



InterviewBit

ETL Testing Interview Questions



To view the live version of the page, [click here](#).

© Copyright by Interviewbit

Contents

ETL Interview Questions for Freshers

1. What is the importance of ETL testing?
2. Explain the process of ETL testing.
3. Name some tools that are used in ETL.
4. What are different types of ETL testing?
5. What are the roles and responsibilities of an ETL tester?
6. What are the different challenges of ETL testing?
7. Explain the three-layer architecture of an ETL cycle.
8. Explain data mart.
9. Explain how a data warehouse differs from data mining.
10. What do you mean by data purging?
11. State difference between ETL and OLAP (Online Analytical Processing) tools.
12. Write about the difference between power mart and power center.
13. What is data source view?
14. Write the difference between ETL testing and database testing.
15. What is BI (Business Intelligence)?
16. What do you mean by ETL Pipeline?
17. Explain the data cleaning process.

ETL Interview Questions for Experienced

18. State difference between ETL testing and manual testing.

ETL Interview Questions for Experienced (.....Continued)

19. Mention some of the ETL bugs.
20. Can you define cubes and OLAP cubes?
21. Explain what is fact and write its type.
22. Define Grain of Fact.
23. What do you mean by ODS (Operational data store)?
24. What do you mean by staging area and write its main purpose?
25. Explain the Snowflake schema.
26. Explain what you mean by Bus Schema.
27. What do you mean by schema objects?
28. What is the benefit of using a Data reader destination adapter?
29. What do you mean by factless table?
30. Explain SCD (Slowly Change Dimension).

ETL Scenario Based Interview Questions

31. Explain partitioning in ETL and write its type.
32. Write different ways of updating a table when SSIS (SQL Server Integration Services) is being used.
33. Write some ETL test cases.
34. Explain ETL mapping sheets.
35. How ETL testing is used in third party data management?
36. Explain how ETL is used in data migration projects.

ETL Scenario Based Interview Questions (.....Continued)

37. What are the conditions under which you use dynamic cache and static cache in connected and unconnected transformations?



Let's get Started

What is ETL Testing?

Almost every business relies heavily on data nowadays, which is good! With subjective and accurate data, we can grasp more than we can comprehend with our human brains. What matters is when. Data processing, like any system, is prone to errors. What is the value of data when there is a possibility that some of it could be lost, incomplete, or irrelevant?



This is where ETL testing comes into play. In business processes today, ETL is considered an important component of data warehousing architecture. Data is extracted from source systems, transformed into a consistent data type, and loaded into a single repository through ETL (Extract, Transform, and Load). Validating, evaluating, and qualifying data is an important part of ETL testing. We conduct ETL testing after extracting, transforming, and loading the data to verify that the final data was appropriately loaded into the system in the correct format. It ensures that data reaches its destination safely and is of high quality before it enters your BI (Business Intelligence) reports.

ETL Interview Questions for Freshers

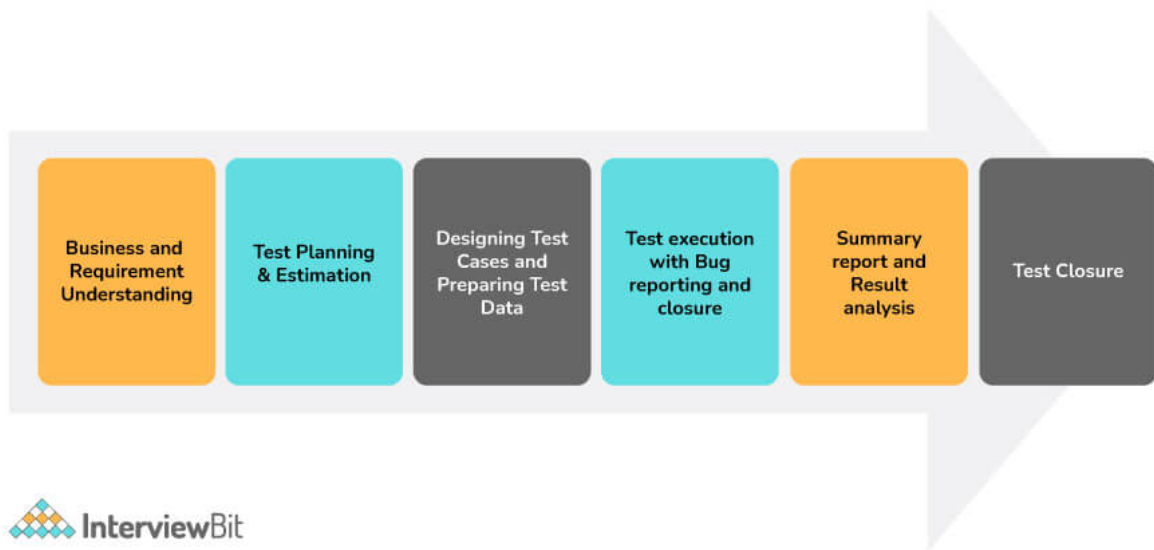
1. What is the importance of ETL testing?

Following are some of the notable benefits that are highlighted while endorsing ETL Testing:

- Ensure data is transformed efficiently and quickly from one system to another.
- Data quality issues during ETL processes, such as duplicate data or data loss, can also be identified and prevented by ETL testing.
- Assures that the ETL process itself is running smoothly and is not hampered.
- Ensures that all data implemented is in line with client requirements and provides accurate output.
- Ensures that bulk data is moved to the new destination completely and securely.

