Ace the upcoming Data Science Interview

You can't anticipate every question an interviewer will ask. However, there are many **critical questions** that you can prepare before the interview.

Our hiring partners have helped us curate a set of interview questions on key skills, which will help you prepare better for the data science job roles.

Filters

1. Given a table(cars) with 4 columns(model_id, model_name,color, price) , perform groupby using model_name and color, order by highest price, get 3rd highest.

Basic SQL

Hint?

Order of execution is extremely important for interviews. Order of query execution: From, Join, Where, Groupby, Having, Select, Order by, Limit/offset.
https://sqlbolt.com/lesson/select_queries_order_of_execution

2. What is the difference between the WHERE and HAVING clauses?

Basic SQL

Hint?

Where filters the base data where as having filters grouped data. Though having is accepted in mysql without a groupby function, it should never be used seperately.
https://sqlbolt.com/lesson/select_queries_order_of_execution

3. Given a table(employee). Find the Second highest salary. Find the 10th highest salary. Find the 25-30th highest salary.

Intermediate SQL

Hint?

When picking a nth value, always use dense_rank. https://www.geeksforgeeks.org/find-nth-highest-salary-table

4. Fetch department-wise salary from an employee table

Basic SQL

Hint?

Simple group-by function. Sum(salary) group-by department

5. Given a table with order-id , order item-id and quantity Find the quantity for distinct order-id

Basic SQL

Hint?

Simple group-by function. Sum(quantity) group-by order-id

6. What are the different type of Joins in Sql and explain them? (Mainly focused on full outer join)

Basic SQL

Hint?

While explaining theoretical concepts, always use examples.

7. Given 2 tables and the following query. What will be the output (select * from table 1 full outer join table 2) where values not in (select * from table 1 inner join table 2)

Intermediate SQL

Hint?

It's essential to imagine what an output would look like like when 2 tables are joined (Outer i.e L/R vs Inner) In this case, outer join will fetch all records, inner join will fetch only common records in both tables. Hence output will join the two tables excluding all common records

8. Given an assumption: There are 2 tables, first table has 10 records and second table has 15 records. There are 5 records common in both the tables. Number of records that would be fetched when you perform left join/right join/inner join/cross-join.

Basic SQL

Hint?

It's essential to imagine what an output would look like like when 2 tables are joined (Outer i.e L/R vs Inner) Full outer - 25, left - 10, right - 15, inner - 5, cross - 150

9. Given a word "JOE", find the word in a given string irrespective of word being upper case or lower case or capitalize?

Intermediate SQL

Hint?

SQL server by default is NOT case sensitive. Other potential questions could include fetching case senitive data. Read about - Upper(), Lower()

10. Find out if the database has any duplicate record names.

Basic SQL

Hint?

Finding duplicate and null values in python and SQL is a very common interview question

Finding duplicate and null values in python and SQL is a very common interview question

11. Differentiate between Implicit vs Explicit Join

Intermediate SQL

Hint?

Implicit join is not advised as per industry standards. https://riptutorial.com/sql/example/938/implicit-join

12. With respect to SQL, which one is more preferable - Subqueries or Joins? Why?

Intermediate SQL

Hint?

This varies on a case to case basis. However, a lot of preference is given to Joins in terms of usage / speed.

13. Does SQL have User Defined Functions?

Basic SQL

Hint?

Read about stored procedures and PL/SQL

14. Query to find the employees in the office given check in and check out as fields.

Intermediate SQL

Hint?

Always solve scenario based questions with example records

15. Given a table of an event having columns date-ts/ event id. Find the event that happened 3rd on every month

Basic SQL

Hint?

Simple date function. Filter day(date-ts) = 3. Date/Time functions are very frequent in interviews.

16. Split a full name into 2. First and last.

Basic SQL

Hint?

string_split()

17. Find the Salary greater than Average salary without using Joins or Sub-Queries

Advanced SQL

Hint?

Fetch average salary in the inner query and filter using outer.

18. What is difference between rownum and dense rank ?

Basic SQL

Hint?

Window functions: rank(), dense_rank(), first_value(), last_value(), nth_value(), lead(), lag()

19. How will you use partition by?

Basic SQL

Hint?

Window functions: General syntax

20. Types of Joins

Basic SQL

Hint?

Always explain joins with examples and venn diagrams. Interveiwers could further this by giving two sample tables and predict output for different types of joins "https://www.geeksforgeeks.org/sql-join-

set-1-inner-left-right-and-full-joins/ https://www.deskbright.com/sql/sql-joins-interview-questions/ https://www.techmixing.com/2019/04/sql-joins-tricky-interview-questions.html"

21. What is inner join?

Basic SQL

The INNER JOIN creates a new result table by combining column values of two tables (table1 and table2) based upon the join-predicate. The query compares each row of table1 with each row of table2 to find all pairs of rows which satisfy the join-predicate.

22. What is left / outer/inner join?

Basic SQL

(INNER) JOIN: Returns records that have matching values in both tables

LEFT (OUTER) JOIN: Returns all records from the left table, and the matched records from the right table

RIGHT (OUTER) JOIN: Returns all records from the right table, and the matched records from the left table

FULL (OUTER) JOIN: Returns all records when there is a match in either left or right table

23. What is Normalization

Basic SQL

Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling.

