<https://mp.weixin.qq.com/s/fF-aqKqFiNxQQPMOJW5zXg>

# 使用mysql数据库，遇到重复数据怎么处理？

## 前言

这是一个基本问题，这篇文章是我很早之前遇到的一种情况，后来在学习视频的时候又遇到了一次，因此给出一个总结。其实解决能否插入重复数据的问题，一般情况下是有两个思路，就像治水一样，第一个就是从源头，第二个就是在水流经的路上。我们带着这两种思路继续往下看：

## 问题

在我们的mysql数据库中，经常会出现一些重复的数据，有些情况我们允许重复数据的存在，但有时候我们也需要删除这些重复的数据。我们如何去处理呢？

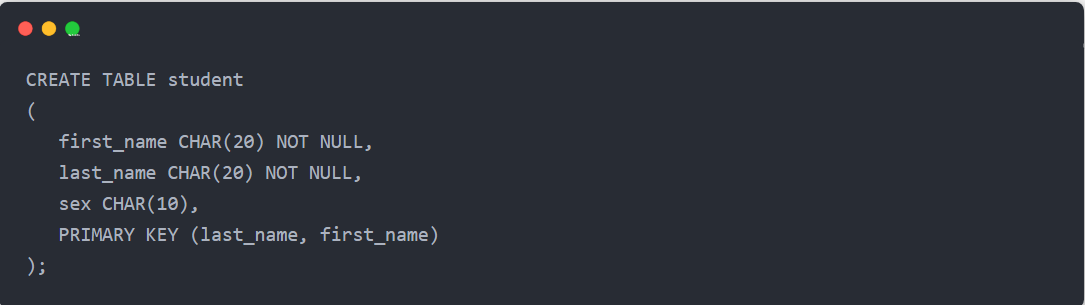
## 方法一：防止出现重复数据

也就是说我们再设计表的时候，就应该对这些数据设置一个UNIQUE 索引，在插入的时候就可以保证其唯一性，也就不存在有重复的数据了。当然你也可以直接设置为PRIMARY KEY（主键）。效果也是一样的。我们看一个案例：下表中无索引及主键，所以该表允许出现多条重复记录。