2. Explain the process of ETL testing.

ETL testing is made easier when a testing strategy is well defined. The ETL testing process goes through different phases, as illustrated below:



- **Analyze Business Requirements:** To perform ETL Testing effectively, it is crucial to understand and capture the business requirements through the use of data models, business flow diagrams, reports, etc.
- **Identifying and Validating Data Source:** To proceed, it is necessary to identify the source data and perform preliminary checks such as schema checks, table counts, and table validations. The purpose of this is to make sure the ETL process matches the business model specification.
- **Design Test Cases and Preparing Test Data:** Step three includes designing ETL mapping scenarios, developing SQL scripts, and defining transformation rules. Lastly, verifying the documents against business needs to make sure they cater to those needs. As soon as all the test cases have been checked and approved, the pre-execution check is performed. All three steps of our ETL processes - namely extracting, transforming, and loading - are covered by test cases.
- **Test Execution with Bug Reporting and Closure:** This process continues until the exit criteria (business requirements) have been met. In the previous step, if any defects were found, they were sent to the developer for fixing, after which retesting was performed. Moreover, regression testing is performed in order to prevent the introduction of new bugs during the fix of an earlier bug.
- **Summary Report and Result Analysis:** At this step, a test report is prepared, which lists the test cases and their status (passed or failed). As a result of this report, stakeholders or decision-makers will be able to properly maintain the delivery threshold by understanding the bug and the result of the testing process.
- **Test Closure:** Once everything is completed, the reports are closed.

3. Name some tools that are used in ETL.

The use of ETL testing tools increases IT productivity and facilitates the process of extracting insights from big data. With the tool, you no longer have to use labor-intensive, costly traditional programming methods to extract and process data.

Technology evolved over time, so did solutions. Nowadays, various ways can be used for ETL testing depending on the source data and the environment. There are several ETL vendors that focus on ETL exclusively, such as Informatica. Software vendors like IBM, Oracle, and Microsoft provide other tools as well. Open source ETL tools have also recently emerged that are free to use. The following are some ETL software tools to consider:

Enterprise Software ETL

- Informatica PowerCenter
- IBM InfoSphere DataStage
- Oracle Data Integrator (ODI)
- Microsoft SQL Server Integration Services (SSIS)
- SAP Data Services
- SAS Data Manager, etc.

Open Source ETL

- Talend Open Studio
- Pentaho Data Integration (PDI)
- Hadoop, etc.

4. What are different types of ETL testing?

Before you begin the testing process, you need to define the right ETL Testing technique. It is important to ensure that the ETL test is performed using the right technique and that all stakeholders agree to it. Testing team members should be familiar with this technique and the steps involved in testing. Below are some types of testing techniques that can be used:

- **Production Validation Testing:** Also known as "production reconciliation" or "table balancing," it involves validating data in production systems and comparing it against the source data.
- **Source to Target Count Testing:** This ensures that the number of records loaded into the target is consistent with what is expected.
- **Source to Target Data Testing:** This entails ensuring no data is lost and truncated when loading data into the warehouse, and that the data values are accurate after transformation.
- **Metadata Testing:** The process of determining whether the source and target systems have the same schema, data types, lengths, indexes, constraints, etc.
- **Performance Testing:** Verifying that data loads into the data warehouse within predetermined timelines to ensure speed and scalability.
- **Data Transformation Testing:** This ensures that data transformations are completed according to various business rules and requirements.
- **Data Quality Testing:** This testing involves checking numbers, dates, nulls, precision, etc. Testing includes both Syntax Tests to report invalid characters, incorrect upper/lower case order, etc., and Reference Tests to check if the data is properly formatted.
- **Data Integration Testing:** In this test, testers ensure the data from various sources have been properly incorporated into the target system, as well as verifying the threshold values.
- **Report Testing:** The test examines the data in a summary report, verifying the layout and functionality, and making calculations for subsequent analysis.

5. What are the roles and responsibilities of an ETL tester?

Since ETL testing is so important, ETL testers are in great demand. ETL testers validate data sources, extract data, apply transformation logic, and load data into target tables. The following are key responsibilities of an ETL tester:

- In-depth knowledge of ETL tools and processes.
- Performs thorough testing of the ETL software.
- Check the data warehouse test component.
- Perform the backend data-driven test.
- Design and execute test cases, test plans, test harnesses, etc.
- Identifies problems and suggests the best solutions.
- Review and approve the requirements and design specifications.
- Writing SQL queries for testing scenarios.
- Various types of tests should be carried out, including primary keys, defaults, and checks of other ETL-related functionality.
- Conducts regular quality checks.

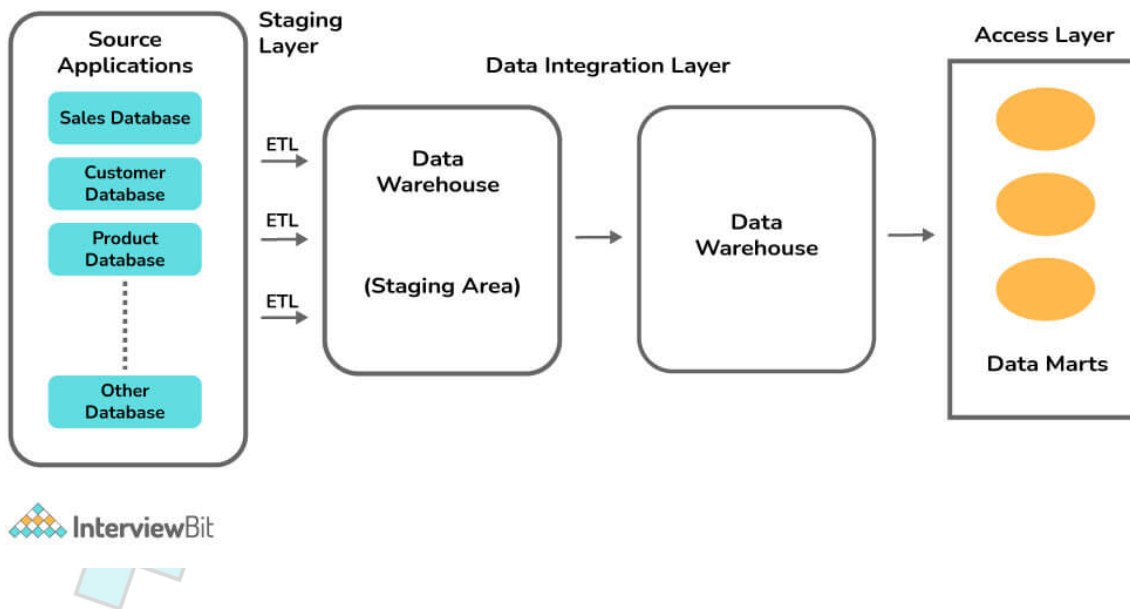
6. What are the different challenges of ETL testing?

