不停服切换数据库

目录

[1.1. 老师，这个热切开关具体什么方式实现呢？ 1](#_Toc29895)

[2. 双写 3](#_Toc31276)

[2.1. 利用数据同步工具Canal （推荐） 3](#_Toc16607)

[3. 迁移和校验脚本 3](#_Toc27346)

[4. 迁移用户数据一行行模式 3](#_Toc10945)

## 老师，这个热切开关具体什么方式实现呢？

作者回复: 这个一般都是通过编码来实现的。具体触发的方法，可以对外暴露一个可供调用的接口，或者通过动态配置下发等方式来触发。

有个问题，既然有比对和补偿程序，可不可以不使用数据实时同步。首先上线观察双写和补偿程序，没问题后先进行数据从旧到新的快照复制，然后开启双写，因有缝衔接而丢失的数据通过补偿程序来做……求教～

作者回复: 考虑到很难实现一个完美的对比补偿程序，还是建议不要这么做。

数据迁移的方案从业务层到数据库层各有不同的迁移方案, 我们先列举一些进行比对:

业务层: 在业务层进行硬编码, 数据双写, 以某个时间点进行划分, 新产生的数据同时写入新表, 运行一段时间后将旧数据迁移至新表. 成本极高, 与业务耦合严重, 不考虑.

连接层: 是方案1的进阶版, 在连接层拦截SQL进行双写, 与业务解耦, 但与1有着同样的一个问题: 周期较长, 要确保旧数据不会产生变更才能进行迁移.

触发器: 通过触发器将新产生的数据同步到新表, 本质上与2差不多.

数据库日志: 从某一时间点T备份数据库, 将备份库的数据迁移至新表, 从时间点T读取日志, 恢复到新表, 并持续写入. 待两份数据保持同步后, 上线新代码.

伪装从库: 相对于方案4的优势是不需要直接去读取日志, 解决了数据库在云上不方便直接读取日志的问题.

相比较之下, 方案4和5都是可选的, 因数据库在云上, 直接读取日志不方便, 且方案5有成熟的开源中间件\*\*[canal](https://github.com/alibaba/canal" \t "_blank)\*\*可用, 故笔者选择了方案5.

数据迁移方案

挂从库

在主库上建一个从库。从库数据同步完成后，将从库升级成主库（新库），再将流量切到新库。

这种方式适合数据结构不变，而且空闲时间段流量很低，允许停机迁移的场景。一般发生在平台迁移的场景，如从机房迁移到云平台，从一个云平台迁移到另一个云平台。大部分中小型互联网系统，空闲时段访问量很低。在空闲时段，几分钟的停机时间，对用户影响很小，业务方是可以接受的。所以我们可以采用停机迁移的方案。步骤如下：

1，新建从库（新数据库），数据开始从主库向从库同步。

2，数据同步完成后，找一个空闲时间段。为了保证主从数据库数据一致，需要先停掉服务，然后再把从库升级为主库。如果访问数据库用的是域名，直接解析域名到新数据库（从库升级成的主库），如果访问数据库用的是IP，将IP改成新数据库IP。

3，最后启动服务，整个迁移过程完成。

这种迁移方案的优势是迁移成本低，迁移周期短。缺点是，切换数据库过程需要停止服务。

# 双写

老库和新库同时写入，然后将老数据批量迁移到新库，最后流量切换到新库并关闭老库读写。

## 利用数据同步工具Canal （推荐）

我们可以看到上面双写的方案比较麻烦，很多数据库写入的地方都需要修改代码。有没有更好的方案呢？

我们还可以利用Canal，DataBus等工具做数据同步。以阿里开源的Canal为例。

# 迁移和校验脚本

准备迁移程序脚本，用于做老数据迁移。

准备校验程序脚本，用于校验新库和老库的数据是否一致。

# 迁移用户数据一行行模式