**Resolving Duplicates in an Oracle Table**

To resolve duplicates one can conceptualize alogorithmically the problem as looping through a table from top down on primary key field or combination primary key checking each subsequent row for an equivalent value, and if

Alternatively, one could read in hash values of primary key before writing to a second table rejecting any repeat of hash value then not writing that row to second table until program reaches last row in source table.

However, Oracle SQL, PL/SQL, and Workspace Manager offer more specific applied solutions:

-----------------------------------------------------------------------------------------

Using SQL to remove duplicates in an Oracle table

------------------------------------------------------------------------------------------

https://asktom.oracle.com/pls/apex/asktom.search?tag=duplicate-rows-200101

How to find duplicate rows in a table ?

select c1, c2, c3, ..., cn, count(\*)

from T

group by c1, c2, c3, ...., cn

HAVING count(\*) > 1

/

will do it. c1, c2, .. cn represents the set of columns that should be "unique" in T.

\

http://www.dba-oracle.com/t\_delete\_duplicate\_table\_rows.htm

Use subquery to delete duplicate rows using SQL to delete duplicate table rows using an SQL subquery to identify duplicate rows, manually specifying the join columns:

DELETE FROM

table\_name A

WHERE

a.rowid >

ANY (

SELECT

B.rowid

FROM

table\_name B

WHERE

A.col1 = B.col1

AND

A.col2 = B.col2

);

Use RANK to delete duplicate rows

This is an example of the RANK function to identify and remove duplicate rows from Oracle tables, which deletes all duplicate rows while leaving the initial instance of the duplicate row:

delete from $table\_name where rowid in

(

select "rowid" from

(select "rowid", rank\_n from

(select rank() over (partition by $primary\_key order by rowid) rank\_n, rowid as "rowid"

from $table\_name

where $primary\_key in

(select $primary\_key from $table\_name

group by $all\_columns

having count(\*) > 1

)

)

)

where rank\_n > 1

)

Use self-join to delete duplicate rows: A most effective way to detect duplicate rows is to

join table against itself as shown below.

select

book\_unique\_id,

page\_seq\_nbr,

image\_key

from

page\_image a

where

rowid >

(select min(rowid) from page\_image b

where

b.key1 = a.key1

and

b.key2 = a.key2

and

b.key3 = a.key3

);

Must specify all of columns that make the row a duplicate in SQL where clause.

Once have detected duplicate rows, may modify SQL statement to remove duplicates

as shown below:

delete from

table\_name a

where

a.rowid >

any (select b.rowid

from

table\_name b

where

a.col1 = b.col1

and

a.col2 = b.col2

)

;

Use analytics functions to delete duplicate rows:

delete from

customer

where rowid in

(select rowid from

(select

rowid,

row\_number()

over

(partition by custnbr order by custnbr) dup

from customer)

where dup > 1);

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Removing duplicate table rows where rows have NULL values. Need to add a null check

because this fails to remove dupe rows where fields match on a null value.

delete from

table\_name a

where

a.rowid >

any (select b.rowid

from

table\_name b

where

a.col1 = b.col1

and

a.col2 = b.col2

)

;

Need to do following to remove all of duplicate table rows:

delete from

table\_name a

where

a.rowid >

any (select b.rowid

from

table\_name b

where

(a.col1 = b.col1 or (a.col1 is null and b.col1 is null))

and

(a.col2 = b.col2 or (a.col2 is null and b.col2 is null))

)

;

---------------------------------------------------------------------------------------------------------

Using PL/SQL to remove duplicate records in an Oracle table

---------------------------------------------------------------------------------------------------------

https://stackoverflow.com/questions/34603720/delete-duplicated-records-using-procedure-in-oracle-plsql

CREATE OR REPLACE PROCEDURE DELETE\_DUPLICATE AS

BEGIN

FOR I IN (SELECT TAB.A, TAB.B, MIN(ROWID) RID

FROM DUPLICATE\_TABLE TAB

GROUP BY TAB.A, TAB.B

HAVING COUNT(\*) > 1) LOOP

DELETE FROM DUPLICATE\_TABLE TAB

WHERE I.RID <> TAB.ROWID

AND TAB.A = I.A

AND TAB.B = I.B;

COMMIT;

END LOOP;

END;

-----------------------------------------------------------------------------------

create or replace procedure dedupe\_sometable

as

begin

delete sometable

where rowid in

( select lag(rowid) over (partition by id order by null)

from sometable );

end dedupe\_sometable;

CREATE OR REPLACE PROCEDURE Dup\_DELETE

AS

BEGIN

DELETE

FROM EMP

WHERE EMP.ROWID IN

-- Assuming that i am trying to segregate the duplicate values on Empno and ename

(SELECT A.ROWID

FROM

(SELECT ROW\_NUMBER() OVER(PARTITION BY EMPNO,ENAME ORDER BY JOB DESC) RNK,

empno,

ename,

rowid

FROM EMP

)A

WHERE A.RNK <> 1

);

END;

DECLARE

TYPE t\_del IS TABLE OF VARCHAR2(100);

l\_del t\_del;

CURSOR c IS

SELECT MIN(ROWID) RID

FROM test\_tbl TAB

GROUP BY TAB.age, TAB.gender

HAVING COUNT(\*) > 1;

BEGIN

OPEN c;

LOOP

FETCH c BULK COLLECT INTO l\_del;

EXIT WHEN l\_del.COUNT = 0;

FORALL i IN l\_del.FIRST..l\_del.last

DELETE FROM test\_tbl WHERE ROWID = l\_del(i);

END LOOP;

END;

-------------------------------------------------------------------------------------------------------------------

Using Workspace Manager to remove duplicates from an Oracle table

-------------------------------------------------------------------------------------------------------------------

https://docs.oracle.com/cd/B19306\_01/appdev.102/b14253/long\_intro.htm#i1009427

If commit bulk loading changes, Workspace Manager ensures that data is updated in required workspace and version. By default, the bulk-loaded data is checked for each unique or referential constraint defined on the table, and any bulk-loaded rows that are in violation of any constraints are moved to a discards table specified as a parameter to CommitBulkLoading procedure. If specified to check for duplicates (that is, records in data to be bulk loaded that have same values in primary key columns), for any duplicate records only record with lowest ROWID value is loaded into table, and rest are moved to discards table.

-------------------------------------------------------------------------------------------------------------------