In spite of the importance of ETL testing, companies may face some challenges when trying to implement it in their applications. The volume of data involved or the heterogeneous nature of the data makes ETL testing challenging. Some of these challenges are listed below:

- Changing customer requirements result in re-running test cases.
- Changing customer requirements may necessitate a tester creating/modifying new mapping documents and SQL scripts, resulting in a long and tedious process.
- Uncertainty about business requirements or employees who are not aware of them.
- During migration, data loss may occur, making it difficult for source-to-destination reconciliation to take place.
- An incomplete or corrupt data source.
- Reconciliation between data sources and targets may be impacted by incorporating real-time data.
- There may be memory issues in the system due to the large volume of historical data.
- Testing with inappropriate tools or in an unstable environment.

7. Explain the three-layer architecture of an ETL cycle.

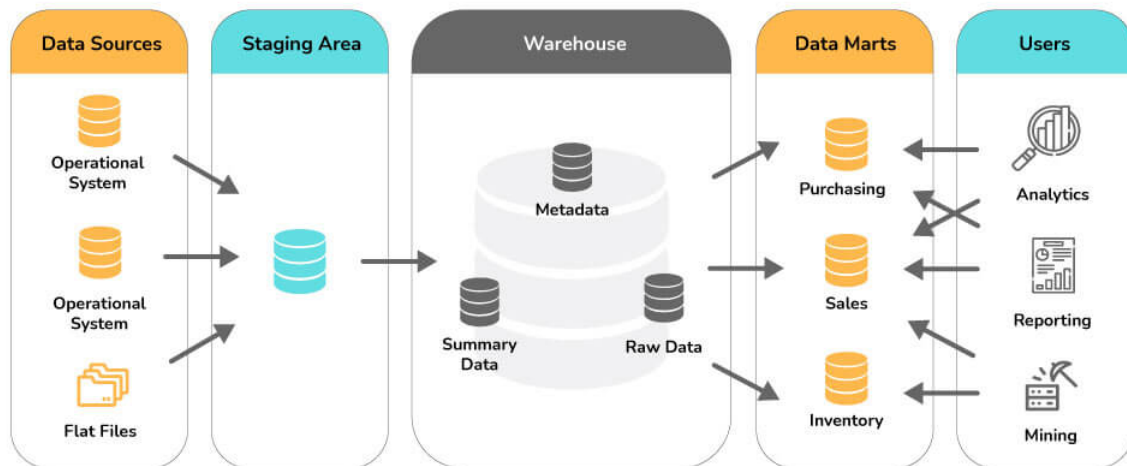
Typically, ETL tool-based data warehouses use staging areas, data integration layers, and access layers to accomplish their work. In general, the architecture has three layers as shown below:



- **Staging Layer:** In a staging layer, or source layer, data is stored that is extracted from multiple data sources.
- **Data Integration Layer:** The integration layer plays the role of transforming data from the staging layer to the database layer.
- **Access Layer:** Also called a dimension layer, it allows users to retrieve data for analytical reporting and information retrieval.

8. Explain data mart.

An enterprise data warehouse can be divided into subsets, also called data marts, which are focused on a particular business unit or department. Data marts allow selected groups of users to easily access specific data without having to search through an entire data warehouse. Some companies, for example, may have a data mart aligned with purchasing, sales, or inventories as shown below:

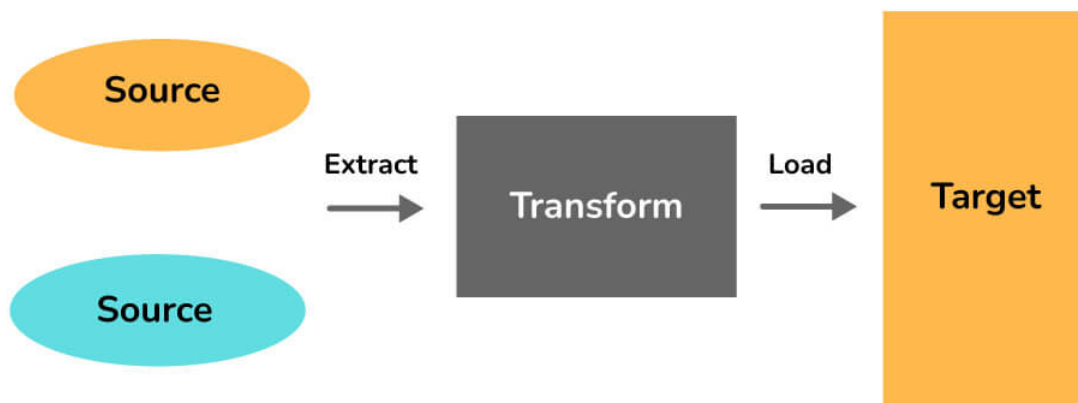


In contrast to data warehouses, each data mart has a unique set of end users, and building a data mart takes less time and costs less, so it is more suitable for small businesses. There is no duplicate (or unused) data in a data mart, and the data is updated on a regular basis.

9. Explain how a data warehouse differs from data mining.

Both data mining and data warehousing are powerful data analysis and storage techniques.

- **Data warehousing:** To generate meaningful business insights, it involves compiling and organizing data from various sources into a common database. In a data warehouse, data are cleaned, integrated and consolidated to support management decision-making processes. Object-oriented, integrated, time-varying, and nonvolatile data can be stored within a Data warehouse.



DATA WAREHOUSING PROCESS



- **Data mining:** Also referred to as KDD (Knowledge Discover in Database), it involves searching for and identifying hidden, relevant, and potentially valuable patterns in large data sets. An important goal of data mining is to discover previously unknown relationships among the data. Through data mining, insights can be extracted that can be used for things such as marketing, fraud detection, and scientific discoveries.



