

[Product tour](#)[Science](#)[Test library](#)[Pricing](#)[Request demo](#)[Buy now!](#)[Log in](#)[Hiring & recruiting](#)[Human resources](#)[Pre-employment testing](#)[Remote hiring](#)[TestGorilla](#)

25 ETL developer interview questions and answers for recruiters



25 ETL developer interview questions and answers for recruiters



Share



In today's data-driven business landscape, few roles are more crucial than an ETL developer. These software specialists are responsible for designing complex data storage systems that are critical to business intelligence. In short, they turn Big Data into Big Wins.

ETL stands for [extract](#), [transform](#), and [load](#), referring to the general procedure for transferring data from multiple sources to a single data warehouse. Data is then used by analysts to make better business decisions.

Needless to say, if you need an ETL developer, it's vitally important to hire the right one for your company. It's no secret that [data-savvy businesses outperform their competitors](#).

Latest Posts



The Big 5 personality traits: How they can influence your hiring



101 B2B sales interview questions to ask your applicants



How to hire MongoDB developers for your business



25 Python debugging interview questions



44 cognitive ability interview questions

The best advice in pre-employment testing, in your inbox.

No spam. Unsubscribe at any time.

Recruiters should be mindful that ETL tools can vary widely. For this reason, you need to tailor the recruitment process to your organization's specific ETL needs where possible.

We recommend using a [skills assessment](#) before the interview, in order to shortlist your best candidates and concentrate on them only—and not spend time interviewing unsuitable applicants.

Ask these ETL developer interview questions to help you hire the best

Below, you'll find 25 of the most important ETL interview questions to help you make the best hire. We've split the questions into three groups—beginner, intermediate, and advanced—for different experience levels.

You'll also find sample answers for your reference, so that you can evaluate applicants' skills confidently, even if you don't have any ETL development skills yourself.

8 beginner ETL developer interview questions

If you're hiring for an entry-level ETL developer with up to 2 years of experience, consider using these beginner interview questions.



1. What are the key steps of the ETL process?

Sample answer:

While the abbreviation implies a neat, three-step process, ETL actually encompasses more steps. ETL refers to data extraction from different sources, its transformation and loading into a data warehouse, and the eventual analysis of the warehouse data.

In short, ETL comprises the transportation of data across all four areas: extraction, transformation, loading, and analysis.

2. Why is data warehousing important?

Sample answer:

Data warehousing is a core component of business intelligence. By bringing different data sources under a single, unified library, analysts can work more efficiently, get more in-depth insights, and spot patterns across different datasets.

Ultimately, it helps businesses be more competitive by improving their decision-making process.

3. What are the key differences between ETL and ELT?

Sample answer:

ETL transforms data before it's loaded into the target system, while ELT transforms data within the data warehouse. Out of the two, ELT is generally considered the better solution for large amounts of data,

Submit



**Hire the best.
No bias. No
stress.**

Our screening tests identify the best candidates and make your hiring decisions faster, easier, and bias-free.

Try for free

offering a more flexible and agile way of working with data.

4. What is meant by partitioning in ETL?

Sample answer:

Partitioning refers to the division of large datasets into smaller, more manageable areas based on shared characteristics. Its purpose is to make the data warehouse easier to navigate and improve query processing performance.

5. What types of applications and tools are used in ETL?

Sample answer:

There are a number of different ETL software tools on the market, but they all share the same purpose of data integration. Some of the most popular ETL tools are:

Informatica PowerCenter

IBM InfoSphere DataStage

Oracle Data Integrator

Microsoft SQL Server Integration Services (SSIS)

Xplenty

6. What are the different types of facts in ETL?

Sample answer:

Facts are quantitative pieces of information about a business, such as sales numbers or account balances. They're stored in a fact table. There are three different types of facts:

1. **Non-additive facts** cannot be summed up across any dimension in the fact table, such as a percentage
2. **Semi-additive facts** can be summed up for some, but not all, dimensions in the fact table, such as a headcount
3. **Additive facts** can be summed up across all dimensions in the fact table, such as sales

7. What is the role of a dimension table and how does it differ from a fact table?

Sample answer:

Fact tables contain measurements and metrics about a business. Dimension tables are adjoined to fact tables by a foreign key and show the descriptive characteristics of the facts within them. While fact tables are granular, dimension tables are wordy and detailed.

8. Why is ETL testing important and how can it be done?

Sample answer:

Regular testing is an essential part of the ETL process and ensures that data arrives in the analytics warehouse smoothly and accurately.

ETL testing can be performed in the following ways:

Review primary sources to make sure they have extracted without any data loss

Verify that the data has been transformed into the appropriate data type for the warehouse

Check that the warehouse accurately reports cases of invalid data

Document any bugs that occur during the ETL process

8 intermediate ETL developer interview questions

Next, you'll find some intermediate ETL developer questions. Use these when recruiting for mid-level roles where applicants already have some hands-on experience working with ETL.



