## How to Remove Duplicate Rows from a SQL Server Table?

There are a number of ways in SQL Server to handle duplicate records in a table based on particular circumstances such as:

### Removing duplicates rows from a unique index SQL Server table

You can use the index to classify the duplicate data in unique index tables then delete the duplicate records. First, we need to [create a database](https://appuals.com/how-to-list-all-databases-and-tables-using-psql/) named “test\_database”, then create a table “[Employee](https://appuals.com/thousands-of-redditors-order-sennheiser-headphones-after-employees-only-coupon-leak/)” with a unique index by using the code given below.

USE master

GO

CREATE DATABASE test\_database

GO

USE [test\_database]

GO

CREATE TABLE Employee

(

[ID] INT NOT NULL IDENTITY(1,1),

[Dep\_ID] INT,

[Name] varchar(200),

[email] varchar (250) NULL,

[city] varchar(250) NULL,

[address] varchar(500) NULL

CONSTRAINT Primary\_Key\_ID PRIMARY KEY(ID)

)

The output will be as below.

Creating the table “Employee”

Now insert data into the table. We will insert duplicate rows also. The “Dep\_ID” 003,005 and 006 are duplicate rows with similar data in all fields except the identity column with a unique key index. Execute the code given below.

USE [test\_database]

GO

INSERT INTO Employee(Dep\_ID,Name,email,city,address) VALUES

(001, 'Aaaronboy Gutierrez', 'aronboy.gutierrez@gmail.com','HILLSBORO','5840 Ne Cornell Rd Hillsboro Or 97124'),

(002, 'Aabdi Maghsoudi', 'abdi\_maghsoudi@gmail.com','BRENTWOOD','987400 Nebraska Medical Center Omaha Ne 681987400'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(004, 'Aabish Mughal', 'abish\_mughal@gmail.com','OMAHA','2975 Crouse Lane Burlington Nc 272150000'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(007, 'Pilar Ackaerman', 'pilar.ackaerman@gmail.com','ATLANTA','5813 Eastern Ave Hyattsville Md 207822201');

SELECT \* FROM Employee

The output will be as follows.

Inserting data into the table named “Employee” and fetching data from the same table.

Now find the no of rows in the table by executing the following code. The count(\*) function will count no of rows.

SELECT Dep\_ID,Name,email,city,address, COUNT(\*) AS duplicate\_rows\_count FROM Employee

GROUP BY Dep\_ID,Name,email,city,address

The output will be as below. Row no (3, 4), (6, 7), (8, 9) highlighted in the red box are duplicate ones.

This figure highlights duplicate rows having row\_no greater than 1

Our task is to enforce uniqueness by removing duplicates for the duplicate columns. It is a little easier to remove duplicate values from the table with a unique index than to remove the rows from a table without it. Given below are two methods to achieve this. The first method gives you duplicate rows from the table using the “row\_number()” function, whereas the second method uses the “NOT IN” function. Thes two methods have their own cost which will be discussed later.

### Method 1: Selecting duplicate records using “ROW\_NUMBER ()” function

select \* from (SELECT

Dep\_ID,Name,email,city,address,

ROW\_NUMBER() OVER (

       PARTITION BY

           Dep\_ID,Name,email,city,address

       ORDER BY

           Dep\_ID,Name,email,city,address

        ) row\_no

       FROM test\_database.dbo.Employee) x

       where row\_no>1

### Method 2: Selecting duplicate records using “NOT IN ()” function

SELECT \* FROM test\_database.dbo.Employee

WHERE ID NOT IN (SELECT MAX(ID)

FROM test\_database.dbo.Employee

GROUP BY Dep\_ID,Name,email,city,address)

Execute the above code and you will see the following output. Both methods give the same result, but they have different costs.

Selecting duplicate rows from the table named “Employee” using method 1 and 2 respectively

Now we will delete the above selected duplicate rows using “CTE” by using the following code. The following code is selecting duplicate rows to be deleted using the “ROW\_NUMBER ()” function.

### Method 1: Deleting duplicate records using “ROW\_NUMBER ()” function

WITH cte\_delete AS (

SELECT

Dep\_ID,Name,email,city,address,

ROW\_NUMBER() OVER (

PARTITION BY

Dep\_ID,Name,email,city,address

ORDER BY

Dep\_ID,Name,email,city,address

) row\_no

FROM

test\_database.dbo.Employee

)

DELETE FROM cte\_delete WHERE row\_no > 1;

The output will be as below.

