SQL Interview Questions: 3 Tech Screening Exercises (For Data Analysts)

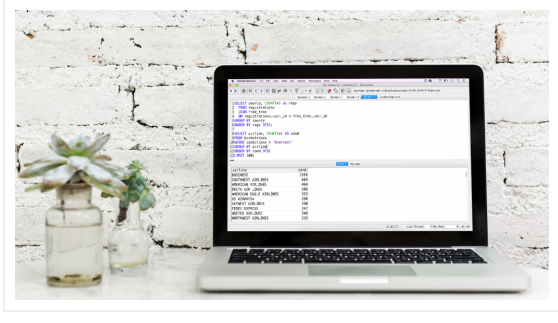
Written by Tomi Mester on August 28, 2018

Last updated on September 04, 2019

Over the last 6 years I've been part of many job interviews – on both sides of the table. The most fun, but also the most feared, part of the process is the technical screening. In this article, I'll show you three SQL test exercises that, in my experience, are quite typical in data analyst job interviews. *(And hey, these are "sample" SQL interview questions but they are heavily based on reality!)*

Before the tasks – What can you expect in an SQL technical screening?

There are two common ways an SQL tech screening can be done.



The simpler but less common way is that you get a computer, a data set and a task. While you are solving the task, the interviewers are watching and asking questions. A little trial and error is totally fine, as long as you can come up with the correct solution in a reasonable amount of time.

The other, more difficult (and by the way much more common) way is the **whiteboard interview.** In this case, you don't get a computer. You have to solve the task and sketch up the code on a whiteboard. This means that you won't get feedback (at least not from a computer) on whether you made a logical or a syntax error in your code. Of course, you can still solve the tasks by thinking iteratively (cracking the different SQL problems one by one), but you have to be very confident with your SQL skills.

Additionally, usually you have to solve the tasks on the fly. Maybe you will get 3-5 minutes of thinking time but that's the maximum you can expect. The good news is that because of that you will get relatively simpler tasks. (See the difficulty level below!)

Note: there are other types of tech screening – like the take-home assignment – where you can prove that you can solve more complex coding challenges, too.

Test yourself!

Here are three SQL interview questions that are really close to what I actually got or gave on data analyst/scientist job interviews!
Try to solve all of them as if they were whiteboard interviews!
In the second half of the article, I'll show you the solutions, too!

SQL Interview Question #1

Let's say you have two SQL tables: `authors` and `books`.
The `authors` dataset has 1M+ rows; here's the first six rows:

author_name	book_name
author_1	book_1
author_1	book_2
author_2	book_3
author_2	book_4
author_2	book_5
author_3	book_6
…	…

The `books` dataset also has 1M+ rows and here's the first six:

book_name	sold_copies
book_1	1000
book_2	1500
book_3	34000
book_4	29000
book_5	40000
book_6	4400
…	…

Create an SQL query that shows the TOP 3 authors who sold the most books in total!

(Note: I got a very, very similar SQL interview question for a data scientist position at a very well-known Swedish IT company.)

SQL Interview Question #2

You work for a startup that makes an online presentation software. You have an event log that records every time a user inserted an image into a presentation. (One user can insert multiple images.) The `event_log` SQL table looks like this:

user_id	event_date_time
7494212	1535308430
7494212	1535308433
1475185	1535308444
6946725	1535308475
6946725	1535308476
6946725	1535308477
…	…

…and it has over one billion rows.

Note: If the `event_date_time` column's format doesn't look familiar, google "epoch timestamp"!

Write an SQL query to find out how many users inserted more than 1000 but less than 2000 images in their presentations!

(Note: I personally created and used this interview question to test data analysts when I was freelancing and my clients needed help in their hiring process.)

SQL Interview Question #3

You have two SQL tables! The first one is called `employees` and it contains the employee names, the unique employee ids and the department names of a company. Sample:

department_name	employee_id	employee_name
Sales	123	John Doe
Sales	211	Jane Smith
HR	556	Billy Bob
Sales	711	Robert Hayek
Marketing	235	Edward Jorgson
Marketing	236	Christine Packard
…	…	…

The second one is named `salaries`. It holds the same employee names and the same employee ids – and the salaries for each employee. Sample:

salary	employee_id	employee_name
500	123	John Doe
600	211	Jane Smith
1000	556	Billy Bob
400	711	Robert Hayek
1200	235	Edward Jorgson
200	236	Christine Packard
…	…	…

The company has 546 employees, so both tables have 546 rows.

Print every department where the average salary per employee is lower than $500!