9. How is data analyzed in ETL?

Sample answer:

Once data has been successfully transported into the data warehouse, analysts typically use third-party Business Intelligence (BI) applications such as Tableau to turn the raw data into graphs and charts, based on which business decisions can be made.

Some of the latest ETL tools are equipped with their own data analysis mechanisms.

10. What are some of the most common types of ETL testing?

Sample answer:

There are several ETL testing methods, each one serving a specific purpose at different points in the ETL process. Some of the most common types of ETL testing are:

Production validation: Data in the target system is compared against the sources for validity purposes

Source to target count testing: The number of loaded records in the data warehouse is checked for consistency against the expected record count

Performance testing: The loading process is tested to make sure it takes place within the expected time frames

Data transformation testing: Data in the target system is checked to verify that it has been correctly transformed in line with business requirements

11. What are snapshots in ETL?

Sample answer:

Snapshots are read-only copies of data from the master table. Before a change to the data warehouse is made, a snapshot is taken and stored locally for reference. This works to preserve records as they were before the change was made.

Each snapshot has three components:

1. The time it was taken
2. A key to identify the snapshot
3. The data relating to the key

12. What is data profiling in ETL?

Sample answer:

Data profiling is the process of reviewing source data to identify its structure, quality, and interrelationships.

It's an important precursor to the analysis stage, as it ensures that the appropriate source data is loaded into the data warehouse and that it is transformed in line with the business requirements.

13. What are some of the most common ETL bugs?**Sample answer:**

ETL testing can reveal a wide variety of problems. Some of these include:

- Loss of data between the source and the target system
- Lack of consistency between datasets
- Lengthy query processing
- System crashes due to the scale of the data warehouse
- Cosmetic bugs related to font, color, or alignment

14. What is meant by the three-tier architecture of ETL?**Sample answer:**

Most data warehouses comprise three separate areas. These are:

1. **The staging area**, where data is extracted from various sources and processed
2. **The data integration area**, where data from the staging area is transformed (sometimes called the OLAP server)
3. **The access area**, where transformed data is retrieved by users for analysis

15. Why is the staging area in ETL important?**Sample answer:**

The staging area is the landing zone for data extracted from sources and sits between the source and target in the ETL process.

Here, the data is cleansed and modified before it is transferred to the data warehouse. This is a more efficient alternative to transforming data in the target system itself.

16. What is the difference between an initial load and an incremental load in ETL?**Sample answer:**

Initial load refers to the process of loading all data from primary sources into the target system for the first time. Once this has been completed, all subsequent loads into the system will be incremental loads, where only new or modified records are brought in.

9 advanced ETL developer interview questions

Below, you can find nine advanced ETL developer questions. You can use these to assess the skills and knowledge of senior ETL developers who have plenty of experience.

17. What are some of the challenges of ETL testing?

Sample answer:

It's important to identify potential challenges early on in the ETL process to avoid bottlenecks further down the pipeline. Some of the most common problems and challenges of ETL testing are:

- Loss, corruption, or duplication of data during transportation
- Underperformance caused by large volumes of historical data
- Unachievable business requirements
- Limited availability of source data
- Outdated ETL tools

18. Can you explain the ETL testing process from start to finish?

Sample answer:

ETL testing is a demanding process that should be completed in the following order:

1. **Define the business requirements:** Liaise with the client to understand their reporting needs and define the scope of the project
2. **Validate data sources:** Perform a data count check and ensure check keys are in place
3. **Design ETL logic:** Design the mapping sheet, SQL script, and transformational codes
4. **Extract source data:** Identify any bugs during the extraction phase
5. **Transform data:** Make sure data is transformed consistently
6. **Load data:** Perform a record count check and verify the accuracy of the loaded data
7. **Review process:** Verify the validity, layout, and export functionality of the summary report
8. **File test report:** Share test results with the relevant stakeholders

19. What are some of the best ETL features that our company should be using?

Sample answer:

A good ETL tool makes the data integration process more efficient and user-friendly. Some particularly useful ETL features are:

- Cloud compatibility**, allowing for greater flexibility and better handling of massive datasets
- Third-party integrations** to connect with ERP platforms and BI tools
- Automatic code generation** to reduce the risks of human errors and speed up development
- Intuitive interface** to improve user navigation
- Sophisticated debugging tools**, which reduce disruption to data flows

20. What is the role of ETL in the data mining process?

Sample answer:

ETL is an important early phase of the data mining process. After data sources have been identified and business requirements are set, ETL is performed to bring all historical data under a single, consistent system.

From here, data is analyzed and modeled using BI tools. Data scientists are then able to evaluate the data to draw conclusions about business decisions.

