

★ Member-only story

SQL Coding Interview With ChatGPT. Would You Pass It?



Gustavo R Santos

9 min read · 6 hours ago



ChatGPT interviewed me about SQL coding and concepts.



LLM Interview | Image generated by AI. Gemini 3, 2025. <https://gemini.google.com> | Not Medium subscriber, [read it here.](#)

Introduction

Structured Query Language (SQL) has existed for about 50 years. It was initially developed in IBM labs in the 1970s, but it was standardized by organizations like ISO only in 1986/87.

I have a feeling that at one time or another during your data journey, you might have faced SQL. After all, despite all the languages existing these days, from Python to R, or from HTML to JavaScript, SQL has endured.

Thinking about why that is, I would point out a couple of things:

- *It is simple to learn and to write*
- *It is stable and has few updates to learn over the years.*
- *It presents structured data, which is the easiest way to handle it.*

As data scientists, we obviously need data for our projects. And not only data but datasets with rows and columns that can be handled and prepared for analytics and modeling.

Knowing all of that, be aware that knowing SQL is almost a requirement if you are or want to be a Data Scientist.

According to data from 365 Data Science, around 82% of data scientists reported using SQL as one of their primary tools. The same research shows that if you are a data engineer, approximately 80% affirm that SQL is an essential tool!

No breaking news here. If you are in the data field, you must keep it sharp or brush up your SQL skills.

So this post is a mock interview designed by ChatGPT to test my knowledge of SQL. There is the conceptual piece and the coding piece. What I liked is that it summarizes concepts and exercises that would allow one to face most job interviews about SQL.

Let's dive into the questions and answers.

The Interview

To create this interview, I asked our known AI friend ChatGPT to simulate it. Here's the prompt:

Can you simulate a SQL coding interview with 5 questions for me to solve? I want it just like in a regular interview. The responses will be given here, without opening any IDE. You ask me something, wait for my answer, and then the next question until the end of the interview. You can create some mock data to show me and ask for the task to solve. Start easy and go to hard questions, like CTE/Windows.

Next, follow along for questions and answers.

Coding Interview

This is the data prepared by the LLM.

Mock Data: Employees Table

We'll use a simple `employees` table for our questions. Here's the structure and sample data:

id	name	department	salary	hire_date
1	Alice	HR	60000	2019-05-21
2	Bob	IT	75000	2021-08-15
3	Charlie	IT	80000	2020-07-10
4	David	HR	65000	2018-12-03
5	Eva	Finance	90000	2017-06-25

Sample table created by ChatGPT. Image by the author.

Reference table for the questions from ChatGPT. Image by the author.

And so the interview begins.

Question 1 (Easy)

Write an SQL query to fetch all employees from the IT department.

Answer:

```
SELECT *  
FROM employees  
WHERE department = 'IT';
```

Question 1 was indeed very easy. It was a simple `SELECT` statement with a `WHERE` filter to get the employees from IT. Nothing fancy. Let's move on.

Question 2 (Medium)

Write a query to find the average salary for each department. The result should show two columns: department and avg_salary.

Answer:

```
SELECT department, AVG(salary) as avg_salary
FROM employees
GROUP BY department
ORDER BY 2 DESC;
```

ChatGPT called it a medium-level question, but it is actually still pretty basic but is also important because it shows you know how to group columns and use grouping functions like `AVG()` on the `SELECT` statement.

Now we will go up a notch.

Question 3 (Intermediate — Using a CTE)

Write a query using a Common Table Expression (CTE) to find employees who earn more than the average salary of their respective departments. The result should include id, name, department, and salary.

Answer:

```
--Creating the CTE to find the average salary by department
WITH avg_sal AS(
SELECT department, AVG(salary) as avg_dpt_sal
FROM employees
GROUP BY department )

--Finding the salaries over > avg dept salary
SELECT e.id, e.name, e.department, e.salary
FROM employees e
INNER JOIN avg_sal a ON e.department = a.department
WHERE e.salary > a.avg_dpt_sal;
```

That is a nice question where we can show our skills and logic in creating CTEs, documenting steps with comments, and joining tables.

Question 4 (Advanced — Window Functions)

Write a query that assigns a rank to each employee within their department based on salary (highest salary = rank 1).

The result should include id, name, department, salary, and rank.