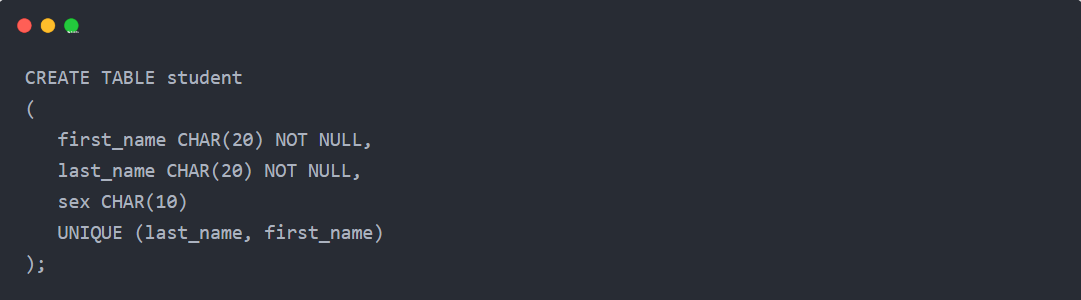


目前first\_name，last\_name是可以重复的，如果不想重复这里有两个解决办法：

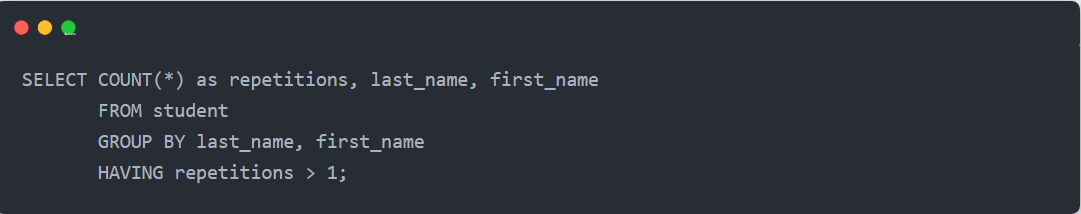
### 1、设置双主键模式



### 2、添加unique索引



这两种看起来形式好像有一点区别，但是能起到相同的作用。此时我们可以插入两条重复的数据，会发现报错。当然我们还可以在数据库中去验证一下：



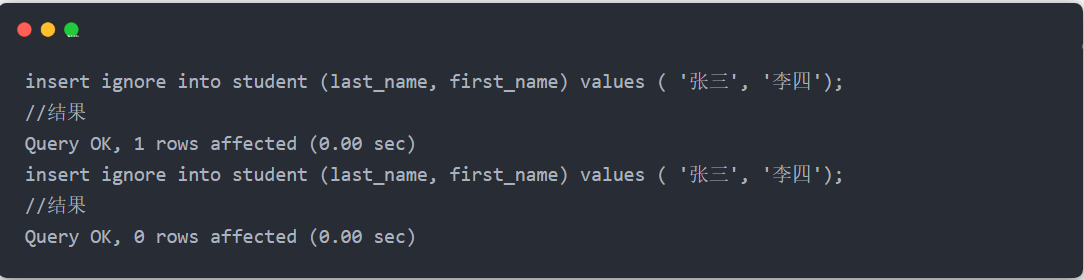
在这里我们统计的是 first\_name 和 last\_name的重复记录数，上面已经用两种方法设置了，这里肯定就是0了。

## 方法二：在插入时指定能否插入重复数据

在这里我们使用的是Insert ignore into 与Insert into指令。

（1）Insert ignore into会忽略数据库中已经存在的数据，如果数据库没有数据，就插入新的数据，如果有数据的话就跳过这条数据。这样就可以保留数据库中已经存在数据，达到在间隙中插入数据的目的。

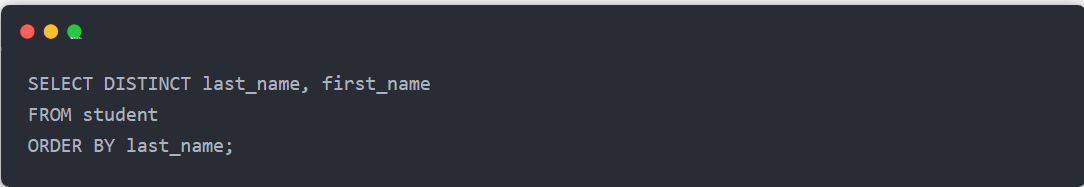
（2）Insert into则直接相反，会直接插入数据，不管数据库里面是否含有重复数据。我们还是举例说明：

现在我们看出来了吧，也就是说在执行第一条插入操作的时候，看到数据库没有，则直接插入一条新纪录，因此一行记录受到影响，但是在第二次插入的时候，数据库已经有一条一样的了，因此便不会插入了，0行受到影响。

当然了还有一个指令也可以完成类似于insert ignore into相似的功能，那就是replace  into。他表示的是如果存在primary 或 unique相同的记录，则先删除掉。再插入新记录。

## 方法三：过滤重复数据

如果你需要读取不重复的数据可以在 SELECT 语句中使用 DISTINCT 关键字来过滤重复数据。

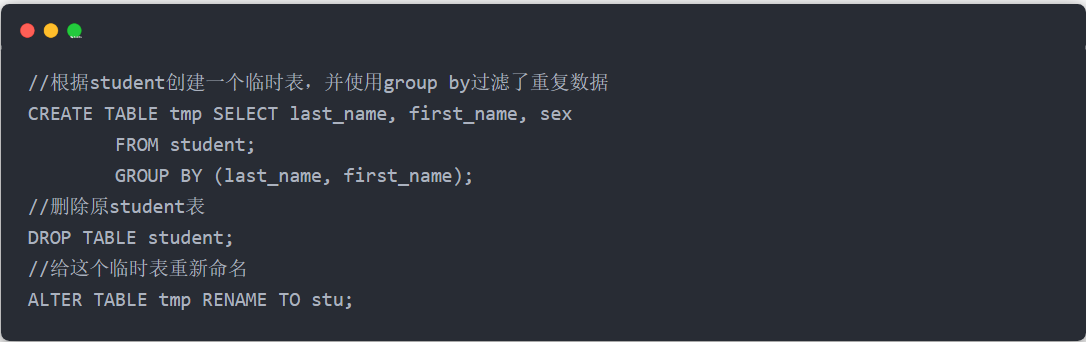


你也可以使用 GROUP BY 来读取数据表中不重复的数据：



## 方法四：删除重复数据

这种情况其实就相当于，在水的终点处去解决。看下面sql语句：



当然你也可以在数据表中添加 INDEX（索引） 和 PRIMAY KEY（主键）这种简单的方法来删除表中的重复记录。方法如下：

