# SQL Articles

## SQL query to delete duplicate rows

By Jagan Mohan (https://plus.google.com/102089965968283459910)

(http://www.facebook.com/sharer.php?u=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://twitter.com/share?text=SQL query to delete duplicate rows&url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (https://plus.google.com/share?url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://www.digg.com/submit?url=www.besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://www.linkedin.com/shareArticle?mini=true&url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://www.stumbleupon.com/submit?url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows&title= SQL query to delete duplicate rows ) (mailto:?Subject=SQL Article - SQL query to delete duplicate rows &Body=Good%20SQL%20article%20with%20examples%20for%20you!%20 http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows)

Create a table EmpDtl1 with some duplicate rows as shown below to understand different methods of delete duplicate rows.

```
create table EmpDup(empid int,name varchar(20))

insert into EmpDup values(1,'Andy')
insert into EmpDup values (1,'Andy')
insert into EmpDup values (2,'Chris')
insert into EmpDup values(3,'Bill')
insert into EmpDup values(3,'Bill')
insert into EmpDup values (3,'Bill')
```

| empid | name |
|---|---|
| 1 | Andy |
| 1 | Andy |
| 2 | Chris |
| 3 | Bill |
| 3 | Bill |
| 3 | Bill |

User Name "Andy" repeated 2 times and User Name "Bill" repeated 3 times. See above picture.

Following are the 3 different methods for deleting duplicate rows.

# Method 1:

Insert the distinct rows from the duplicate rows table to new temporary table. Then delete all the data from duplicate rows  table then insert all data from temporary table which has no duplicates as shown below.

```
select distinct * into #tmp From EmpDup
delete from EmpDup
insert into EmpDup
select * from #tmp drop table #tmp

select * from EmpDup
```

Output:

| empid | name |
|---|---|
| 1 | Andy |
| 2 | Chris |
| 3 | Bill |

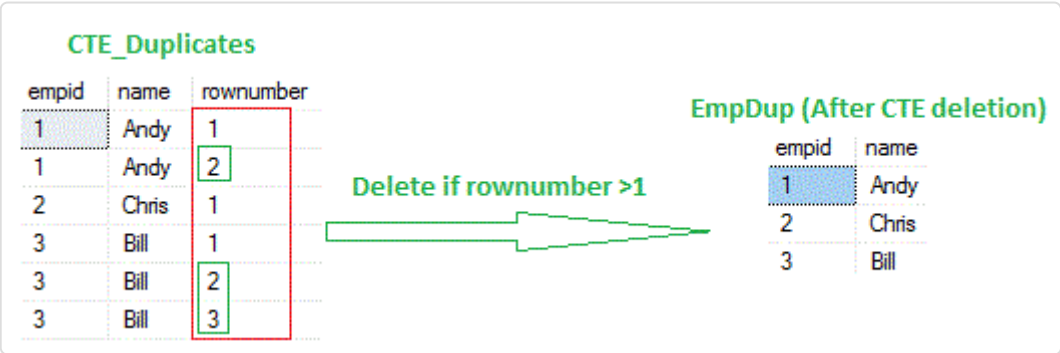
# Method 2:

If you want to consider only few columns in a table for duplication criteria to delete rows then Method1 doesn't work. In our example assume if EMDup has one more column "hobbies" extra apart from empid , name but you want to delete duplicate records if empid and name are repeated irrespective of "hobbies" data column, in this case Method1 will not work and follow **"Method2"**.

**Delete duplicate rows using Common Table Expression(CTE)**

```
With CTE_Duplicates as
(select empid,name , row_number() over(partition by empid,name order by empid,name ) rownumber
from EmpDup  )
delete from CTE_Duplicates where rownumber!=1
```

In above query, the function row_number generates a row number in each row for the same empid,name group of result set. See screenshot of CTE table CTE_Duplicates. If row number is greater than 1 means it is a duplicate row and to be deleted. The query "delete from CTE_Duplicates where rownumber!=1" will delete the duplicate rows from table EmpDup.




