

SACC

卓越 5周年 变迁

SequeMedia
盛拓传媒

IT168
www.it168.com

ChinaUnix

ITPUB

2013中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2013

大数据下的IT架构变迁

内存作为统一存储实践

@尹伟铭

案例：微博系统

- 拉模式的 user_timeline

- 查询出关注的人的列表

```
SELECT id FROM following WHERE follower_id=$UID
```

- 根据关注的人查询出微博列表

```
SELECT * FROM weibo WHERE user_id IN $FOLLOWING_LIST  
ORDER BY created_at LIMIT 0, 20;
```

案例：数据更新

- 关注很多人
 - 每个被关注的人说话都会引起数据更新
 - 收听的队列总是无法稳定

案例：微博系统

- 推模式的 user_timeline

- 查询出关注的人的列表

```
SELECT id FROM following WHERE follower_id=$UID
```

- 根据关注的人更新收听微博列表

```
INSERT INTO liestening (user_id, follower_id,  
weibo_id, created_at) VALUE ($UID, $FOLLOWER_ID,  
$WEIBO_ID, $CREATED_AT)
```

案例：数据更新

- 大V
 - 每句话都会引起千万级的数据队列更新
 - 更新延时容易导致时序问题

利用缓存

- 如何缓存
 - 存储关注的人列表？
 - 存储 user_timeline ？

利用缓存

- 遇到数据更新
 - 数据频繁更新
 - 缓存命中率极低

缓存的特点

- 高速交换数据
- 只存储少部分数据
- 数据更新往往采取废弃重读方式
- 随机IO

案例：解决办法

- 存储业务结构在内存中
 - uid -> [weibo, weibo,]
- 数据更新
 - 寻找需要更新的用户列表
 - uid -> [\$NEW_WEIBO, weibo, weibo, ...]

存储的特点

- 长期的持续化存储
- 数据有具体的更新方式
- 存储结构化数据，如 **RDBMS**
- 连续IO

使用内存

- 速度对比

CPU	1
内存	10
硬盘	1,000

使用内存的优点

- 随机访问速度快
- 扩展方便
- 灵活的数据模型

使用内存的限制

- 空间小
 - 无法存储全部数据
- 宕机数据丢失
- 运维难度增加

使用内存作为存储

- 命中率
 - 非命中情况要构建完整的业务相关数据结构
- 数据填充
 - 避免雪崩

内存数据的更新

- 构建业务相关的数据结构
- 减少对象关系映射
- 在内存中直接进行数据操作

构建二级存储

- 哪些放入内存
 - 热数据监控
 - 区分缓存和存储
- 应用程序
 - 构建内存存储数据结构

可选的技术

- memcached
- redis

Thanks!

SequeMedia
盛拓传媒


www.it168.com

 ChinaUnix^{.net}