Deleting duplicate records from indexed table using “ROW\_NUMBER ()” function

### Method 2: Deleting duplicate records using “NOT IN ()” function

Now in order to test another method, we need to truncate the table which will remove all the rows from the table. Then insert command will add values to the table. Execute the following code now.

USE [test\_database]

GO

truncate table test\_database.dbo.Employee

INSERT INTO Employee(Dep\_ID,Name,email,city,address) VALUES

(001, 'Aaaronboy Gutierrez', 'aronboy.gutierrez@gmail.com','HILLSBORO','5840 Ne Cornell Rd Hillsboro Or 97124'),

(002, 'Aabdi Maghsoudi', 'abdi\_maghsoudi@gmail.com','BRENTWOOD','987400 Nebraska Medical Center Omaha Ne 681987400'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(004, 'Aabish Mughal', 'abish\_mughal@gmail.com','OMAHA','2975 Crouse Lane Burlington Nc 272150000'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(007, 'Pilar Ackaerman', 'pilar.ackaerman@gmail.com','ATLANTA','5813 Eastern Ave Hyattsville Md 207822201');

SELECT \* FROM Employee

The output will be as given below.

**PRO TIP:** **If the issue is with your computer or a laptop/notebook you should try using Restoro Repair which can scan the repositories and replace corrupt and missing files. This works in most cases, where the issue is originated due to a system corruption. You can download Restoro by**[**Clicking Here**](https://www.restoro.com/includes/route.php?tracking=WassabiRes&exec=run&keyword=middleofpost)

Inserting data into the table named “Employee” and fetching data from the same table.

Execute the code given below to delete all the duplicate rows from the table “Employee”.

Delete FROM test\_database.dbo.Employee

WHERE ID NOT IN (SELECT MAX(ID)

FROM test\_database.dbo.Employee

GROUP BY Dep\_ID,Name,email,city,address)

The output will be as follows.

Delete all the duplicate rows from the indexed table named “Employee

### Execution Plan and Query Cost for deleting duplicate rows from the indexed table:

Now we have to check which method will be cost-effective and taking fewer resources. Select the code and click on the execution plan. The following screen will appear showing all executing plans along with cost percentage.

We can see that method 1 “deleting duplicate records using “ROW\_NUMBER ()” function” have 33% cost and method 2 “deleting duplicate records using NOT IN () function” has 67% cost. So the method one is most cost-effective as compared to method two.

Method 1 has 33% cost and method 2 has a 67% cost, revealing that method 1 is more cost-effective.

### Removing duplicates from a SQL Server table without a unique index:

It’s a bit more difficult to remove duplicate rows or tables without a unique index. In this scenario, using a common table expression (CTE) and the ROW NUMBER() function helps us in removing the duplicate records. To remove duplicates from the table without a unique index we need to generate unique row identifiers.

Execute the following code in order to create the table without a unique index.

USE [test\_database]

GO

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[Employee\_with\_out\_index](

[Dep\_ID] [int] NULL,

[Name] [varchar](200) NULL,

[email] [varchar](250) NULL,

[city] [varchar](250) NULL,

[address] [varchar](500) NULL,

)

GO

The output will be as follows.

Creating the table named “Employee\_with\_out\_index” without a unique index

Now insert records into the created table named “Employee\_with\_out\_index” by executing the following code.

USE [test\_database]

GO

INSERT INTO Employee\_with\_out\_index(Dep\_ID,Name,email,city,address) VALUES

(001, 'Aaaronboy Gutierrez', 'aronboy.gutierrez@gmail.com','HILLSBORO','5840 Ne Cornell Rd Hillsboro Or 97124'),

(002, 'Aabdi Maghsoudi', 'abdi\_maghsoudi@gmail.com','BRENTWOOD','987400 Nebraska Medical Center Omaha Ne 681987400'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(004, 'Aabish Mughal', 'abish\_mughal@gmail.com','OMAHA','2975 Crouse Lane Burlington Nc 272150000'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(007, 'Pilar Ackaerman', 'pilar.ackaerman@gmail.com','ATLANTA','5813 Eastern Ave Hyattsville Md 207822201');

SELECT \* FROM Employee\_with\_out\_index

The output will be as follows.