Difference between Data Warehouse and Data Mining -

Data Warehousing	Data Mining
It involves gathering all relevant data for analytics in one place.	Data is extracted from large datasets using this method.
Data extraction and storage assist in facilitating easier reporting.	It identifies patterns by using pattern recognition techniques.
Engineers are solely responsible for data warehousing, and data is periodically stored.	Data mining is carried out by business users in conjunction with engineers, and data is analyzed regularly.
In addition to making data mining easier and more convenient, it helps sort and upload important data to databases.	Analyzing information and data is made easier.
It is possible to accumulate a large amount of irrelevant and unnecessary data. Loss and erasure of data can also be problematic.	Not doing it correctly can create data breaches and hacking since data mining isn't always 100% accurate.
Data mining cannot take place without this process, since it compiles and organizes data into a common database.	Because the process requires compiled data, it always takes place after data warehousing.
Data warehouses simplify every type of business data.	Comparatively, data mining techniques are inexpensive.

10. What do you mean by data purging?

When data needs to be deleted from the data warehouse, it can be a very tedious task to delete data in bulk. The term data purging refers to methods of permanently erasing and removing data from a data warehouse. Data purging, often contrasted with deletion, involves many different techniques and strategies. When you delete data, you are removing it on a temporary basis, but when you purge data, you are permanently removing the data and freeing up memory or storage space. In general, the data that is deleted is usually junk data such as null values or extra spaces in the row. Using this approach, users can delete multiple files at once and maintain both efficiency and speed.

11. State difference between ETL and OLAP (Online Analytical Processing) tools.

- **ETL tools:** The data is extracted, transformed, and loaded into the data warehouse or data mart using ETL tools. Several transformations are necessary before data is loaded into the target table in order to implement business logic. *Example:* Data stage, Informatica, etc.
- **OLAP (Online Analytical Processing) tools:** OLAP tools are designed to create reports from data warehouses and data marts for business analysis. It loads data from the target tables into the OLAP repository and performs the required modifications to create a report. *Example:* Business Objects, Cognos etc.

12. Write about the difference between power mart and power center.

Power Mart	Power Center
It only processes small amounts of data and is considered good if the processing requirements are low.	It is considered good when the amount of data to be processed is high, as it processes bulk data in a short period of time.
ERP sources are not supported.	ERP sources such as SAP, PeopleSoft, etc. are supported.
Currently, it only supports local repositories.	Local and global repositories are supported.
There are no specifications for turning a local repository into a global repository.	It is capable of converting local repositories into global ones.
Session partitions are not supported.	To improve the performance of ETL transactions, it supports session partitioning.

13. What is data source view?

Several analysis services databases rely on relational schemas, and the Data source view is responsible for defining such schemas (logical model of the schema). Additionally, it can be easily used to create cubes and dimensions, thus enabling users to set their dimensions in an intuitive way. A multidimensional model is incomplete without a DSV. In this way, you are given complete control over the data structures in your project and are able to work independently from the underlying data sources (e.g., changing column names or concatenating columns without directly changing the original data source). Every model must have a DSV, no matter when or how it's created.

Using the Data Source View Wizard to create a DSV

You must run the Data Source View Wizard from Solution Explorer within SQL Server Data Tools to create the DSV.

- In solution explorer, Right Click Data source view folder -> Click New Data Source View.
- Choose one of the available data source objects, or add a new one.
- Click Advanced on the same page to specifically select schemas, apply a filter, or exclude information about table relationships.
- Filter Available Objects (If we use a string as a selection criterion, it is possible to prune the list of the available objects).
- A Name Matching page appears if there are no table relationships defined for the relational data source, and you can choose the appropriate method for matching names by clicking on it.

14. Write the difference between ETL testing and database testing.

Data validation is involved in both ETL testing and database testing, however, the two are different. The ETL testing procedure normally involves analyzing data stored in a warehouse system. On the other hand, the database testing procedure is commonly used to analyze data stored in transactional systems. The following are the distinct differences between ETL testing and Database testing.

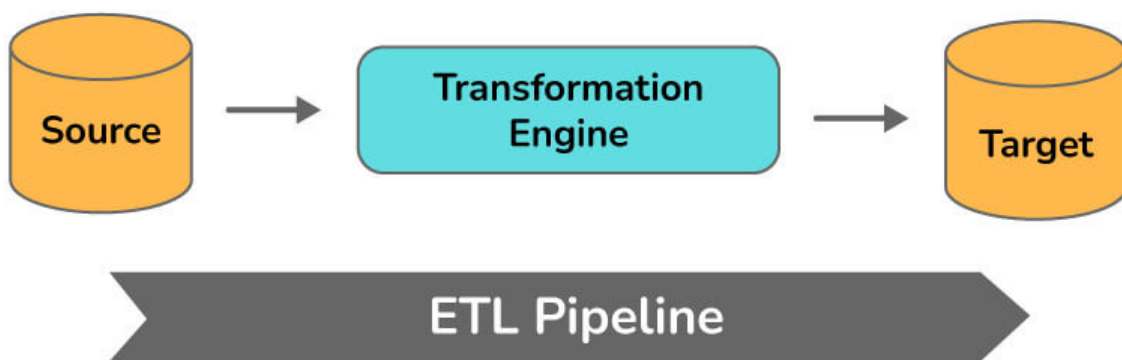
ETL Testing	Database Testing
The ETL process is used to test data extraction, transformation, and loading for BI reporting purposes.	Data is validated and integrated by performing database testing.
Data movement is being checked to determine if it is going as expected	This test is primarily designed to verify that data follows the rules or standards defined in the Data Model.
It verifies whether the counts and data in the source and target match or not.	It ensures that foreign key relationships are maintained and no orphan records are present, as well as that a column in the table has valid values.
This technique is applied to OLAP systems.	This technique is applied to OLTP systems.
The approach utilizes denormalized data with fewer joins, more indexes, and more aggregates.	The approach utilizes normalized data with joins.
Some of the most common ETL tools are QuerySurge, Informatica, Cognos, etc.	Some of the most common database testing tools are Selenium, QTP, etc.