Understanding CTE query is bit difficult so if you still don't understand how CTE is working, no worries see **Method3** same job but different approach.

To understand how CTE works in depth , see my article on CTE here (http://www.besttechtools.com/articles/article/sql-common-table-expression)

# Method 3:

There are two methods here to delete duplicates they are using "group by" and "Rank()"

**Using "group by and min()":**

Add an identity column to the duplicate table as a serial number that acts as a row unique identifier(auto incremental ascending order). Left join EmpDup table with aliasing table "T" which has lowest "sno" from each duplicates sets. From the result of left join, if T.sno is null means it is a duplicate and to be deleted.

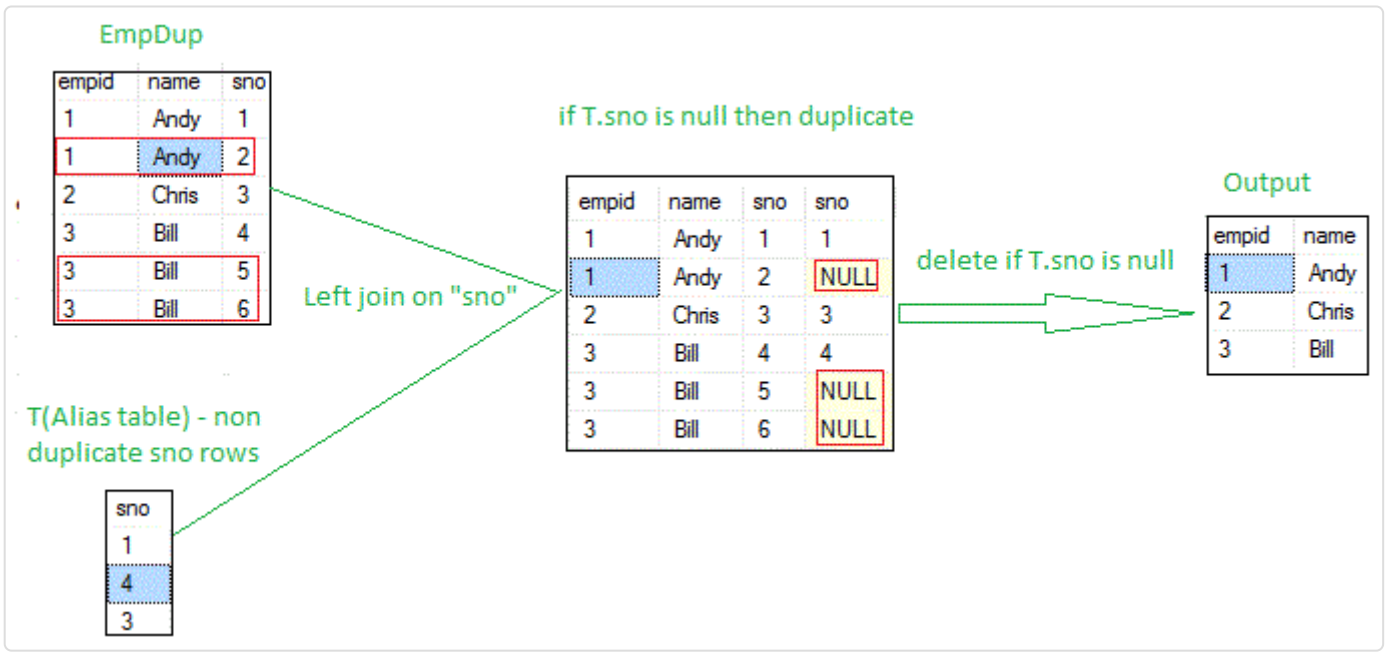
See below example with query and screenshot.

