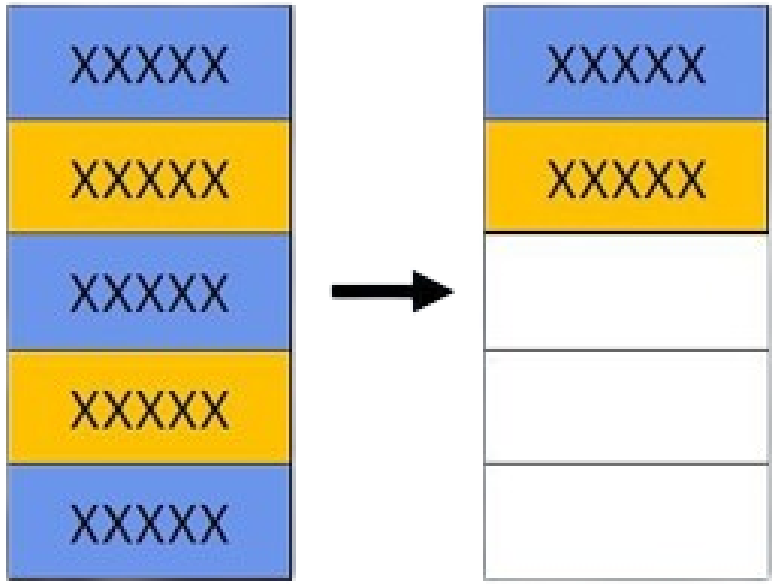
SQL SERIES

# How to Remove Duplicate Data in SQL

Duplicates can be a big problem in SQL databases as they can slow down query performance and waste valuable storage space.

Fortunately, there are several ways to remove duplicate data in SQL.

In this article, we will explore some of the most effective methods for removing duplicate data in SQL, including using the DISTINCT keyword, the GROUP BY clause, and the INNER JOIN statement.



## How to Remove Duplicates in SQL Using DISTINCT Keyword

One of the easiest ways to remove duplicate data in SQL is by using the DISTINCT keyword. You can use the DISTINCT keyword in a SELECT statement to retrieve only unique values from a particular column.

Here's an example of how to use the DISTINCT keyword to remove duplicates from a table:

```
SELECT DISTINCT column_name
FROM table_name;
```

For example, if we have a table called "customers" with columns "customer_id" and "customer_name", we can use the following SQL query to remove duplicates from the "customer_name" column:

```
SELECT DISTINCT customer_name
FROM customers;
```

## How to Remove Duplicates in SQL Using GROUP BY Clause

Another way to remove duplicates in SQL is by using the GROUP BY clause. The GROUP BY clause groups rows based on the values in a specific column and returns only one row for each unique value.

Here's an example of how to use the GROUP BY clause to remove duplicates from a table:

```
SELECT column_name
FROM table_name
GROUP BY column_name;
```

For example, if we have a table called "orders" with columns "order_id", "customer_id", and "order_date", we can use the following SQL query to remove duplicates from the "customer_id" column:

```
SELECT customer_id
FROM orders
GROUP BY customer_id;
```

## How to Remove Duplicates in SQL Using INNER JOIN Statement

Another way to remove duplicates in SQL is by using the INNER JOIN statement. The INNER JOIN statement combines rows from two or more tables based on a related column between them. By joining a table with itself, we can compare rows and remove duplicates.

Here's an example of how to use the INNER JOIN statement to remove duplicates from a table:

```
SELECT a.column_name
FROM table_name a
INNER JOIN table_name b ON
a.column_name = b.column_name
WHERE a.primary_key > b.primary_key;
```

For example, if we have a table called "employees" with columns "employee_id", "employee_name", and "department_id", we can use the following SQL query to remove duplicates from the "department_id" column:



```
SELECT a.department_id
FROM employees a
INNER JOIN employees b ON
a.department_id = b.department_id
WHERE a.employee_id >
b.employee_id;
```

## Conclusion

Removing duplicate data in SQL can help improve query performance and save storage space.

By using the DISTINCT keyword, the GROUP BY clause, and the INNER JOIN statement, we can remove duplicates from a table in SQL.

Remember to always make a backup of your data before modifying it to avoid any potential data loss.

# HELPFUL? REPOST

linkedin.com/in/ileonjose