

MySQL 批量插入：如何不插入重複數據？

Java後端 昨天

收錄於話題

#程序員 69 #MySQL 7 #Java 61

知識這個東西，看來真的要溫故而知新，一直不用，都要忘記了☹

業務很簡單：需要批量插入一些數據，數據來源可能是其他數據庫的表，也可能是一個外部excel的導入

那麼問題來了，是不是每次插入之前都要查一遍，看看重不重複，在代碼裡篩選一下數據，重複的就過濾掉呢？

向大數據數據庫中插入值時，還要判斷插入是否重複，然後插入。如何提高效率

看來這個問題不止我一個人苦惱過。

解決的辦法有很多種，不同的場景解決方案也不一樣，數據量很小的情況下，怎麼搞都行，但是數據量很大的時候，這就不是一個簡單的問題了。

幾百萬的數據，不可能查出來，做去重處理

說一下我Google到的解決方案☺

1、insert ignore into

當插入數據時，如出現錯誤時，如重複數據，將不返回錯誤，只以警告形式返回。所以使用ignore請確保語句本身沒有問題，否則也會被忽略掉。例如：

```
INSERT IGNORE INTO user (name) VALUES ('telami')
```

這種方法很簡便，但是有一種可能，就是插入不是因為重複數據報錯，而是因為其他原因報錯的，也同樣被忽略了～

2、on duplicate key update

當primary或者unique重複時，則執行update語句，如update後為無用語句，如id=id，則同1功能相同，但錯誤不會被忽略掉。

例如，為了實現name重複的數據插入不報錯，可使用一下語句：

```
INSERT INTO user (name) VALUES ('telami') ON duplicate KEY UPDATE id = id
```

這種方法有個前提條件，就是，需要插入的約束，需要是主鍵或者唯一約束（在你的業務中那個要作為唯一的判斷就將那個字段設置為唯一約束也就是unique key）。

3、insert ... select ... where not exist

根據select的條件判斷是否插入，可以不光通過primary 和unique來判斷，也可通過其它條件。例如：

```
INSERT INTO user (name) SELECT 'telami' FROM dual WHERE NOT EXISTS (SELECT id FROM user WHERE id = 1)
```

這種方法其實就是使用了mysql的一個臨時表的方式，但是裡面使用到了子查詢，效率也會有一點點影響，如果能使用上面的就不使用這個。

4、replace into

如果存在primary or unique相同的記錄，則先刪除掉。再插入新記錄。

```
REPLACE INTO user SELECT 1, 'telami' FROM books
```

這種方法就是不管原來有沒有相同的記錄，都會先刪除掉然後再插入。

實踐

選擇的是第二種方式

```
<insert id="batchSaveUser" parameterType="list">
    insert into user (id,username,mobile_number)
    values
    <foreach collection="list" item="item" index="index" separator=",">
        (
        #{item.id},
        #{item.username},
        #{item.mobileNumber}
        )
    </foreach>
ON duplicate KEY UPDATE id = id </insert>
```

這裡用的是Mybatis，批量插入的一個操作，mobile_number 已經加了唯一約束。這樣在批量插入時，如果存在手機號相同的話，是不會再插入了的。

來源：www.telami.cn

推薦閱讀

1. [GitHub上有什麼好玩的項目？](#)
2. [這個牛逼哄哄的數據庫開源了](#)
3. [SpringSecurity + JWT實現單點登錄](#)
4. [100道Linux常見面試題](#)



喜歡此內容的人還喜歡

MySQL詳細知識點總結可以收藏啦

軟件測試面試匯總



MySQL 億級數據分頁的優化

架構師



InnoDB 中一棵B+ 樹可以存多少行數據？

架構之美