Answer:

```
SELECT id, name, department, salary,  
       RANK() OVER( PARTITION BY department ORDER BY salary DESC) AS rank  
FROM employees;
```

In this question, the interviewer would see that you can use Window functions. Since this function takes a little longer to write, it causes some confusion and not everyone uses it frequently or understands it. If you do, you can stand out in SQL interviews.

Great, let's check the last coding question.

Final Question

Write a query that returns only the highest-paid employee from each department using a CTE and a window function.

Your result should include id, name, department, and salary.

Answer:

```
--CTE to rank the employees by department  
WITH rank_ems AS (  
    SELECT id, department, salary,  
           RANK() OVER( PARTITION BY department ORDER BY salary DESC) AS rank  
    FROM employees  
)  
  
--Join Rank to the Main table  
--Show only highest paid employee by dept  
SELECT e.id, e.name, e.department, e.salary
```

```
FROM employees e
INNER JOIN rank_emps r ON e.id = r.id
WHERE rank = 1;
```

This last question brings it all together. You can show the interviewer that you can create a CTE, use Window functions, and join tables, which are most of what you will need to use on your daily job.

Well, this piece is finished. But ChatGPT still had the conceptual SQL interview for me. So here we go.

SQL Concepts

In this interview, we will review some SQL concepts that may be asked in most interviews. It summarizes a great deal of knowledge about the query language.

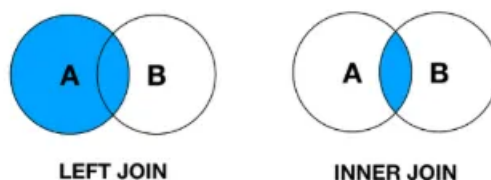
Question 1

Can you explain the difference between INNER JOIN and LEFT JOIN? When would you use each of them?

INNER JOIN returns only the rows where there's a match in both tables.

LEFT JOIN returns all rows from the left table, and matches from the right table where possible, returning **NULL** for non-matching rows. I would use **LEFT JOIN** if I need to know how many times a customer bought something, but I want to keep all the customers, even those with 0 transactions.

INNER JOIN would be used if I wanted to pull only the customers who bought and lose the ones who didn't.



Question 2

What is a subquery, and can you give an example of when you would use one?

A subquery is a *query within another query*, usually inside a `WHERE`, `FROM`, or `SELECT` clause. It allows you to perform an operation that depends on and/or can be used by the main query.

For example: if you're calculating the average spend for clients who bought product A but filtering by a store that's in a separate table, a subquery would be ideal to first identify the store and then use that to filter the clients.

Question 3

Can you explain what normalization is in the context of a relational database, and why it's important?

Normalization is the process of organizing a database to reduce redundancy and dependency. It involves dividing large tables into smaller, more manageable ones and establishing relationships between them. This ensures that each piece of data is stored only once, which prevents anomalies like duplication and inconsistency.

It is like organizing your closet. You should fold and arrange clothes in separate drawers, avoiding duplicate items. Then matching items together like shirts in one drawer, and pants in another. Finally list the brands of similar clothes (let's say shoes) just once, instead of repeating them over and over for every pair of shoes.

This keeps your database clean, consistent, and with better performance.

Question 4

What is a primary key, and what are its main properties? Can a table have more than one primary key?

A **primary key** is a column or set of columns in a table that uniquely identifies each row in that table. It ensures that no two rows have the same value in that key field.

It's also used to establish relationships between tables (for example, linking a `customer_id` in table 1 to a `customer_id` in table 2).

A table can only have **one primary key**.

Question 5

What do indexes do in a relational database, and why would you create an index on a column? Are there any potential downsides to creating too many indexes?