15. What is BI (Business Intelligence)?

Business Intelligence (BI) involves acquiring, cleaning, analyzing, integrating, and sharing data as a means of identifying actionable insights and enhancing business growth. An effective BI test verifies staging data, ETL process, BI reports, and ensures the implementation is reliable. In simple words, BI is a technique used to gather raw business data and transform it into useful insight for a business. By performing BI Testing, insights from the BI process are verified for accuracy and credibility.

16. What do you mean by ETL Pipeline?

As the name suggests, ETL pipelines are the mechanisms to perform ETL processes. This involves a series of processes or activities required for transferring data from one or more sources into the data warehouse for analysis, reporting and data synchronization. It is important to move, consolidate, and alter source data from multiple systems to match the parameters and capabilities of the destination database in order to provide valuable insights.



Among its benefits are:

- They reduce errors, bottlenecks, and latency, ensuring the smooth flow of information between systems.
- With ETL pipelines, businesses are able to achieve competitive advantage.
- The ETL pipeline can centralize and standardize data, allowing analysts and decision-makers to easily access and use it.
- It facilitates data migrations from legacy systems to new repositories.

17. Explain the data cleaning process.

There is always the possibility of duplicate or mislabeled data when combining multiple data sources. Incorrect data leads to unreliable outcomes and algorithms, even when they appear to be correct. Therefore, consolidation of multiple data representations as well as elimination of duplicate data become essential in order to ensure accurate and consistent data. Here comes the importance of the data cleaning process.

Data cleaning can also be referred to as data scrubbing or data cleansing. This refers to the process of removing incomplete, duplicate, corrupt, or incorrect data from a dataset. As the need to integrate multiple data sources becomes more apparent, for example in data warehouses or federated database systems, the significance of data cleaning increases greatly. Because the specific steps in a data cleaning process will vary depending on the dataset, developing a template for your process will ensure that you do it correctly and consistently.

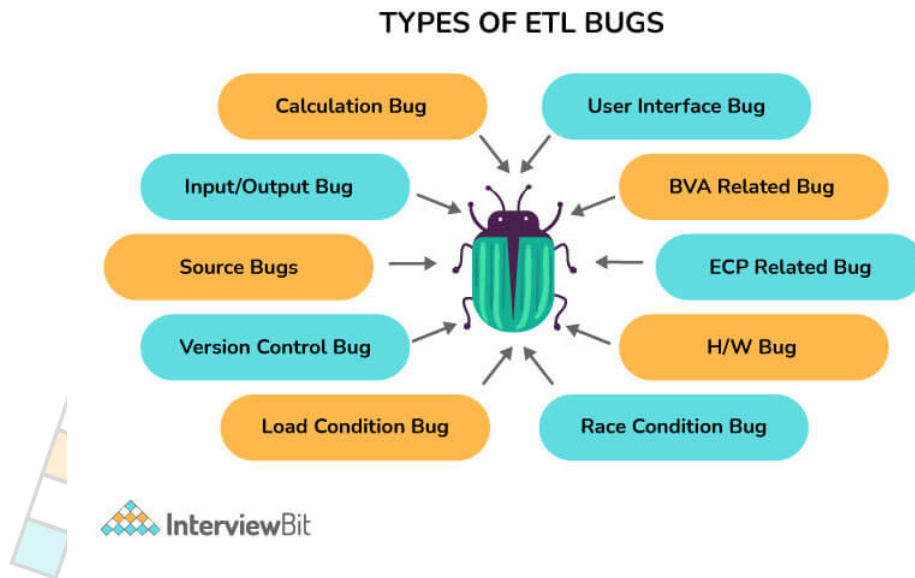
ETL Interview Questions for Experienced

18. State difference between ETL testing and manual testing.

ETL Testing	Manual Testing
The test is an automated process, which means that no special technical knowledge is needed aside from understanding the software.	It requires technical expertise in SQL and Shell scripting since it is a manual process.
It is extremely fast and systematic, and it delivers excellent results.	In addition to being time-consuming, it is highly prone to errors.
Databases and their counts are central to ETL testing.	Manual testing focuses on the program's functionality.
Metadata is included and can easily be altered.	It lacks metadata, and changes require more effort.
It is concerned with error handling, log summary, and load progress, which eases the developer's and maintainer's workload.	From a maintenance perspective, it requires maximum effort.
It is very good at handling historical data.	As data increases, processing time decreases.

19. Mention some of the ETL bugs.

Following are a few common ETL bugs:

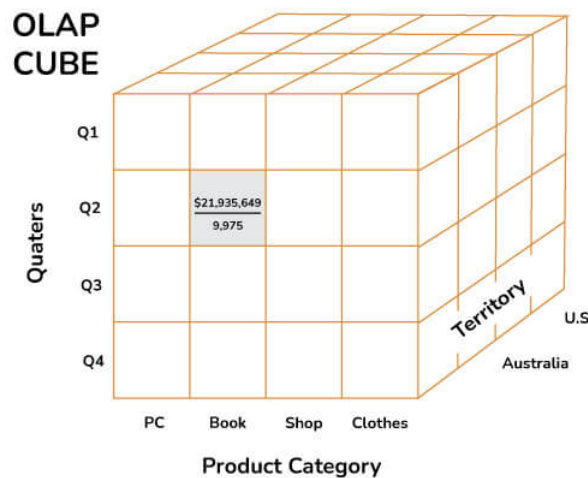


- **User Interface Bug:** GUI bugs include issues with color selection, font style, navigation, spelling check, etc.
- **Input/Output Bug:** This type of bug causes the application to take invalid values in place of valid ones.
- **Boundary Value Analysis Bug:** Bugs in this section check for both the minimum and maximum values.
- **Calculation bugs:** These bugs are usually mathematical errors causing incorrect results.
- **Load Condition Bugs:** A bug like this does not allow multiple users. The user-accepted data is not allowed.
- **Race Condition Bugs:** This type of bug interferes with your system's ability to function properly and causes it to crash or hang.
- **ECP (Equivalence Class Partitioning) Bug:** A bug of this type results in invalid types.
- **Version Control Bugs:** Regression Testing is where these kinds of bugs normally occur and does not provide version details.
- **Hardware Bugs:** This type of bug prevents the device from responding to an application as expected.
- **Help Source Bugs:** The help documentation will be incorrect due to this bug.

