importance of data validation in ETL testing.

Answer:

Data validation checks that extracted data is accurate, transformations are correctly applied, and the data is loaded as intended in the target system.

24. What is metadata testing in ETL?

Answer:

Verifies that metadata, like data type and length specifications, is consistent between source and target to prevent loading errors.

25. How do you validate data transformation logic in ETL testing?

Answer:

Test each transformation rule individually by applying it to sample data and checking that the output matches expected results.

26. What is performance testing in ETL?

Answer:

Evaluates ETL process efficiency, focusing on load time, data processing speed, and system stability under high data volumes.

27. What is exception handling in ETL testing?

Answer:

Identifies how the ETL process handles unexpected data or processing errors, such as data truncation, and ensures they're logged or handled per requirements.

28. How do you handle incremental data load validation in ETL?

Answer:

Validate that only new or updated records are added to the target, based on timestamp or other unique markers.

29. What are some common data validation SQL queries used in ETL testing?

Answer:

Common queries include row count checks, duplicate checks, null validation, and referential integrity checks (JOIN-based checks).

30. What is schema validation in ETL testing?

Answer:

Ensures that the schema in the target matches the source schema in terms of field names, data types, lengths, and constraints.

31. What is meant by data loading validation?

Answer:

Confirms data is correctly loaded into the target without loss, truncation, or duplication and that data types and formats match target specifications.