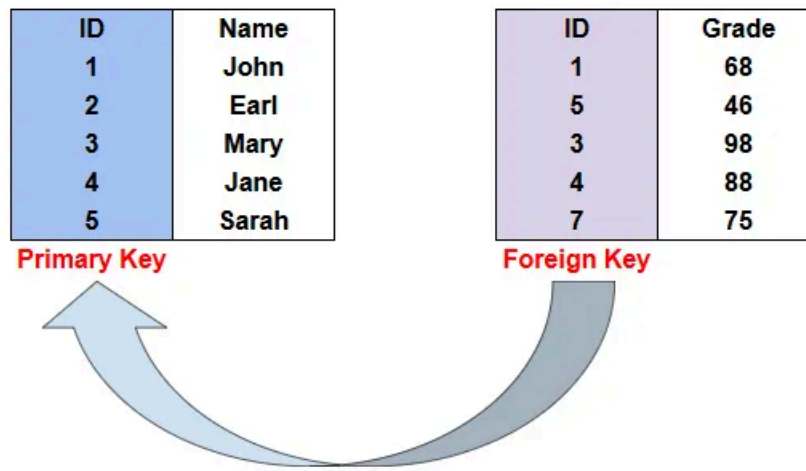
Indexes are used to improve the performance of queries by providing a faster way to look up data in a table. Think of an index for SQL just like a book index. That means that we don't need to scan page by page of a book to find a given subject. Instead, we should check the index and open the book right on the page we need. In a database, it's the same. We can use the index to directly jump to the location of the value.

Indexes are useful on columns that are frequently used in `WHERE`, `JOIN`, and `ORDER BY` clauses. However, we should use caution as too many indexes can increase the size of the DB. They are great for read-heavy workloads but can hurt write performance if overused.

Question 6

What is a foreign key, and how does it differ from a primary key? How is it used to maintain referential integrity in a database?

A **foreign key** is a column (or columns) of data in one table that refers to the unique data values (usually the primary key data) from another table. It establishes a relationship between the two tables by ensuring that each record in Table B refers to a valid record in Table A.



Foreign Key [ID] links records on Table_B to Table_A. Image by the author.

Final Question 7

What is ACID in the context of databases, and why is it important for ensuring data reliability?

ACID stands for:

1. **Atomicity** — Ensures that each transaction is a single piece. Either everything succeeds or fails, preventing corrupted or incomplete data.
2. **Consistency** — Ensures integrity of the data. The rules will be followed, like unique values for ID or only positive values for column *N*.
3. **Isolation** — Transactions are executed in isolation from one another, meaning that one transaction won't interfere with another.
4. **Durability** — Once a transaction is complete, it stays saved — even if the system crashes.

ACID is important because it guarantees that the database remains reliable, even when multiple transactions are occurring simultaneously or if there are unexpected failures. Without these properties, data could become inconsistent, lost, or corrupted.

And that ends our interview.

Before You Go

Now that we have gone through both parts, how do you feel about it?

Would you pass or fail this interview?

I believe that ChatGPT was able to condense a large portion of what's needed to know about SQL to be a good Data Scientist. So, if you understand the topics covered in this post, you are in a good spot for a SQL interview.

Obviously, it is not a guarantee of success, but I believe it covers more than half of the SQL interviews out there.

If you failed in any given topic, that's your cue to open Google (or an LLM) and study a little more.

If you liked this content, find out more about me and my work on my website.

<p>Gustavo R Santos</p> <p>I am a Data Scientist specializing in data analysis, machine learning, and visualization using Python, R, SQL, and...</p> <p>gustavorsantos.me</p>	
---	--

References

<p>Data Engineer Job Market in 2024: Analysis, Trends, and Opportunities - 365 Data Science</p> <p>We researched 1,000 job postings to identify the opportunities and requirements for data engineers in an AI-driven era...</p> <p>365datascience.com</p>	
---	--

<p>W3Schools.com</p> <p>W3Schools offers free online tutorials, references and exercises in all the major languages of the web. Covering...</p> <p>www.w3schools.com</p>	
--	--

--	--

ACID Properties in DBMS - GeeksforGeeks

ACID properties\u2014Atomicity, Consistency, Isolation, and Durability\u2014are essential for ensuring data...

www.geeksforgeeks.org

Sql

Job Interview

Data

ChatGPT

Data Science



Written by Gustavo R Santos

Edit profile

2.9K Followers · 33 Following

Data Scientist | I solve business challenges through the power of data. | Visit my site: <https://gustavorsantos.me>

No responses yet



What are your thoughts?

Respond

More from Gustavo R Santos



In Data Science Collective by Gustavo R Santos

Learn FastAPI with a Full-Stack Project: Movie Recommendation

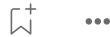


In TDS Archive by Gustavo R Santos

Why You Must Use Plotly Express for Data Visualization

Learn how FastAPI serves a front-end Streamlit application.

