

# 如何设计一个短链系统？

我平时经常看极客时间上的专栏，上面的每一个专栏 URL 地址都有一个短链与之对应。比如你使用下面两个链接打开的都是《MySQL 实战 45 讲》这门课程。

- 原始链接：<https://time.geekbang.org/column/intro/100020801>
- 短链：<http://gk.link/a/10q2l>

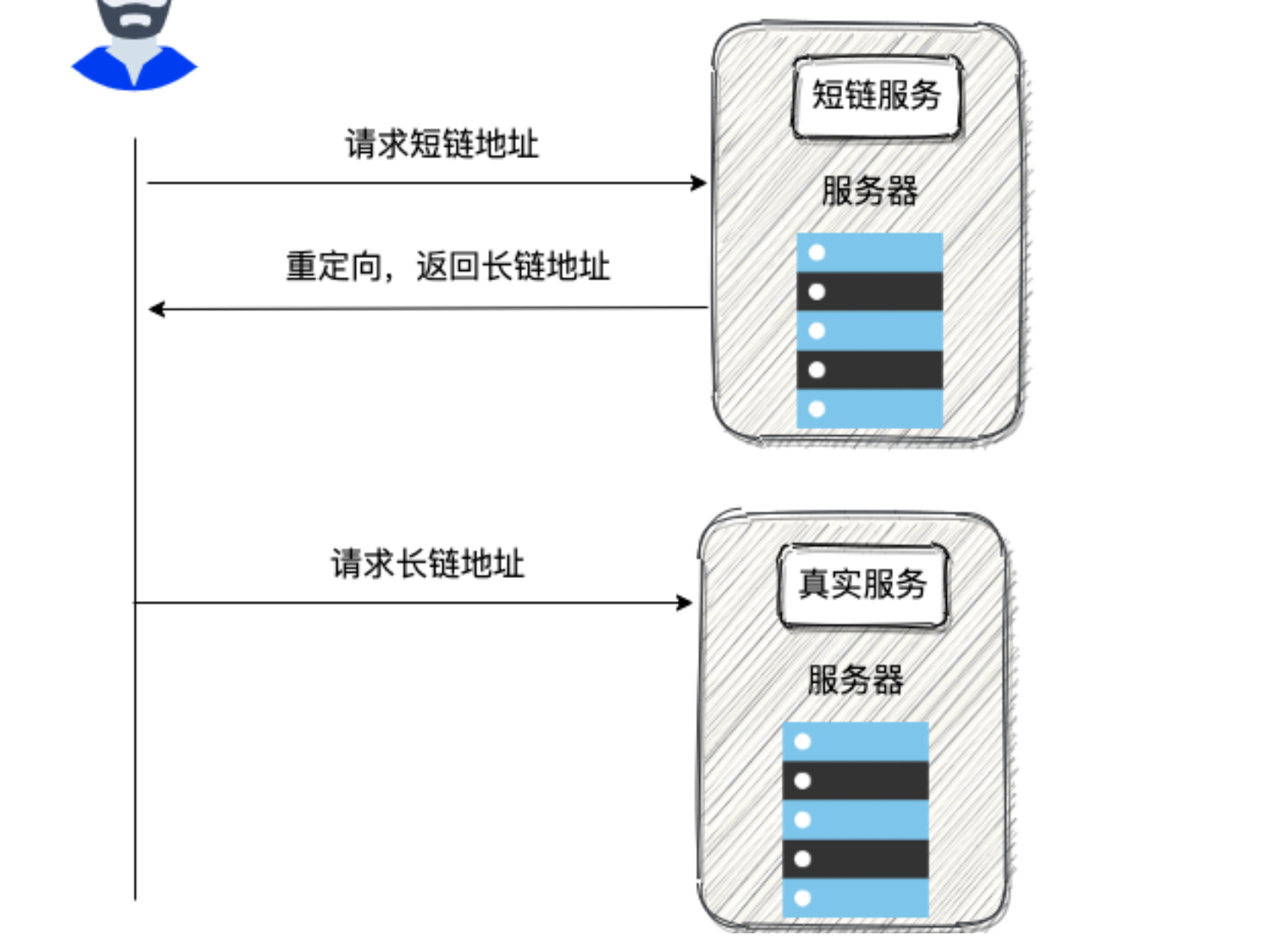
有了长链，为什么还要再弄一个短链呢？

1. 短链更简洁，更方便传播：过长的链接不利于在互联网传播；
2. 方便对链接的点击情况做后续追踪：比如查看短链最近一周的访问量、访客数、访问来源.....；
3. 对于短信等限制字数的场景来说更加友好：很多社交平台发表动态是有字数限制的，如果你直接使用长链的话，那留给你自己想表达的其他内容的文字就少了很多；
4. ....

