

FREE · 2026 EDITION

ETL Testing Interview Kit

50 interview questions with answers · 15 SQL queries · skills checklist · 30-day plan

By Asim Noaman Lodhi · etltesting.aims-ai.com

How to use this kit

1. Answer each question aloud before reading the short answer.
2. Type every SQL query yourself and run it on practice data.
3. Tick the checklist honestly; follow the 30-day plan for the gaps.
4. Take a timed practice test before your real interview.

Want 500+ questions with detailed explanations?

ETL Testing Interview Questions & Answers – Practice Tests: Python, automation, SQL and scenario questions, freshers to experienced.

[Practise on Udemty →](#)

50 ETL testing interview questions

ETL and data warehouse basics

1. What is ETL testing?

Verifying that data extracted from sources, transformed by business rules and loaded into a target is complete, accurate and consistent.

2. ETL vs ELT?

ETL transforms before loading; ELT loads raw data first and transforms inside the warehouse.

3. What is a data warehouse?

A central database designed for reporting and analysis, built from many source systems.

4. What is a staging area?

A temporary layer where raw source data lands before transformation.

5. Fact vs dimension table?

Facts store measures (amount, quantity); dimensions store descriptive context (customer, product, date).

6. Star vs snowflake schema?

Star: dimensions joined directly to the fact. Snowflake: dimensions normalised into sub-tables.

7. What is a surrogate key?

A system-generated key in the warehouse, independent of the source business key.

8. What is SCD Type 1?

Overwrite the old value; no history kept.

9. What is SCD Type 2?

Insert a new row for each change, closing the old row with an end date or flag; full history kept.

10. What is a data mart?

A subject-focused subset of the warehouse, for example sales or finance.

Testing process and test design

11. What is a source-to-target mapping document?

The document listing each target column, its source and the transformation rule; the basis of ETL test cases.

12. Main types of ETL testing?

Count, completeness, transformation, data quality, duplicate, referential integrity, incremental load, performance and report testing.

13. How do you test a full load?

Compare counts and totals, check every source key exists in target, validate transformations and data quality.

14. How do you test an incremental load?

Insert, update and delete source rows after the last load, run the job and confirm only those changes were applied.

15. What is data profiling?

Analysing source data for nulls, ranges, distinct values and patterns before writing tests.

16. What is data reconciliation?

Comparing counts, sums and keys between source and target to prove nothing was lost or changed.

17. What is referential integrity testing?

Checking every foreign key in a fact has a matching dimension row.

18. What is metadata testing?

Verifying table and column names, data types, lengths and constraints match the design.

19. How do you test SCD Type 2?

Change a tracked attribute, run the load, confirm the old row is closed and a new current row exists, with one current row per key.

20. What goes in an ETL defect report?

The query used, expected vs actual result, sample rows, environment and severity.

SQL for ETL testers

21. How do you compare counts?

SELECT COUNT(*) from source and target and compare, ideally per load date.

22. How do you find missing records?

Source MINUS (Oracle) or EXCEPT target on the key columns.

23. How do you find duplicates?

GROUP BY the key HAVING COUNT(*) > 1.

24. WHERE vs HAVING?

WHERE filters rows before grouping; HAVING filters groups after.

25. INNER vs LEFT JOIN?

INNER keeps matches only; LEFT keeps all left rows with NULLs where no match.

26. How do you find NULLs in a mandatory column?

SELECT COUNT(*) WHERE column IS NULL; expect 0.

27. UNION vs UNION ALL?

UNION removes duplicates; UNION ALL keeps them and is faster.

28. How do you pick the latest row per key?

ROW_NUMBER() OVER (PARTITION BY key ORDER BY updated_at DESC) and keep row 1.

29. How do you compare large tables quickly?

Compare counts, sums and hash totals per partition first; drill into mismatches only.

30. How do you test a transformation rule in SQL?

Recalculate the rule from source columns and select target rows where it does not match.

Python and automation

31. Why automate ETL tests?

Repeated checks (counts, duplicates, reconciliation) run after every load without manual effort.

32. Which Python libraries are used?

pandas, pytest, SQLAlchemy or pyodbc, and Great Expectations for data quality rules.

33. How do you compare two DataFrames?

Merge on the key with how="outer", indicator=True, then inspect left_only and right_only rows.

34. What is a pytest fixture?

Reusable setup such as a database connection, shared across tests through conftest.py.

35. How do you run a check on many tables?

Use @pytest.mark.parametrize with a list of tables.

36. How do you integrate tests with CI/CD?

Trigger pytest after the ETL job in Jenkins, GitHub Actions or Azure DevOps and publish the report.

37. What should not be automated?

One-off checks, unstable requirements and judgement-based reviews.

38. How can AI help ETL testers?

Draft test cases and SQL from mapping documents; the tester must review and verify them.

39. Selenium for ETL testing?

Only for report screens; data validation needs SQL, Python or ETL testing tools.

40. Name ETL automation tools.

QuerySurge, Datagaps ETL Validator, Informatica DVO, Great Expectations, dbt tests.