Note: If you have ID column in your duplicate table then no need to add serial number column then delete later as shown in our example.

```
alter table EmpDup add  sno int identity(1,1)

delete E from EmpDup E
left join
(select min(sno) sno From EmpDup group by empid,name ) T on E.sno=T.sno
where T.sno is null

alter table EmpDup
drop  column sno
```




**Using "Rank()":**

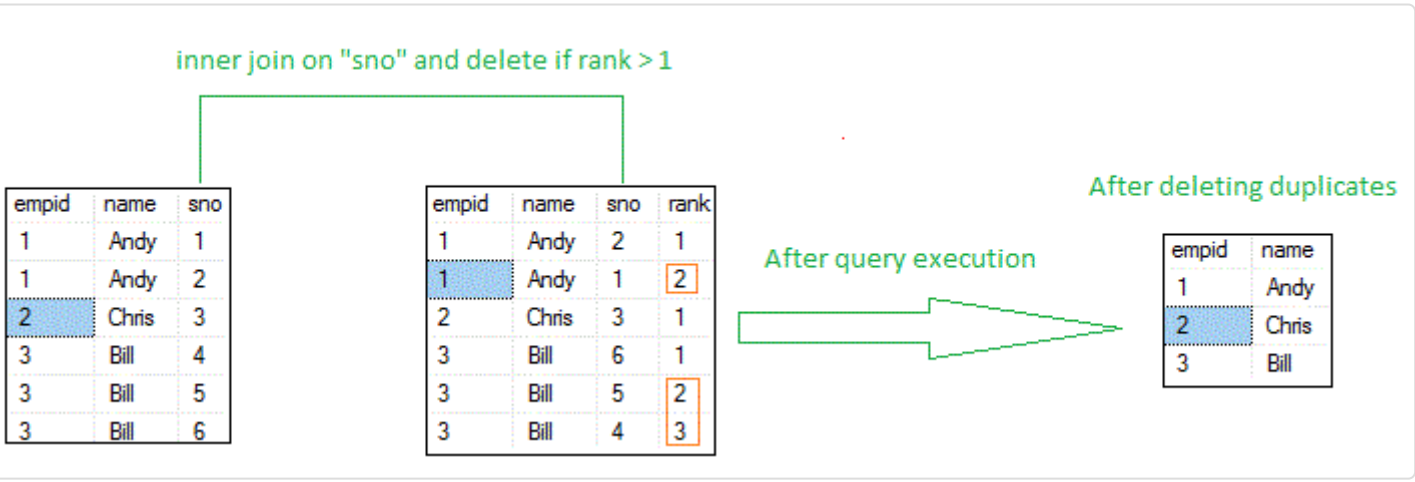
Add an identity column to the table as a serial number that acts as a row unique identifier(auto incremental ascending order).Then get the Rank against each empid,name based on serial number. If Rank is greater than 1 means it is a duplicate row and to be deleted. After deleting the duplicate rows, remove the identity column which is used for rank. See the below example.

Note: If you have ID column in your duplicate table then no need to add serial number column then delete later as shown in our example.

```
alter table EmpDup add  sno int identity(1,1)

delete E
from  EmpDup E
inner join
(select *,
RANK() OVER ( PARTITION BY empid,name ORDER BY sno DESC )rank
From EmpDup )T on E.sno=t.sno
where T.Rank>1

alter table EmpDup
drop  column sno
```




Click here (http://www.besttechtools.com/articles/article/sql-rank-functions) to know how RANK() works.

If you feel this query is little difficult to understand then use following query, same logic but in different way! that is instead of using "INNER JOIN" to delete duplicates(above query) use "IN" statement in query.

```
alter table EmpDup add  sno int identity(1,1)
delete from EmpDup where sno in
(
select sno from (
select *,
RANK() OVER ( PARTITION BY empid,name ORDER BY sno DESC )rank
From EmpDup
)T
where rank>1
)

alter table EmpDup
drop  column sno
```

Use any one of above 3 methods to delete duplicate rows

**Good to know about other scenarios of deleting duplicates:**
If you want to delete duplicates of a particular ID(here empid) then use"Delete Top( )" clause as shown below

```
delete top(2) From EmpDup where empid=2
```

OR

```
delete top(select count(*)-1 From EmpDup x where x.empid=2) From EmpDup where empid=2
```

If you want to delete all the rows if the selected columns repeated more than 1 time then use below query.