(Note: I created this test question based on a real SQL interview question that I heard from a friend, who applied at one of the biggest social media companies (name starts with 'F.' :))

Solution of SQL Interview Question #1

The solution code is:

```
SELECT authors.author_name, SUM(books.sold_copies) AS sold_sum
FROM authors
JOIN books
ON books.book_name = authors.book_name
GROUP BY authors.author_name
ORDER BY sold_sum DESC
LIMIT 3;
```

And here is a short explanation:

1. First you have to initiate the `JOIN`. I joined the two tables by using:

```
SELECT *
FROM authors
JOIN books
ON books.book_name = authors.book_name;
```

2. After that, I used a `SUM()` function with a `GROUP BY` clause. This means that in the `SELECT` statement I had to replace the `*` with the `author_name` and `sold_copies` columns. (It's not mandatory to indicate from which table you are selecting the columns, but it's worth it. That's why I used **`authors.`**`author_name` and **`books.`**`sold_copies`.)

3. Eventually, I `ORDER` ed the results in `DESC` ending order. (Just for my convenience, I also renamed the `sum` column to `sold_sum` using the `AS sold_sum` method in the `SELECT` statement.)

Solution of SQL Interview Question #2

The SQL query is:

```
SELECT COUNT(*) FROM
  (SELECT user_id, COUNT(event_date_time) AS image_per_user
  FROM event_log
  GROUP BY user_id) AS image_per_user
WHERE image_per_user < 2000 AND image_per_user > 1000;
```

The trick in this task is that you had to use the `COUNT()` function two times: first, you had to count the number of images per user, then the number of users (who fulfill the given condition). The easiest way to do that is to use a subquery.

1. Write the inner query first! Run a simple `COUNT()` function with a `GROUP BY` clause on the `event_log` table.
2. Make sure that you create an alias for the subquery (`AS image_per_user`). It's a syntax requirement in SQL.

3. Eventually, in an outer query, apply a `WHERE` filter and a `COUNT()` function on the result of the subquery.

Solution of SQL Interview Question #3

Solution:

```
SELECT department_name, AVG(salaries.salary) AS avg_salaries
FROM employees
JOIN salaries
ON employees.employee_id = salaries.employee_id
GROUP BY department_name
HAVING AVG(salaries.salary) < 500;
```

Note: You can solve this task using a subquery, too – but in an interview situation the committee will like the above solution better.

Brief explanation:

1. First JOIN the two tables:

```
SELECT *
FROM employees
JOIN salaries
ON employees.employee_id = salaries.employee_id
```

Watch out! Use the `employee_id` column – not the `employee_name` . (You can always have two John Does at a company, but the employee id is unique!)

2. Then use an `AVG()` function with a `GROUP BY` clause — and replace the `*` with the appropriate columns. (Just like in the first task.)

3. And the last step is to use a `HAVING` clause to filter by the result of the `AVG()` function. (Remember: `WHERE` is not good here because it would be initiated before the `AVG()` function.)
Watch out: in the `HAVING` line, you can't refer to the alias – you have to use the whole function itself again!

Prepare for SQL tech screenings by practicing!

If you managed to solve all these questions properly, you are probably ready for a junior or even for a mid-level Data Analyst SQL technical screening.
If not, let me recommend my new online course: SQL for Aspiring Data Scientists (7-day online course) – where you can level up (or brush up) your SQL skills in only 7 days. When you finish the course, just come back to this article and I guarantee that you will be able to solve these questions!

SQL for Aspiring Data Scientists (7-day online course)

And if you are just about to start with SQL, start with my SQL For Data Analysis series on the blog!

Conclusion

The hard part of these SQL interview questions is that they are abstract. The tasks say to "imagine the data sets" and show only a few lines of them. When you get an exercise like that, it helps a lot if you have seen similar datasets and solved similar problems before. I hope solving the tasks in this article will boost your confidence!

If you have questions or alternative solutions, don't hesitate to leave a comment.

- If you want to learn more about how to become a data scientist, take my 50-minute video course: How to Become a Data Scientist. (It's free!)
- Also check out my 6-week online course: The Junior Data Scientist's First Month video course.

Cheers,
Tomi Mester

August 28, 2018 In Coding In Data Science and Analytics
#analytics #data coding #data science #learn data science #learn sql #postgresql #sql #sql for data analysis #tomi mester

← PREVIOUS POST NEXT POST →

28 Comments

Xiao
AUGUST 30, 2018

Thanks Tomi,

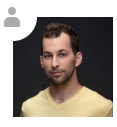
These three questions are very good.

For the second question, could it use the having function like Q3:

```
SELECT user_id, COUNT(event_date_time) AS image_per_user
FROM event_log
GROUP BY user_id
Having COUNT(event_date_time) > 1000 and COUNT(event_date_time) < 2000;
```

Regards,
Xiao

REPLY



Tomi Mester
AUGUST 30, 2018

hi Xiao,

good question – and you are spot-on!
You can use HAVING in the inner query instead of using WHERE in the outer query.