21. What are the different types of partitioning in ETL, and when should they be used?

Sample answer:

The two main types of partitioning in ETL are:

Hash partitioning: Rows are assigned using a hash key, meaning partitions depend on the specified hashing algorithm

Round-robin partitioning: Rows are assigned in a round-robin manner, meaning each partition contains approximately the same number of rows

22. What is meant by regression testing in ETL?

Sample answer:

Regression testing is used after developing functional repairs to the data warehouse. Its purpose is to check if said repairs have impaired other areas of the ETL process.

Regression testing should always be performed after system modifications to see if they have introduced new defects.

23. What is the purpose of data purging and archiving?

Sample answer:

Data purging is the process of permanently deleting obsolete data from the data warehouse. For example, data may be purged once it becomes ten years old. This is done to free up space on the server and improve performance.

Data purging is usually accompanied by archiving, where data is moved to a separate storage device for long-term retention, typically for legal purposes.

24. What are the key differences between connected and unconnected lookups in ETL?

Sample answer:

There are several key differences between connected and unconnected lookups:

Connected lookups return multiple columns from the same row, whereas unconnected lookups return one column from each row

Connected lookups receive values directly from the mapping pipeline, whereas unconnected lookups receive values from a separate transformation

Connected lookups use a dynamic or static cache, whereas unconnected lookups only use a static cache

25. When would you perform a lookup transformation in ETL?

Sample answer:

Lookup transformation is used to retrieve values from a data source based on specific lookup conditions. There are a few scenarios when this may be necessary, for example, to:

- Update a dimension table
- Check if records already exist in a table
- Find a specific value from a table

At which stage of the hiring process should you use ETL developer interview questions?

Interviews are just one part of what should be a rigorous, multi-stage candidate selection process. It's important to draw on other recruitment tools and techniques when hiring your next ETL developer so that you can gain a more rounded view of each candidate's skill set.

Interviews work best at the back end of the selection process—using skills assessments and tasks beforehand allows you to rule out underqualified applicants at an earlier stage, helping you save time by only interviewing your best candidates.

Here is an outline for an effective ETL hiring process:

1. **Job posting:** Make sure to state the ETL tools used by your organization in the job advert and list the responsibilities of the role
2. **Resume screening:** Sift through CVs, shortlisting candidates that might have the right ETL skills required for the position
3. **Skills assessment:** Invite shortlisted candidates to complete an online skills assessment of up to five different tests—you could assess their ETL skills, cognitive abilities, or even their personality traits or culture add potential
4. **Screening call:** Have an informal chat with candidates to see if your expectations and goals align
5. **Role-specific task:** Present candidates with an ETL task similar to what they would do in the role, such as creating an automated data cleansing process
6. **Interview:** Draw on interview questions above, but also remember to assess the candidate's soft skills
7. **Reference check:** Perform checks on the candidate's background and references
8. **Final decision and job offer:** Make your final decision based on skills, personality and culture add potential before making a job offer

The highly specialized nature of ETL makes it even more important to include hands-on programming skills assessments and role-specific tasks as a part of the selection process.

Work samples are one of the most effective predictors of a candidate's likely performance and are used by many of the world's biggest corporations, including Google. Structured interviews, where you're asking all interviewees the same questions in the same order will help you accurately evaluate their performance and limit bias.

If you've already got a good idea of the candidate's ETL expertise by the time the interview rolls around, ask a few specific ETL questions and focus on the candidate's soft skills and personality fit. Any hiring decision should take into account attitude just as much as aptitude.

Finding your next ETL developer

By using interview questions alongside role-specific assessments, you can create a bulletproof selection process capable of unearthing the best ETL developer for your organization.

Remember to tailor the recruitment process to your company's ETL needs and the specific requirements of the role. This will help you to find a candidate with the right skills and, ultimately, maximize your chances of making a mutually beneficial long-term hire.

Try TestGorilla for free today and start making better hiring decisions.

Related posts

The Big 5 personality traits: How they can influence your hiring

You may think that personality doesn't matter much in the...

101 B2B sales interview questions to ask your applicants

If you're searching for a B2B sales expert to join...

How to hire MongoDB developers for your business

image listing why should your business hire MongoDB developers

Share

Leave a comment

Comment

Name

Email

☐ I agree to the terms and conditions laid out in the Privacy Policy

☐ Save my name, email, and website in this browser for the next time I comment.

Post a comment

Hire the best candidates with TestGorilla.

Create pre-employment assessments in minutes to screen candidates, save time, and hire the best talent.

Try for free

Free resources

Anti-cheating checklist

This checklist covers key features you should look for when choosing a skills testing platform.

Checklist

Download now

Onboarding checklist

This resource will help you develop an onboarding checklist for new hires.

Checklist

Download now

How to find candidates with strong attention to detail

How to assess your candidates' attention to detail.

Ebook

Download now

- ©

TestGorilla

2022. All

rights

reserved.
- Help center

Information for candidates

For subject-matter experts
- Careers

Blog

Contact
- Hiring!

Legal stuff

Privacy policy

Sitemap