24. Use count function in a query

Basic SQL

```
SELECT COUNT(*) FROM dataset;
```

25. Difference between count(column_name) and count(*)

Intermediate SQL

COUNT(*) will count the number of records.

COUNT(column_name) will count the number of records where column_name is not null.

26. Write SQL query to find the cumulative price of each customer in a table?

Intermediate SQL

Step1: Partition data using OVER Clause

```
SELECT CustomerID
,TransactionDate
,TransactionAmount
,SUM(TransactionAmount) OVER(PARTITION BY TransactionDate) RunningTotal
FROM Sales.CustomerTransactions T1
WHERE TransactionTypeID = 1
GROUP BY CustomerID
ORDER BY InvoiceID
,TransactionAmount
```

Step2: Order Partitions with Order BY

```
SELECT CustomerID
,TransactionDate
,TransactionAmount
,SUM(TransactionAmount) OVER(PARTITION BY TransactionDate ORDER BY InvoiceID) RunningTotal
FROM Sales.CustomerTransactions T1
WHERE TransactionTypeID = 1
GROUP BY CustomerID
```

```
ORDER BY InvoiceID
,TransactionAmount
```

27. Write a query to delete duplicate records in a table

Intermediate SQL

```
SELECT [FirstName],
[LastName],
[Country],
COUNT(*) AS CNT
FROM [SampleDB].[dbo].[Employee]
GROUP BY [FirstName],
[LastName],
[Country]
HAVING COUNT(*) > 1;
```

28. What is a constraint in SQL?

Intermediate SQL

Hint?

Constraints are the rules enforced on the data columns of a table. These are used to limit the type of data that can go into a table. This ensures the accuracy and reliability of the data in the database.

29. What are the constraints type available in SQL

Intermediate SQL

Hint?

Some of the constraints in SQL are – Primary Key, Foreign Key, Unique Key, SQL Not Null, Default, Check and Index constraint.

30. What is a Primary Key

Basic SQL

The PRIMARY KEY constraint uniquely identifies each record in a table.

Primary keys must contain UNIQUE values, and cannot contain NULL values.

A table can have only ONE primary key; and in the table, this primary key can consist of single or multiple columns (fields).

31. Can we have multiple keys for primary key

Basic SQL

A table can have only one primary key, which may consist of single or multiple fields. When multiple fields are used as a primary key, they are called a composite key. If a table has a primary key defined on any field(s), then you cannot have two records having the same value of that field(s).

32. What is a Unique Key ?

Basic SQL

Unique key constraints identify an individual tuple uniquely in relation or table. A table can have more than one unique key, unlike primary key. Unique key constraints can accept only one NULL value for column. Unique constraints are also referenced by the foreign key of another table. It can be used when someone wants to enforce unique constraints on a column and a group of columns which is not a primary key.

33. What is a Foreign Key ?

Basic SQL

A FOREIGN KEY is a key used to link two tables together.

A FOREIGN KEY is a field (or collection of fields) in one table that refers to the PRIMARY KEY in another table.

The table containing the foreign key is called the child table, and the table containing the candidate key is called the referenced or parent table

34. Can a table contain multiple FOREIGN KEY's?

Basic SQL

A table may have multiple foreign keys, and each foreign key can have a different parent table.

35. What is SQL NOT NULL constraint?

Intermediate SQL

A NOT NULL constraint in SQL is used to prevent inserting NULL values into the specified column, considering it as a not accepted value for that column. This means that you should provide a valid SQL NOT NULL value to that column in the INSERT or UPDATE statements, as the column will always contain data

36. What is a CHECK constraint?

Intermediate SQL

The CHECK constraint is used to limit the value range that can be placed in a column. If you define a CHECK constraint on a single column it allows only certain values for this column. If you define a CHECK constraint on a table it can limit the values in certain columns based on values in other columns in the row.

37. What is a DEFAULT constraint?

Intermediate SQL

The DEFAULT constraint is used to provide a default value for a column. The default value will be added to all new records IF no other value is specified.

38. What is the difference between NULL value, Zero, and Blank space?

Intermediate SQL

A NULL value is not same as zero or a blank space. A NULL value is a value which is 'unavailable, unassigned, unknown or not applicable'. Whereas, zero is a number and blank space is a character

39. What is a Composite key ?

Intermediate SQL

A composite key is a combination of two or more columns in a table that can be used to uniquely identify each row in the table when the columns are combined uniqueness is guaranteed, but when it is taken individually it does not guarantee uniqueness.

40. How do you restrict the data at columns level ?

Intermediate SQL

The SQL SELECT LIMIT statement is used to retrieve records from one or more tables in a database and limit the number of records returned based on a limit value

41. write a query f_name and l_name fields from table emp and allow a space in between the 2 columns

Intermediate SQL

```
select ID, f_name + ' ' + l_name from emp
```

42. Write a query to rename the column name id as emp_id, name as emp_name for the table emp;

Basic SQL

```
select id as emp_id, name as emp_name from emp;
```

43. select * from dual, what does the dual mean and what is the default data types ?

