Atitit 数据库 垃圾数据缺失重复数据产生原因与解决总结

目录

[1. 原因 1](#_Toc17367)

[1.1. 缺少数据约束校验 1](#_Toc21997)

[1.2. 表关系关联设计错误 1](#_Toc152)

[2. 约束种类 2](#_Toc26947)

[2.1. 分类 表级约束vs列级别约束 2](#_Toc4848)

[2.2. 主键 外键约束 2](#_Toc20946)

[2.3. 单列约束vs多列组合约束 2](#_Toc27808)

[3. Other 2](#_Toc13383)

[3.1.1. 表级约束，给多个字段联合约束 2](#_Toc31182)

[4. 解决法总结 3](#_Toc26039)

[4.1. 增加外键约束，防止失联的垃圾数据 3](#_Toc30389)

[4.2. -------数据落地约束 3](#_Toc26367)

[4.3. 增加主键，外键 3](#_Toc2055)

[4.4. 增加unique索引 3](#_Toc31786)

[4.5. Trigger约束 实现检查约束 3](#_Toc21309)

[4.6. Merge 融合语句（需要unique或union unique索引 3](#_Toc18437)

[5. ------查询方面排重 3](#_Toc32345)

[5.1. 关联limit1 3](#_Toc21829)

[5.2. distinct 3](#_Toc31479)

[5.3. 适当groupby 分组运算 配合聚合函数max等 3](#_Toc494)

[5.4. 放弃表join，使用select字段udf join模式 4](#_Toc24217)

[6. Other 4](#_Toc9942)

[6.1. 通过view视图纠正原表过滤查询 4](#_Toc23182)

# 原因

## 缺少数据约束校验

## 表关系关联设计错误

导致left join多条对应

# 约束种类

* 非空约束(not null)
* 唯一性约束(unique)
* 主键约束(primary key) PK
* 外键约束(foreign key) FK
* 检查约束(目前MySQL不支持、Oracle支持)

## 分类 表级约束vs列级别约束

多列一般啊属于表级约束

**外键约**束(foreign key)FK只能是**表级定义**

## 主键 外键约束

## 单列约束vs多列组合约束

# Other

联合约束 可以使用unique实现

-> unique(name,email)

联合约束，表示两个或以上的字段同时与另一条记录相等，则报错

### **表级约束，给多个字段联合约束**

# 解决法总结

## 增加外键约束，防止失联的垃圾数据

## -------数据落地约束

## 增加主键，外键

## 增加unique索引

## Trigger约束 实现检查约束

建议trigger仅仅trig，主体检查放入sp udf 方便测试。。

## Merge 融合语句（需要unique或union unique索引

# ------查询方面排重

## 关联limit1

## distinct

## 适当groupby 分组运算 配合****聚合函数max等****

groupby count配合having

思路, group by 分组可以对多个列进行分组, 分组后可以过滤掉重复的数据

count数查询出来了,那么count>1的自然是重复的数据 配合having

SELECT id,`name`,age,count(1) as c

FROM user GROUP BY `name`,age having c > 1

## 放弃表join，使用select字段udf join模式

# Other

## 通过view视图纠正原表过滤查询