

Duplicates

```
--delete duplicate data in SQL..
select * from employee
```

	emp_id	emp_name	salary	create_timestamp
1	1	Ankit	10000	2022-12-25 09:28:46.000000
2	2	Rahul	20000	2022-12-25 09:29:02.000000
3	3	Agam	30000	2022-12-25 09:29:47.000000
4	1	Ankit	15000	2022-12-25 09:43:32.000000

- 1.
2. Here we have duplicate of emp_id = 1
3. Also we can identify duplicates by using group by and count

```
--delete duplicate data in SQL..
select emp_id,count(1) from employee group by emp_id having count(1)>1
```

emp_id	count
1	2
2	1
3	1

4. One way of removing duplicates is to make use of timestamp columns...and if we want to keep latest timestamp id..then we use max(time stamp) ..or if we want to store the first entry then we use min(time stamp)

```
select emp_id,min(create_timestamp) as create_timestamp from employee group by emp_id
```

5. Now we can use delete command to delete the unwanted records

```
delete from employee where (emp_id,create_timestamp) in (select emp_id,min(create_timestamp) as create_timestamp from employee group by emp_id)
```

6. Also in realtime we'll always have the create timestamp column
7. What if we have exactly the same duplicates?

emp_id	emp_name	salary	create_timestamp
2	Rahul	20000	2022-12-25 09:29:02.000000
3	Agam	30000	2022-12-25 09:29:47.000000
1	Ankit	20000	2022-12-25 00:00:00.000000
1	Ankit	20000	2022-12-25 00:00:00.000000

8. Here to eliminate the duplicates in the output...we have used distinct keyword

```
insert into employee select distinct * from employee_back;
```

	emp_id	emp_name	salary	create_timestamp
1	3	Agam	30000	2022-12-25 09:29:47.000000
2	2	Rahul	20000	2022-12-25 09:29:02.000000
3	1	Ankit	20000	2022-12-25 00:00:00.000000

- 9.