

MySQL 处理重复数据

有些 MySQL 数据表中可能存在重复的记录，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。本章节我们将为大家介绍如何防止数据表出现重复数据及如何删除数据表中的重复数据。

防止表中出现重复数据

你可以在MySQL数据表中设置指定的字段为 **PRIMARY KEY（主键）** 或者 **UNIQUE（唯一）** 索引来保证数据的唯一性。让我们尝试一个实例：下表中无索引及主键，所以该表允许出现多条重复记录。

```
CREATE TABLE person_tbl
(
    first_name CHAR(20),
    last_name CHAR(20),
    sex CHAR(10)
);
```

如果你想设置表中字段first_name，last_name数据不能重复，你可以设置双主键模式来设置数据的唯一性，如果你设置了双主键，那么那个键的默认值不能为NULL，可设置为NOT NULL。如下所示：

```
CREATE TABLE person_tbl
(
    first_name CHAR(20) NOT NULL,
    last_name CHAR(20) NOT NULL,
    sex CHAR(10),
    PRIMARY KEY (last_name, first_name)
);
```

如果我们设置了唯一索引，那么在插入重复数据时，SQL语句将无法执行成功并抛出错。

INSERT IGNORE INTO与INSERT INTO的区别就是INSERT IGNORE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。

以下实例使用了INSERT IGNORE INTO，执行后不会出错，也不会向数据表中插入重复数据：

```
mysql> INSERT IGNORE INTO person_tbl (last_name, first_name)
-> VALUES( 'Jay', 'Thomas');
Query OK, 1 row affected (0.00 sec)
mysql> INSERT IGNORE INTO person_tbl (last_name, first_name)
-> VALUES( 'Jay', 'Thomas');
Query OK, 0 rows affected (0.00 sec)
```

INSERT IGNORE INTO当插入数据时，在设置了记录的唯一性后，如果插入重复数据，将不返回错误，只以警告形式返回。而REPLACE INTO如果存在primary 或 unique相同的记录，则先删除掉。再插入新记录。

另一种设置数据的唯一性方法是添加一个UNIQUE索引，如下所示：

```
CREATE TABLE person_tbl
(
    first_name CHAR(20) NOT NULL,
    last_name CHAR(20) NOT NULL,
    sex CHAR(10),
    UNIQUE (last_name, first_name)
);
```

统计重复数据

以下我们将统计表中 first_name 和 last_name的重复记录数：

```
mysql> SELECT COUNT(*) as repetitions, last_name, first_name
-> FROM person_tbl
-> GROUP BY last_name, first_name
-> HAVING repetitions > 1;
```

以上查询语句将返回 person_tbl 表中重复的记录数。一般情况下，查询重复的值，请执行以下操作：

- 确定哪一列包含的值可能会重复。
- 在列选择列表使用COUNT(*)列出的那些列。
- 在GROUP BY子句中列出的列。
- HAVING子句设置重复数大于1。

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。

```
mysql> SELECT DISTINCT last_name, first_name
-> FROM person_tbl;
```

你也可以使用 GROUP BY 来读取数据表中不重复的数据：

```
mysql> SELECT last_name, first_name
-> FROM person_tbl
-> GROUP BY (last_name, first_name);
```

删除重复数据

如果你想删除数据表中的重复数据，你可以使用以下的SQL语句：

```
mysql> CREATE TABLE tmp SELECT last_name, first_name, sex FROM person_tbl GROUP BY (last_name, first_name, sex);  
mysql> DROP TABLE person_tbl;  
mysql> ALTER TABLE tmp RENAME TO person_tbl;
```

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

```
mysql> ALTER IGNORE TABLE person_tbl  
-> ADD PRIMARY KEY (last_name, first_name);
```

[← MySQL 序列使用](#)[MySQL 及 SQL 注入 →](#)

1 篇笔记

[写笔记](#)

```
select 列名1, count (1) as count  
from 表名  
group by 列名1  
having count>1 and 其他条件  
  
select 列名1, 列名2, count (1) as count  
from 表名  
group by 列名1, 列名2  
having count>1 and 其他条件
```

原理：先按照要查询出现重复数据的列，进行分组查询。count > 1 代表出现 2 次或 2 次以上。

示例：

```
/*查询重复数据*/  
select serialnum,cdate,count(*) as count  
from m_8_customer_temp_20180820bak  
group by serialnum,cdate having count>1 and cdate>='2018-08-20 00:00:00';
```

宋新峰 2周前 (03-05)