But you will still need the subquery solution to actually count the number of the users. (Your query currently only returns the users but it doesn't count them.)

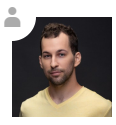
Cheers,
Tomi

REPLY

Peter Pan
SEPTEMBER 11, 2018

The answer by Xiao is correct. You do not need a subquery for what you are asking for. Tomi, no offense, but you fail this question by simply overcomplicating things.

REPLY



Tomi Mester
SEPTEMBER 17, 2018

hey Peter,

thanks for the comment and none taken – comments with better solutions and with bug fixes are always welcome.
But I think in this case my solution is the correct one:
Xiao's answer returns the correct users but it doesn't count them as requested by the task. (The output is a list and not a number.)
I agree though that the WHERE clause from the outer query can go as HAVING in the inner query.

Cheers,
Tomi

REPLY

David
OCTOBER 25, 2018

These two commenters need to read the question a little better!

Fun practice questions, thanks Tomi.

Tomi Mester
OCTOBER 31, 2018

Thanks David!

Charles
MARCH 18, 2019

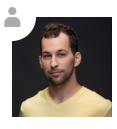
Also, agree that these are great examples.

To Xiao's point, given that the selection is being grouped by user_id, you could simply look at the number of records in his query to find the count.

Appreciate you posting these questions!

Best,
Charles

REPLY



Tomi Mester
MARCH 22, 2019

Thanks Charles!
Cheers!

REPLY

vanny
NOVEMBER 15, 2018

thank you for the blog..very useful information for the job applier.

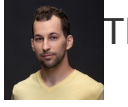
REPLY

Rida N
FEBRUARY 18, 2019

Super helpful questions, Tomi. Thank you! It would be great if you can add a couple more practice SQL questions.

REPLY

Tomi Mester
MARCH 6, 2019



Thanks, I'll add more!

REPLY

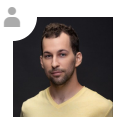
SUNAINA SONI

FEBRUARY 18, 2019

Solution by Tomi Mester for question#2 is absolutely correct!

Thanks Tomi for these practice questions. It helps alot.

REPLY



Tomi Mester

MARCH 6, 2019

Thanks! :) Glad you liked it!

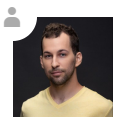
REPLY

Ketki

MARCH 14, 2019

Great set Tomi. Would appreciate if you can add more SQL questions.

REPLY



Tomi Mester

OCTOBER 18, 2020

I have more in my SQL course here: https://data36.com/sql-for-aspiring-data-scientists-7-day-online-course/

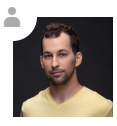
REPLY

John Howard

APRIL 6, 2019

Nice questions; Very helpful as I'm brushing up on my SQL abilities. Thank you so much for these.

REPLY



Tomi Mester

OCTOBER 19, 2020

thank you! more is yet to come!

REPLY

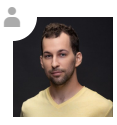
Rahul Kumrawat

JUNE 18, 2019

Really Helpful, please add some more queries for practice.

Thanks Tomi!

REPLY



Tomi Mester
OCTOBER 19, 2020

will do that soon!

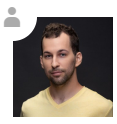
REPLY

Michelle
MARCH 31, 2020

Hi there, is there any way to access all the SQL for Aspiring Data Scientists (7-day online course) materials at once? My interview is in two days and I was hoping this will be a crash course! Could you please unlock all my sessions?

Hope to hear from you. Thanks!

REPLY



Tomi Mester
OCTOBER 13, 2020

Yeah, it's possible!
When you enroll, I can manually give you access for the full course if you ping me in email right after the registration!
Tomi

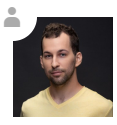
REPLY

Modupe
MAY 18, 2020

This was great Tomi, i have an interview tomorrow and this has been a really helpful practice. More questions will definitely be appreciated.

Mo.

REPLY



Tomi Mester
OCTOBER 13, 2020

Cheers — and I hope that it went well!

REPLY

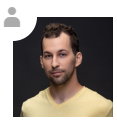
Akanksha Kamra
MAY 21, 2020

Hi Tomi,

Thanks for sharing these questions, can you please add more practice questions.

Best Regards,
Akanksha

REPLY



Tomi Mester

OCTOBER 13, 2020

Sure, I'll come up with a few more soon!

Tomi

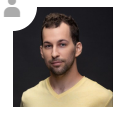
REPLY

santhosh

AUGUST 12, 2020

Very good questions and good illustration. Good way to test ourselves before we see your solutions.

REPLY



Tomi Mester

OCTOBER 4, 2020

Thanks!!

REPLY

Leave a Reply

COMMENT

NAME *

EMAIL *

WEBSITE

Post Comment