Intermediate SQL

The DUAL is special one row, one column table present by default in all Oracle databases. The owner of DUAL is SYS (SYS owns the data dictionary, therefore DUAL is part of the data dictionary.) but DUAL can be accessed by every user.

The table has a single VARCHAR2(1) column called DUMMY that has a value of 'X'. MySQL allows DUAL to be specified as a table in queries that do not need data from any tables. In SQL Server DUAL table does not exist, but you could create one.

44. How do you get the current system date using dual table?

Intermediate SQL

```
SELECT sysdate FROM DUAL ;
```

45. Write a query to get the number of records from a table emp

Basic SQL

```
SELECT COUNT(*) FROM emp;
```

46. Explain DDL with examples.

Basic SQL

DDL is Data Definition Language and is used to define the structures like schema, database, tables, constraints etc. Examples of DDL are create and alter statements.

47. Explain DML with examples.

Basic SQL

DML is Data Manipulation Language and is used to manipulate data. Examples of DML are insert, update and delete statements.

48. Explain DCL and TCL with examples

Intermediate SQL

DCL is Data Control Language

TCL is Transaction Control Language

Examples under DCL: GRANT, REVOKE

Examples under TCL: START TRANSACTION, COMMIT, ROLLBACK

49. How to get only the delhi records from emp table, and handle all types of case sensitive issues: Delhi, delhi, DELHI, DELhi

Intermediate SQL

select * from emp where upper(city)='DELHI'

50. Write a query to change the format of the date to (YYYY-MON-DD) in dual table ?

Intermediate SQL

select to_char(date_column,'YYYY-MM-DD') from dual;

51. How to remove duplicate in the col from a table ?

Basic SQL

SELECT DISTINCT column FROM table1;

52. Write a query to find count of unique id in the retail_shopping table

Basic SQL

SELECT count(DISTINCT ID) FROM retail_shopping;

53. Write a query to select only the id, name, city,country and phone from the table customer and restrict the record only to india

Basic SQL

select id, name, city, country, phone from customer where country='india'

54. Write a query to update the table emp, where the city name is Madras to chennai

Basic SQL

```
UPDATE emp
SET city= 'chennai'
WHERE city= 'Madras';
```

55. Write a query to remove record whose salary is great than or equal to 50000 and city is chennai

Basic SQL

```
DELETE FROM emp WHERE salary>=50000 and city='chennai'
```

56. Write a query to select all the students from table stud whose name begins with 'S'

Intermediate SQL

```
select * from stud where name like "S%"
```

57. Write a query to display all the records from table emp where the age is between 18 and 58

Basic SQL

```
select * from emp where age between 18 and 58
```

58. Select all the record for emp, in which gender is female or age > 18

Basic SQL

select * from emp where gender='female' or age>18

59. Write a query to exact all the record for which payment_detail col is null for the table payment_detail

Basic SQL

select * from payment_detail where payment_detail is null

60. Write a query to get the top 5 salary from the table emp

Basic SQL

select top 5 * from emp

61. Query the records from emp where order is descending for name and ascending for salary

Intermediate SQL

select * from emp order by name descending, salary

62. What is the difference between union and union all in SQL

Intermediate SQL

UNION removes duplicate records (where all columns in the results are the same), UNION ALL does not.

There is a performance hit when using UNION instead of UNION ALL, since the database server must do additional work to remove the duplicate rows, but usually, you do not want the duplicates (especially when developing reports).

63. What is an execution plan? When would you use it? How would you view the execution plan

Intermediate SQL

An execution plan is a window in SQL Server Management Studio to shows you how SQL Server breaks down a query and also identifies where issues might exist within the execution plan. By identifying the statements that take a long time to complete, you can then look at the execution plan to determine tuning needs.

When do you use it?

You can use it anytime you write a query. Most developers use execution plan when they have database queries consumes a lot of resources and takes time.

How do you view it in SQL Server?

SQL Server can create execution plans in two ways:

Actual Execution Plan - (CTRL + M) - is created after execution of the query and contains the steps that were performed

Estimated Execution Plan - (CTRL + L) - is created without executing the query and contains an approximate execution plan

Execution plans can be presented in these three ways.

Text Plans

Graphical Plans

XML Plans

64. How can you select all the even number records from a table? All the odd number records?

Intermediate SQL

Select * from table where id % 2 = 0

Select * from table where id % 2 != 0

65. What is the difference between the RANK() and DENSE_RANK() functions? Provide an example.

Intermediate SQL

The one and the only difference between the DENSE_RANK() and RANK() functions is the fact that RANK() will assign non-consecutive ranks to the values in a set in the case of a tie, which means that with RANK() there will be gaps between the integer values when there is a tie. But the DENSE_RANK() will assign consecutive ranks to the values in the case of a tie, so there will be no gaps between the integer values in the case of a tie.

66. What is the difference between char and varchar2?

Intermediate SQL

CHAR is used for storing fix length character strings. It will waste a lot of disk space if this type is used to store variable-length strings.

VARCHAR2 is used to store variable-length character strings.