★ 5d ago 🖱 37



In Data Hackers by Gustavo R Santos

Crie Dashboards Simples e Estáticos no R

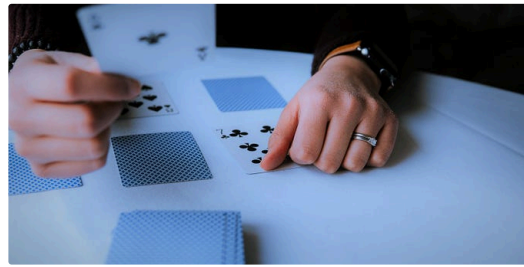
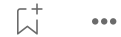
Use as bibliotecas ggplot2 e patchwork para criar ótimas visualizações

★ Jun 10, 2021 🖱 2



This visualization library enhances your productivity and provides better insights

★ Apr 3, 2022 🖱 99



In gustavorsantos by Gustavo R Santos

Introduction to Python deque()

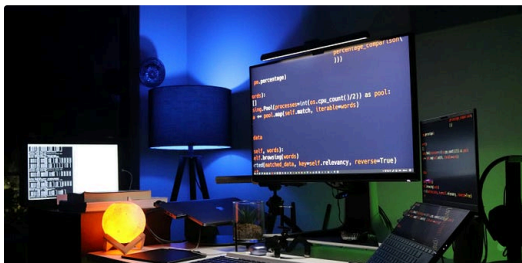
Understand how this type of collection works.

★ Sep 24, 2021 🖱 60



See all from Gustavo R Santos

Recommended from Medium



Vijay Gadhave

When to Use COUNT(*) vs COUNT(1) in SQL Queries

Note: If you're not a medium member, [CLICK HERE](#)



In Level Up Coding by Attila Vágó

The Meta Mistake And Why Nobody Is Immune To Low...

Not even the rockstar developers, ninja coders, or the 10x engineers...

Lists



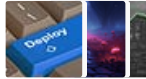
ChatGPT prompts
51 stories · 2599 saves



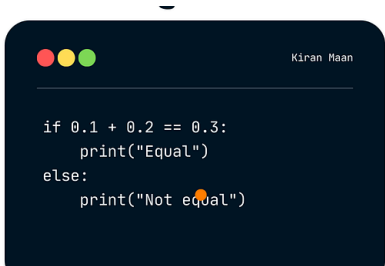
The New Chatbots: ChatGPT, Bard, and Beyond
12 stories · 557 saves



ChatGPT
21 stories · 979 saves



Predictive Modeling w/ Python
20 stories · 1838 saves

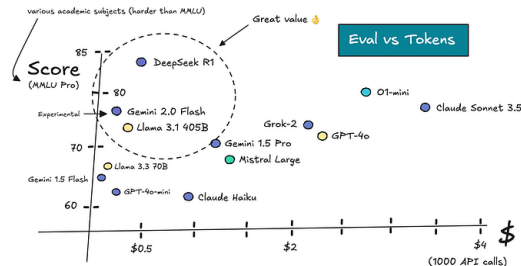


PY In Python in Plain English by Kiran Maan

Why 0.1 + 0.2 ≠ 0.3 in Python (How I Fix It)

Shocking truth about Python Floats

3d ago 337 7

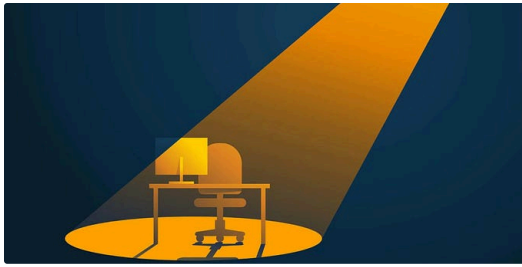


DSC In Data Science Collective by Ida Silfverskiöld

Economics of LLMs: Evaluations vs Pricing

Looking at which model to use for which task

5d ago 633 5



In Entrepreneurship Handbook by Joe Procopio

The Real Reason No One Is Hiring or Getting Hired

All the dots are connecting to one person, and that person needs to start doing their job



Feb 17



4.3K



100



Harshiha Blogs

Deutsche Bank Interview Experience — Java Backend

The interview was held face-to-face in their office. The Rounds were scheduled as follows:

Feb 9



138



2



See more recommendations