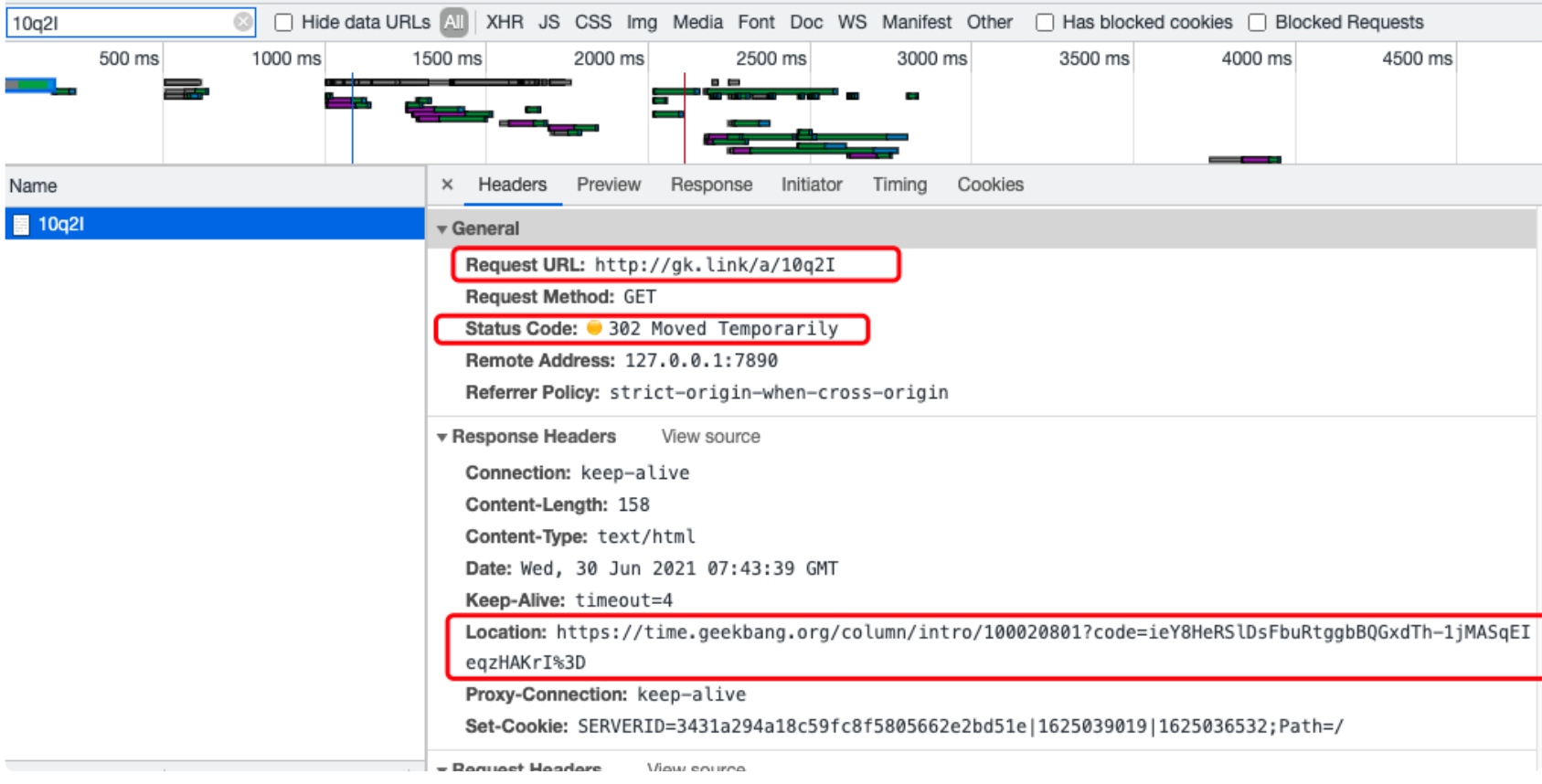
## 短链原理

短链的具体原理其实比较简单，说白了就是：**通过短链找到长链（原始链接），然后再重定向到长链地址即可！**

我画了一个简单的示意图：



🔴 举个例子：我们来访问“<http://gk.link/a/10q2l>”这个链接，从 HTTP 请求信息可以看到请求被重定向了，返回的状态码为“302”。



另外还有一个比较常用的重定向状态“301”，**我们应该用“301”还是“302”作为状态码更好呢？**

答案是：“302”，绝大部分短链系统也都是使用的“302”作为状态码。

这是因为“301”状态码代表永久重定向，只要浏览器拿到长链之后就会对其缓存，下次再请求短链就直接从缓存中拿对应的长链地址。这样的话，我们就没办法对短链进行相关分析了。

