

## 业务背景

私信：单发+群发

1. 查看联系人：单发联系人+群发联系人。
2. 查看私信：取两个用户的单发、群发私信索引聚合形成会话。
3. 删除私信：删除会话私信、删除全部私信、删除单条私信。
4. 发送私信：
  - 接收者在一定周期内只能收到发送者的n条私信。
  - 发送者除了发粉丝以外还可以发送订阅用户、按条件筛选一部分订阅用户、指定一批uid发送、发给完全没有关系的人（新浪新闻）等。
  - 私信箱里收到的私信不会随着关系的变化而改变（比如两人原来为关注关系，取消关注以后仍然可以看见私信）。

## 群发索引现状

1. 群发cell处理机主从各16台，每次每天发送数约1亿条。
2. 采用推模式，推送到接收方的收件箱，不存发件箱。
3. 每次群发包含1条内容，n条索引；内容永久存储；索引按天分表，7天滚动删除。mysql索引库8端口+ssd。
4. 在redis里保存活跃用户收到的最近100条群发索引，聚合时剔除7天前数据。
5. 读取：线上查询会话(包含群发和单发)qps约5000/s，访问历史数据的请求<100/s，其中：

新浪新闻群发后1小时内对新浪新闻的查询会话接口请求qps峰值约3000  
央视新闻和人民日报下发后，对应的接口查询qps峰值约100  
对应db8端口峰值总计约500

6. 写入：
  - 1. 判断用户是否在索引缓存中（活跃用户），存在则写入索引缓存，>100w/s。
  - 2. 每个cell有一个单独线程异步更新所有用户的索引db，批插，16个cell总计约20w索引/s。
7. 删除：新浪新闻群发后删除峰值qps约50

文本

## 需求

能查看7天之前的群发私信（保留7天之前的索引）。

性能不低于当前速度:

活跃用户写入速度>100w/s

非活跃用户写入速度>=20w/s

读取速度>1000/s

容量目前约为400亿/年，考虑后续增长，按1500亿/年估算

---

## 方案:

---

### 1. hbase+redis counter

#### hbase

rowkey: owneruid

两个column family:

第一个用于查询两人会话，qualifier结构: otheruid+dmid

第二个用于查询收件箱，qualifier结构: dmid+otheruid

索引redis保存最近接收100条，查看会话时如果超过缓存里的100条则查询hbase。

每条索引占用 $8+(16+4)*2$ 约50byte，总容量约1500亿\*50=7.5T/年，线上hbase配置的冗余因子为3，共计约22.5T/年

目前了解到的hbase一些指标：单机写入>4w/s，读取约800/s，单机容量约3T。

考虑容量，使用9台hbase regionserver+1台master，共10台机器，最大写入峰值约9\*4w，约36w/s，最大读取峰值约7200/s，可线性扩展。

#### redis counter

保存群发会话索引计数(用于在收件箱中翻页)

key:fromuid(8byte)+toid(8byte),value:4byte，按每人平均收到20个V用户的群发私信计算，约为280G，每台cell启动15G redis counter，cell剩余内存约30G。在更新db的线程中更新缓存。

---

### 2. mysql+mc

#### 索引表

```
uid | otheruid | dmid | time
```

按uid hash与时间两个维度分库分表，每年32库，每库按uid hash分512表。按每年1500亿估

算，平均每表约1000w数据。

## 二级索引表：

```
uid | otheruid | year | count
```

由于聚合需要取总索引数，增加二级索引库，按uid hash分库分表。共8库，每库512表。每年每人平均收到20人的群发私信，总计约120亿数据。

mysql16主+32从共48台机器。

索引redis保存最近接收100条，查看会话时如果超过缓存中的100条则查询db。

## MC：

增加mc会话计数缓存，key为uid\_otheruid,value为每会话每年计数，在更新db的线程中写入活跃用户计数缓存。

每个会话计数30byte，7天过期，每次群发时更新活跃用户数约占总量的1/4，7天群发35亿条，其中活跃用户约8亿条，占用空间24G。

---

## 总结：

倾向于采用hbase方案。

1. hbase方案使用机器数量少；
2. 群发数据量大，对存储扩展性要求高，mysql大数据量时扩展比hbase复杂。
3. 群发高写入、低读取的场景与hbase契合度较高。
4. 目前更改存储不需要洗数据，双写7天即可。

## 待完善：

1. hbase版本选择：
  - 0.94：未来运维想统一成0.96版本(无法在线升级、随机读性能低50%)。
  - 0.96：hbase client与api-commons依赖包冲突。