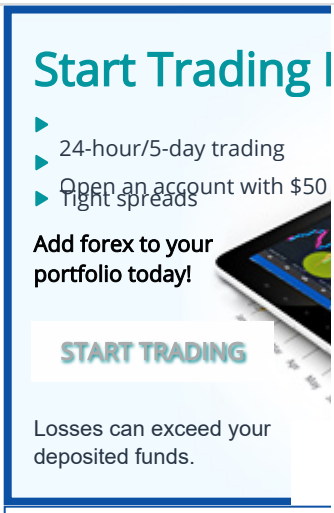
Query to delete 3 duplicated rows (in our example table) or repeated more than 1 time using "GROUP BY" clause.

```
 delete from EmpDup where EmpID in(select EmpID from EmpDup group by EmpId having
count(*) >1)
```

(http://www.facebook.com/sharer.php?u=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://twitter.com/share?text=SQL query to delete duplicate rows&url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (https://plus.google.com/share?url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://www.digg.com/submit?url=www.besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://www.linkedin.com/shareArticle?mini=true&url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows) (http://www.stumbleupon.com/submit?url=http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows&title= SQL query to delete duplicate rows ) (mailto:?Subject=SQL Article - SQL query to delete duplicate rows &Body=Good%20SQL%20article%20with%20examples%20for%20you!%20 http://besttechtools.com/articles/article/sql-query-to-delete-duplicate-rows)

Thay Ổ Cứng Laptop, SSD Laptop
Ổ Cứng Laptop, SSD Laptop Giá Rẻ. Hàng Chính Hãng- Miễn Phí Công Lắ
Thành Vinh Center

(http://www.kusalasolutions.com)



How did I lose 10 Kgs weight in 4 months naturally (/life/WeightLoss)

(/life/WeightLoss)

My Paintings (/life/paintings)

(/life/paintings)

All articles

SQL SERVER Ranking Functions - RANK, DENSE_RANK, NTILE, ROW_NUMBER (/articles/article/sql-rank-functions)

SQL query to delete duplicate rows (/articles/article/sql-query-to-delete-duplicate-rows)

SQL Server Isolation Levels with examples (/articles/article/sql-server-isolation-levels-by-example)

SQL query to update from join (/articles/article/sql-query-to-update-from-join)

SQL Server CTE(Common Table Expression) and Recursive Queries (/articles/article/sql-common-table-expression)

CTE Recursive query for data hierarchy(Parent Child hierarchy) (/articles/article/cte-recursive-query-for-data-hierarchy)

Different methods of SQL queries to insert data in tables (/articles/article/sql-different-methods-to-insert-data-into-tables)

SQL query to display all columns with datatypes for a given Table name (/articles/article/sql-query-to-display-all-columns-of-a-table)

SQL query to check two tables have identical data (/articles/article/sql-query-to-check-two-tables-have-identical-data)

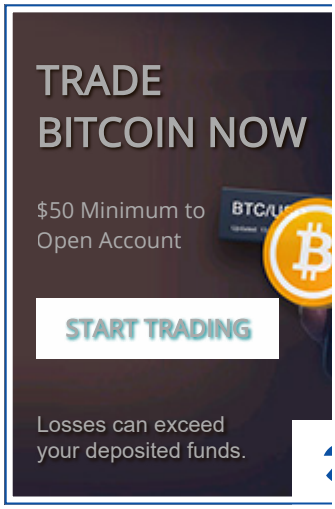
SQL query to search a string in database Schema (/articles/article/sql-query-to-search-a-string-in-database-Schema)

SQL query to display total number of rows for each table in database (/articles/article/sql-query-to-display-total-number-of-rows-for-each-table)

SQL query to search a column name in all tables (/articles/article/sql-query-to-search-column-name-in-database)

SQL server query to search object name in database (/articles/article/sql-server-query-to-search-object-name-in-database)

SQL Server performance tip - Do not use NOT IN clause in Subquery use LEFT OUTER JOIN instead (/articles/article/sql-performance-tip-Do-not-use-NOT-IN-clause)



(http://www.kusalasolutions.com)