Scenarios and HR

41. Counts match but totals differ. Why?

Transformation, rounding, currency, time zone or duplicate-join issues; compare at row level.

42. Target has fewer rows than source. What do you check?

Filters in the job, rejected rows, failed batches and key mismatches.

43. A rerun doubled the data. What is wrong?

The load is not idempotent; reruns must not create duplicates.

44. Report total does not match source. How do you investigate?

Compare totals layer by layer: source, staging, warehouse, report.

45. Release is tomorrow and you found a critical defect.

Inform the lead with evidence and impact immediately; let decision-makers choose.

46. Developer says it is not a bug.

Show the requirement, query and sample rows; escalate to the BA if still unclear.

47. Explain your ETL project.

Domain, sources, target, your role, 3 checks you ran, 1 defect you found and its impact.

48. Tell me about yourself.

Name and role, experience, core skills, one achievement with a number, why this role. 60–90 seconds.

49. Why ETL testing?

You enjoy validating data, it uses SQL and analysis, and data quality matters more every year.

50. Where do you see yourself in 3 years?

Leading data testing with automation and cloud data platforms.

Ready for a timed practice test?

Practise 500+ questions with detailed explanations before your real interview.

[Start the practice tests on Udemy →](#)

15 SQL queries every ETL tester must know

1. Row count

```
SELECT COUNT(*) FROM stg_orders;
SELECT COUNT(*) FROM fact_orders;
```

2. Missing in target

```
SELECT order_id FROM stg_orders
EXCEPT
SELECT order_id FROM fact_orders;
```

3. Extra in target

```
SELECT order_id FROM fact_orders
EXCEPT
SELECT order_id FROM stg_orders;
```

4. Duplicates

```
SELECT order_id, COUNT(*) FROM fact_orders
GROUP BY order_id HAVING COUNT(*) > 1;
```

5. NULL check

```
SELECT COUNT(*) FROM fact_orders WHERE
customer_key IS NULL;
```

6. Sum reconciliation

```
SELECT SUM(amount) FROM stg_orders;
SELECT SUM(amount) FROM fact_orders;
```

7. Transformation rule

```
SELECT * FROM fact_orders
WHERE ROUND(net_amount,2) <>
ROUND(gross_amount - discount,2);
```

8. Orphan facts

```
SELECT f.order_id FROM fact_orders f
LEFT JOIN dim_customer d ON d.customer_key =
f.customer_key
WHERE d.customer_key IS NULL;
```

9. Latest row per key

```
SELECT * FROM (SELECT t.*, ROW_NUMBER() OVER
(PARTITION BY id ORDER BY updated_at DESC)
rn FROM t) x WHERE rn = 1;
```

10. One current row per key (SCD2)

```
SELECT customer_id FROM dim_customer
WHERE current_flag = 'Y'
GROUP BY customer_id HAVING COUNT(*) > 1;
```

11. Counts by load date

```
SELECT load_date, COUNT(*) FROM fact_orders
GROUP BY load_date ORDER BY load_date;
```

12. Value range

```
SELECT * FROM fact_orders WHERE amount < 0
OR amount > 1000000;
```

13. Invalid codes

```
SELECT * FROM dim_customer
WHERE gender NOT IN ('M','F','U');
```

14. Blank vs NULL

```
SELECT COUNT(*) FROM dim_customer
WHERE TRIM(email) = '';
```

15. Schema check

```
SELECT column_name, data_type FROM
information_schema.columns
WHERE table_name = 'fact_orders';
```

Learn ETL testing hands-on

ETL Testing: From Beginners to Advanced – 79 lectures on ETL and data warehouse testing, data quality, Python, Power Query and AI agents. Rated 4.5 by 251 students.

[See the course on Udemy →](#)

Skills self-check

Tick what you can do today without help. 8–10: start applying. 5–7: 2–3 weeks of practice. Under 5: follow the plan below.

- ☐ Write a count and EXCEPT/MINUS query between two tables
- ☐ Find duplicates and NULLs in a key column
- ☐ Explain fact vs dimension with an example
- ☐ Explain SCD Type 1 vs Type 2
- ☐ Turn one mapping rule into test cases
- ☐ Describe the ETL testing life cycle
- ☐ Log a data defect with query, expected and actual result
- ☐ Explain full vs incremental load testing
- ☐ Compare two files in Excel or Python
- ☐ Explain one ETL project end to end

30-day plan

When	Focus
Days 1–7	SQL: SELECT, joins, GROUP BY, NULLs, duplicates
Days 8–12	Data warehouse concepts, star schema, SCD
Days 13–18	Mapping documents and test case design
Days 19–23	Source-to-target validation and defect writing
Days 24–27	Python or Excel automation of repeated checks
Days 28–30	Mock interviews, self introduction, resume

Which course fits you?

You are...	Course
New to testing	Manual Software Testing Mastery with AI
Manual tester moving to data	ETL Testing: From Beginners to Advanced
Interview coming up	ETL Testing Interview Questions & Answers – Practice Tests
Moving to data analysis / BI	Claude AI for Data Analysis & Business Intelligence

More free guides: <https://etltesting.aims-ai.com/blog/start-here.html>