

SSO存储迁移MySQL方案

现状

MongoDB

表名	数据量	使用场景	是否迁移	备注
service_control	8	未知, 代码未使用	N	
applicant	17	授权账号	Y	
permission	2669	权限	Y	
attribute_map	2	SAML登录返回给SP的attribute, 已废弃	N	
role_group	18	权限组	Y	
login_history	388368145	登录记录	Y	保留近三个月数据, 老数据归档
sms_code	140698	短信验证码	Y	
account_lockdown	4830	账户锁定记录	Y	
sp_provider	40	SAML SP metadata	Y	
user	4393906	账号表	Y	
operation	16329419	操作记录	Y	
resource	4595222	权限资源	Y	
email_template	4	邮件模板	Y	
role	24469	权限角色	Y	
access_control_unit	726596	访问控制单元	Y	
role_assignment	11451940	角色授权	Y	
removed_user	31615	已删除用户, 备份表	Y	
subsystem	94	子系统	Y	
access_control_list	18	访问控制组	Y	
config	228	配置表, 目前只有电话区号	Y	
password_reset	1165330	重置密码记录	Y	

访问来源

<http://mongos-kibana.prod.xiaohongshu.com/goto/c5af643ec71a8aec555367802dbdee2e>

根据MongoDB访问日志client_host查询，请求都来自account的服务

目标

1. MongoDB数据全量迁移到MySQL
2. 服务以及运维脚本的数据库读写操作全量切换到MySQL
3. 迁移过程中不停机，对用户无感，有问题可迅速降级或回滚；

迁移方案

整体方案

对于增量的数据，采用数据双写方式同步到MySQL。对于存量的数据，采用[RED-DTS](#)同步到MySQL。

整体流程如下：

1. 完成前期工作包括MySQL建表，读写MySQL代码开发，测试环境验证；
2. 双写上线（包括staging和prod环境）；
3. 开启DTS存量同步，同步过程中可能出现错误，需要人工介入修复；
4. 同步完成后，逐步将读流量切换到MySQL，验证MySQL业务功能是否正常；
5. 读流量全量迁移到MySQL并验证完毕后，停止写MongoDB，迁移完成；

上线阶段	数据库读写状态	流程	
MongoDB	MySQL		
一	读/写	写	双写上线，DTS同步存量数据
二	写	读/写	读流量迁移至MySQL
三		读/写	停写MongoDB

准备工作

1. MySQL建表：[SSO MySQL表定义](#)（字段和索引和MongoDB基本保持一致，部分字段可能微调）
2. 开发双写逻辑：
 - a. ORM框架采用 Mybatis，方便自定义SQL；使用[MyBatisGenerator](#)生成Mapper和Model代码；
 - b. 双写时需要考虑，当数据不存在的时候插入，存在的时候则更新修改，使用[MySQL INSERT ... ON DUPLICATE KEY UPDATE Statement](#)；

- c. 双写覆盖account服务中所有插入、更新、删除操作;
 - d. 读操作的相关代码也可以同时开发;
3. 配置DTS同步迁移逻辑

PS:

● sso-09上有一些运维脚本，里面的逻辑都是写入MongoDB。为了减少修改脚本的工作，要求在DTS全量同步完成前不允许操作这些脚本，等待DTS全量同步完成后，这些脚本写入的数据会通过DTS自动同步到MySQL。

测试验证

1. account 新增一个全局MongoDB读/写开关，用于控制是否请求 MongoDB。新增MongoDB请求日志（通过拦截器实现）。
2. 在测试环境做接口测试：
 - a. 第一轮测试，验证上线第一阶段，打开MongoDB读/写开关，测试读MongoDB和双写模式下是否有问题，包括业务功能和数据;
 - b. 第二轮测试，验证上线第二阶段，关闭MongoDB读开关，测试读MySQL和双写模式下是否正常，包括业务功能和数据;
 - c. 第三轮测试，验证上线第三阶段，关闭MongoDB读/写访问开关，测试单独MySQL读写的逻辑，包括业务功能和数据;
 - d. 每轮测试需要检查日志，确保没有MongoDB读或写的访问;

*部分场景可以考虑使用[QueryDiff](#)回放线上流量来对比MySQL和MongoDB数据是否一致。

迁移流程

1. 服务双写上线（staging和prod）；
2. 开启DTS存量同步，同步过程中可能出现错误，需要人工介入修复；
3. 同步完成后，逐步将读流量切换到MySQL，开始测试验证；
4. 读流量全量迁移到MySQL并验证完毕后，停止写MongoDB，迁移完成；

开发任务拆解&排期

序号	任务	计划完成时间	状态	负责人
1	MySQL表结构设计	2023-01-29		@高俊康 (实习) @周凡 (实习) @郑剑 (实习)
2	MySQL读写代码开发sit环境配置DTS	2023-02-03		@高俊康 (实习) @周凡 (实习) @郑剑 (实习)
3	测试案例评审	2023-02-03		@涵珊(钟心琳)@黄国秦
4	测试: DTS和服务	2023-02-11		@涵珊(钟心琳)@黄国秦
6	双写上线	2023-02-13		
7	开始DTS同步	2023-02-13		
8	读切MySQL	2023-02-16		
9	停写MongoDB	2023-02-21		

其他依赖方迁移

大数据: 数仓porch相关离线任务

QA团队: datafactory数据工厂porch账号相关功能、qa_sync_task定时任务, 每天定时把线上员工账号同步到sit环境

企效: rbac调接口失败查SSO数据库

TODO

1. DTS全量同步模式下, delete可能产生脏数据。(将delete记录起来, 全量同步完成后, 再回放delete?)

TODO

2. MySQL当前版本5.7 腾讯云MySQL, 确认一下对DTS和其他迁移工作的影响 @廖立标

公司目前都用5.7版本MySQL(已支持JSON), 腾讯云MySQL确认过没有问题可以使用, 现在配置是4C8G比较低, 年后回来扩容

3. login_history数据量大, 确定数据拆分(分表or分区), 定期归档的方案 @廖立标

- a. 公司有成熟的[MySQL数据归档方案](#)，可以设置按时间归档；
- b. 在迁移本方案中，DTS配置同步近三月数据到线上库，老数据迁到归档库。后续login_history保留近3个月数据，老数据归档。
4. 年后找业务方QA排期测试验证@镜悬(王志斌)
5. DTS怎么测试 @廖立标

DTS有sit 环境，现在sit环境配置测试后再上prod环境 [DTS SIT](#)

每个表的数据规模，使用场景，读写方式，迁移方式

整理读写mongodb的client ip列表

迁移过程中，原先通过sso-09脚本操作的流程怎么处理（依靠DTS同步到mysql）

python服务的读写怎么处理（根据saml迁移项目计划，python服务那个时候已下线）