Inserting data into the table with an out index named “Employee\_with\_out\_index”

### Method 1: Deleting duplicate rows from a table using the “ROW\_NUMBER ()” function and JOINS.

Execute the following code which is using ROW\_NUMBER () function and JOIN to remove duplicate rows from the table without index. IT first creates a unique identity to assigns row\_no to all the rows and keep only one-row removing duplicate ones.

WITH temp\_tablr\_with\_row\_ids AS

(

SELECT ROW\_NUMBER() OVER (ORDER BY Dep\_ID,Name,email,city,address) AS row\_no,

Dep\_ID,Name,email,city,address

FROM test\_database.dbo.Employee\_with\_out\_index

)

DELETE a FROM temp\_tablr\_with\_row\_ids a

WHERE row\_no < (SELECT MAX(row\_no) FROM temp\_tablr\_with\_row\_ids i WHERE  a.Dep\_ID=i.Dep\_ID and

a.Name=i.Name and a.email=i.email and a.city=i.city and a.address=i.address

GROUP BY  Dep\_ID,Name,email,city,address)

The output will be as follows.

Deleting duplicate rows from a table without index using “ROW\_NUMBER ()” function and JOINS

### Method 2: Deleting duplicate rows from a table using the “ROW\_NUMBER ()” function and PARTITION BY.

Now, in this method, we are using ROW\_NUMBER function along with partition by clause in order to assign row\_no to all the rows and then delete duplicates ones. First of all, we need to truncate the same table we have created earlier so that all the data gets deleted from the table. Then, insert records into the table including the duplicates records. The third query will delete duplicate rows from the table named “Employee\_with\_out\_index”.

truncate table Employee\_with\_out\_index

INSERT INTO Employee\_with\_out\_index(Dep\_ID,Name,email,city,address) VALUES

(001, 'Aaaronboy Gutierrez', 'aronboy.gutierrez@gmail.com','HILLSBORO','5840 Ne Cornell Rd Hillsboro Or 97124'),

(002, 'Aabdi Maghsoudi', 'abdi\_maghsoudi@gmail.com','BRENTWOOD','987400 Nebraska Medical Center Omaha Ne 681987400'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(003, 'Aabharana, Sahni', 'abharana.sahni@gmail.com','HYATTSVILLE','2 Barlo Circle Suite A Dillsburg Pa 170191'),

(004, 'Aabish Mughal', 'abish\_mughal@gmail.com','OMAHA','2975 Crouse Lane Burlington Nc 272150000'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(005, 'Aabram Howell', 'aronboy.gutierrez@gmail.com','DILLSBURG','868 York Ave Atlanta Ga 303102750'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(006, 'Humbaerto Acevedo', 'humbaerto.acevedo@gmail.com','SAINT PAUL','895 E 7th St Saint Paul Mn 551063852'),

(007, 'Pilar Ackaerman', 'pilar.ackaerman@gmail.com','ATLANTA','5813 Eastern Ave Hyattsville Md 207822201');

Selecting duplicate records into the temp table

; WITH temp\_tablr\_with\_row\_ids AS

(

SELECT ROW\_NUMBER() OVER (PARTITION BY Dep\_ID,Name,email,city,address

ORDER BY  Dep\_ID,Name,email,city,address) AS row\_no, Dep\_ID,Name,email,city,address

FROM Employee\_with\_out\_index

)

Deleting duplicate records from the temp table

DELETE a FROM temp\_tablr\_with\_row\_ids a WHERE row\_no > 1

The output will be as follows.

Truncating, inserting, deleting duplicate rows from a table without index and selecting resulting records.

Furthermore, we need to know about query execution costs in order to understand which one is an optimized solution. So you need to select all the relevant queries and click on the execution plan. The image below shows the execution plan for the queries along with execution cost. Delete queries are highlighted in the red box. The first query which is using “ROW\_NUMBER ()” and JOIN clause has 56% execution cost, whereas the second query is using “ROW\_NUMBER ()” and “PARTITION BY” has 31% cost. So the second method is a more optimized one and we should follow an optimized solution.