20. Can you define cubes and OLAP cubes?

The **cube** is one of the things on which data processing relies heavily. In their simplest form, cubes are just data processing units that contain dimensions and fact tables from the Data warehouse. It provides clients with a multidimensional view of data, querying, and analysis capabilities.

On the other hand, **Online Analytical Processing** (OLAP) is software that allows you to analyze data from several databases at the same time. For reporting purposes, an OLAP cube can be used to store data in the multidimensional form. With the cubes, creating and viewing reports becomes easier, as well as smoothing and improving the reporting process. The end users are responsible for managing and maintaining these cubes, who have to manually update their data.



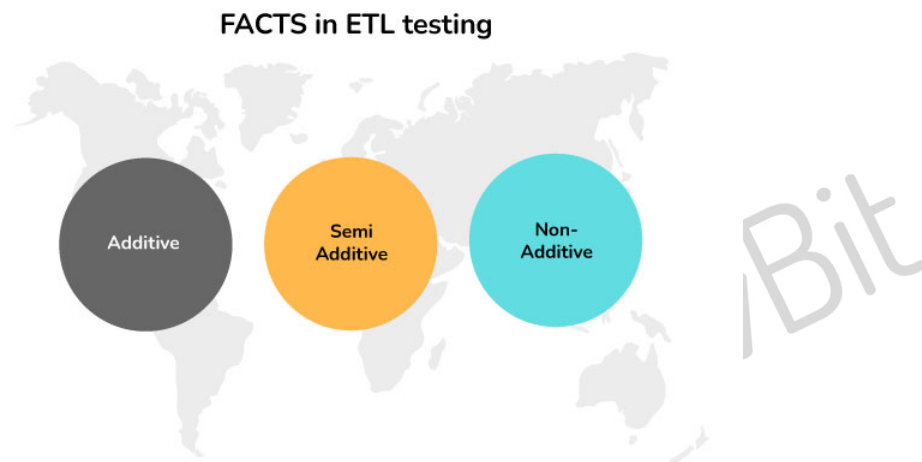
21. Explain what is fact and write its type.

An important aspect of data warehousing is the fact table. A fact table basically represents the measurements, metrics, or facts of a business process. In fact tables, facts are stored, and they are linked to a number of dimension tables via foreign keys. Facts are usually details and/or aggregated measurements of a business process which can be calculated and grouped to address the business question. Data schemas like the star schema or snowflake schema consist of a central fact table surrounded by several dimension tables. The measures or numbers like sales, cost, profit and loss, etc., are some examples of facts.

Fact tables have two types of columns, foreign keys and measures columns. Foreign keys store foreign keys to dimensions, while measures contain numeric facts. Other attributes can be added, depending on the business need and necessity.

Types of Facts

Facts can be divided into three basic types, as follows:



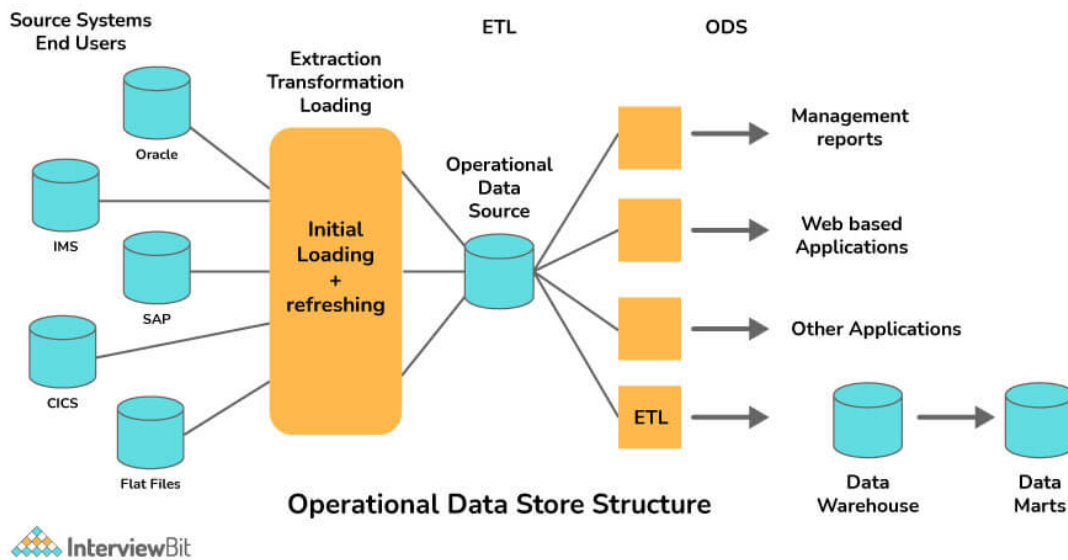
- **Additive:** Facts that are fully additive are the most flexible and useful. We can sum up additive facts across any dimension associated with the fact table.
- **Semi-additive:** We can sum up semi-additive facts across some dimensions associated with the fact table, but not all.
- **Non-Additive:** The Fact table contains non-additive facts, which cannot be summed up for any dimension. The ratio is an example of a non-additive fact.

22. Define Grain of Fact.

Accordingly, grain fact refers to the level of storing fact information. Alternatively, it is known as Fact Granularity.

