**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!**

Answer:

SELECT author\_name

FROM authors, books

WHERE authors.book\_name = books.book\_name

GROUP BY author\_name

ORDER BY sum(sold\_copies) DESC

LIMIT 3

**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 many 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: The*event\_date\_time*column’s format is an epoch timestamp*

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

Answer:

SELECT count(user\_id)

FROM event\_log

GROUP BY user\_id

HAVING count(event\_data\_time) >1000 and count(event\_data\_time) < 2000

**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!**

Answer:

SELECT department\_name

FROM employees, salaries

WHERE employees.employee\_id = salaries.employee\_id

and employees.employee\_name = salaries. employee\_name

GROUP BY department\_name

HAVING avg(salary) < 500