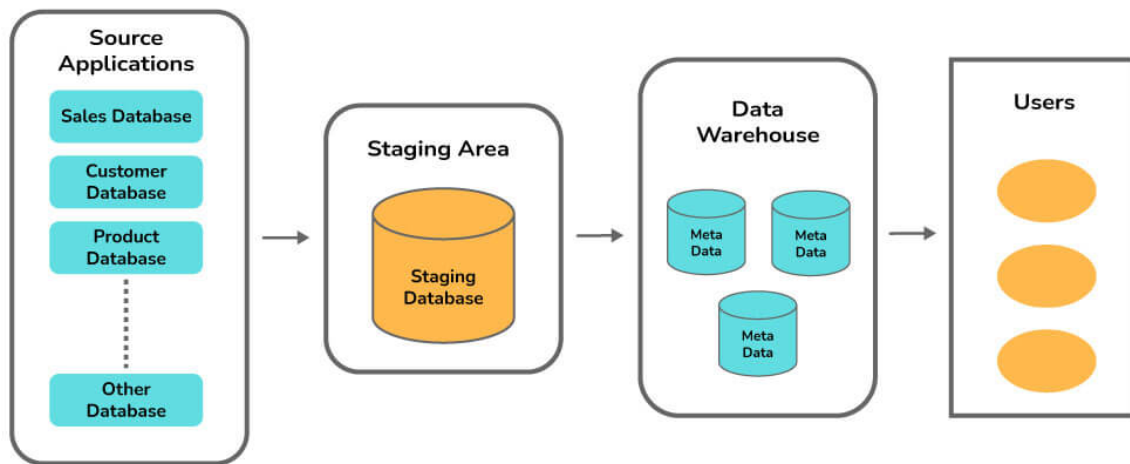
23. What do you mean by ODS (Operational data store)?

Between the staging area and the Data Warehouse, ODS serves as a repository for data. Upon inserting the data into ODS, ODS will load all the data into the EDW (Enterprise data warehouse). The benefits of ODS mainly pertain to business operations, as it presents current, clean data from multiple sources in one place. Unlike other databases, an ODS database is read-only, and customers cannot update it.



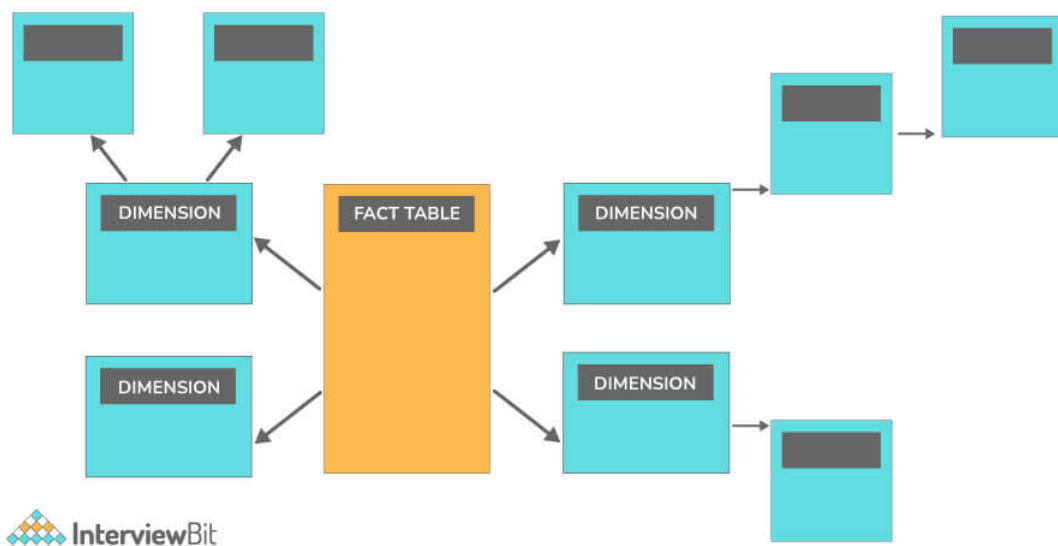
24. What do you mean by staging area and write its main purpose?

During the extract, transform, and load (ETL) process, a staging area or landing zone is used as an intermediate storage area. It serves as a temporary storage area between data sources and data warehouses. Staging areas are primarily used to extract data quickly from their respective data sources, therefore minimizing the impact of those sources. Using the staging area, data is combined from multiple data sources, transformed, validated, and cleaned after data has been loaded.



25. Explain the Snowflake schema.

Adding additional dimension tables to a Star Schema makes it a Snowflake Schema. In the Snowflake schema model, multiple hierarchies of dimension tables surround a central fact table. Alternatively, a dimension table is called a snowflake if its low-cardinality attribute has been segmented into separate normalized tables. These normalized tables are then joined with referential constraints (foreign key constraints) to the original dimensions table. Snowflake schema complexity increases linearly with the level of hierarchy in the dimension tables.



Advantages

- Data integrity is reduced because of structured data.
- Data are highly structured, so it requires little disk space.
- Updating or maintaining Snowflaking tables is easy.

Disadvantages

- Snowflake reduces the space consumed by dimension tables, but the space saved is usually insignificant compared with the entire data warehouse.
- Due to the number of tables added, you may need complex joins to perform a query, which will reduce query performance.

26. Explain what you mean by Bus Schema.

An important part of ETL is dimension identification, and this is largely done by the Bus Schema. A BUS schema is actually comprised of a suite of verified dimensions and uniform definitions and can be used for handling dimension identification across all businesses. To put it another way, the bus schema identifies the common dimensions and facts across all the data marts of an organization just like identifying conforming dimensions (dimensions with the same information/meaning when being referred to different fact tables). Using the Bus schema, information is given in a standard format with precise dimensions in ETL.

27. What do you mean by schema objects?

Generally, a schema comprises a set of database objects, such as tables, views, indexes, clusters, database links, and synonyms, etc. This is a logical description or structure of the database. Schema objects can be arranged in various ways in schema models designed for data warehousing. Star and snowflake schemas are two examples of data warehouse schema models.

28. What is the benefit of using a Data reader destination adapter?

ADO Recordset holds a collection of records (records and columns) from a database table. The Data Reader Destination Adapter is very useful when it comes to populating them in a simple manner. Using the ADO.NET DataReader interface, it exposes the data in a data flow for other applications to consume it.

29. What do you mean by factless table?

Factless tables do not contain any facts or measures. It contains only dimensional keys and deals with event occurrences at the informational level but not at the calculational level. As the name implies, factless fact tables capture relationships between dimensions but lack any numerical or textual data. Factual fact tables can be categorized into two categories: one that describes events, and the other one that describes conditions. Both may have a significant impact on your dimensional modeling.

30. Explain SCD (Slowly Change Dimension).

SCD (Slowly Changing Dimensions) basically keep and manage both current and historical data in a data warehouse over time. Rather than changing regularly on a time-based schedule, SCD changes slowly over time. SCD is considered one of the most critical aspects of ETL.

ETL Scenario Based Interview Questions

31. Explain partitioning in ETL and write its type.

Essentially, partitioning is the process of dividing up a data storage area for improved performance. It can be used to organize your work. Having all your data in one place without organization makes it more difficult for digital tools to find and analyze the data. It is easier and faster to locate and analyze data when your data warehouse is partitioned. The following reasons make partitioning important:

- Facilitate easy data management and enhance performance.
- Ensures that all of the system's requirements are balanced.
- Backups/recoveries made easier.
- Simplifies management and optimizes hardware performance.

Types of Partitioning -

- **Round-robin Partitioning:** This is a method in which data is evenly spread among all partitions. Therefore, each partition has approximately the same number of rows. Unlike hash partitioning, the partitioning columns do not need to be specified. New rows are assigned to partitions in round-robin style.
- **Hash Partitioning:** With hash partitioning, rows are evenly distributed across partitions based on a partition key. Using a hash function, the server creates partition keys to group data.

32. Write different ways of updating a table when SSIS (SQL Server Integration Services) is being used.

In order to update a table in SSIS, the following steps can be taken:

- Use the SQL command.
- For storing stage data, use staging tables.
- Keep data in a cache that occupies a limited amount of space and needs to be refreshed frequently.
- Scripts can be used for scheduling tasks.
- When updating MSSQL, use the full database name.

33. Write some ETL test cases.

Among the most common ETL test cases are:

- **Mapping Doc Validation:** Determines whether the Mapping Doc contains ETL information.
- **Data Quality:** In this case, every aspect of the data is tested, including number Check, Null Check, Precision Check, etc.
- **Correctness Issues:** Tests for missing, incorrect, non-unique, and null data.
- **Constraint Validation:** Make sure that the constraints are properly defined for each table.

34. Explain ETL mapping sheets.

Typically, ETL mapping sheets include full information about a source and a destination table, including every column as well as their lookup in reference tables. As part of the ETL testing process, ETL testers may need to write big queries with multiple joins to validate data at any point in the testing process. Data verification queries are significantly easier to write using ETL mapping sheets.

35. How ETL testing is used in third party data management?

Different kinds of vendors develop different kinds of applications for big companies. Consequently, no single vendor manages everything. Consider a Telecommunication project in which billing is handled by one company and CRM by another. For instance, if a CRM requires data from the company that is managing the billing, now that company will be able to receive the data feed from another company. In this case, we will use the ETL process to load data from the feed.

36. Explain how ETL is used in data migration projects.

Data migration projects commonly use ETL tools. As an example, if the organization managed the data in Oracle 10g earlier and now they want to move to SQL Server cloud database, the data will need to be migrated from Source to Target. ETL tools can be very helpful for carrying out this type of migration. The user will have to spend a lot of time writing ETL code. The ETL tools are therefore very useful since they make coding simpler than P-SQL or T-SQL. Hence, ETL is a very useful process for data migration projects.

37. What are the conditions under which you use dynamic cache and static cache in connected and unconnected transformations?

- In order to update the master table and slowly changing dimensions (SCD) type 1, it is necessary to use the dynamic cache.
- In the case of flat files, a static cache is used.

Conclusion

With abundant job opportunities and lucrative salary options, ETL testing has become a popular trend. ETL Testing has an extensive market share and is one of the cornerstones of data warehousing and business analytics. To make this process more organized and simpler, many software vendors have introduced ETL testing tools. Most employers who seek ETL testers look for candidates with specific technical skills and experience that meet their needs. No worries, this platform is a great resource for both beginners and professionals. In this article, we have covered 35+ ETL testing interview questions ranging from freshers to experienced level questions typically asked during interviews. Preparation is key before you go for your job interview.

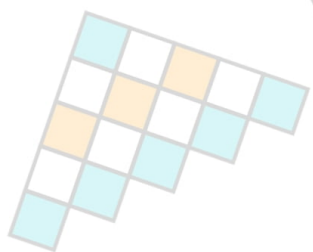
Recommended Resources:

[SQL](#)

[Python](#)

[Java](#)

[Informatica](#)



InterviewBit

Links to More Interview Questions

[C Interview Questions](#)

[Php Interview Questions](#)

[C Sharp Interview Questions](#)

[Web Api Interview Questions](#)

[Hibernate Interview Questions](#)

[Node Js Interview Questions](#)

[Cpp Interview Questions](#)

[Oops Interview Questions](#)

[Devops Interview Questions](#)

[Machine Learning Interview Questions](#)

[Docker Interview Questions](#)

[Mysql Interview Questions](#)

[Css Interview Questions](#)

[Laravel Interview Questions](#)

[Asp Net Interview Questions](#)

[Django Interview Questions](#)

[Dot Net Interview Questions](#)

[Kubernetes Interview Questions](#)

[Operating System Interview Questions](#)

[React Native Interview Questions](#)

[Aws Interview Questions](#)

[Git Interview Questions](#)

[Java 8 Interview Questions](#)

[Mongodb Interview Questions](#)

[Dbms Interview Questions](#)

[Spring Boot Interview Questions](#)

[Power Bi Interview Questions](#)

[Pl Sql Interview Questions](#)

[Tableau Interview Questions](#)

[Linux Interview Questions](#)

[Ansible Interview Questions](#)

[Java Interview Questions](#)

[Jenkins Interview Questions](#)