而“302”状态码代表资源被临时重定向了，不会存在上面说的这种问题。

🔴 举个例子：你的活动链接通过短链发送给了 10w+用户，你想知道短链后续的点击情况的话，你使用“301”状态码就不行了。

## 唯一短链生成

原始链接必定是唯一的，我们也要确保生成的短链唯一。

如何生成唯一的短链呢？换言之就是我们如何通过唯一的字符串来表示长链。

比较常见的一种方法就是：**通过哈希算法对长链去哈希。**

一般建议使用用非加密型哈希算法比如 `MurmurHash`。因为，相比于 `MD5`，`SHA` 等加密型哈希算法，非加密型哈希算法往往效率更高！

我们拿 `MurmurHash` 来说，`MurmurHash` 当前最新的版本是 `MurmurHash3`，它能够产生出 32-bit 或 128-bit 哈希值。对于绝大部分场景来说，32-bit 的一般就已经够用了。

生成的哈希值是 10 进制的，为了缩短它的长度，我们可以将其转变为 62 进制即可。10 进制的 3394174629 转换为 62 进制就是 3HHBS5。

我们将 3HHBS5 作为短链的唯一标识拼接即可。

既然使用了哈希算法，那不可避免会出现哈希冲突（不同的长链生成的短链是一样的），虽然概率比较小，但是我们也还是要解决。

**如何判断是否发生了哈希冲突呢？**

判断是否发生哈希冲突也就是看我们生成的短链是否是唯一的。

如果我们使用的是 MySQL，PostgreSQL 这类关系型数据库的话，我们可以给存放短链的字段 `sort_url` 添加唯一索引。

不过，为了提高性能以及应对高并发，还是建议利用布隆过滤器解决这个问题。

**如何解决哈希冲突呢？**

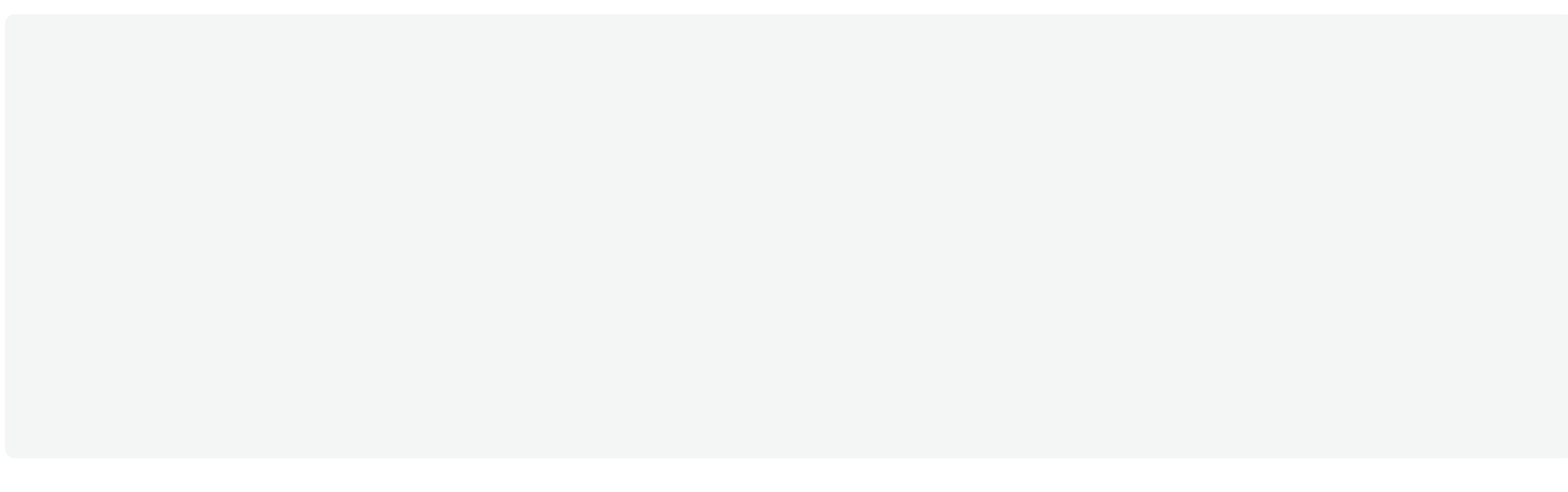
解决办法其实也很简单。如果发生哈希冲突，我们就在长链后拼接一个随机字符串。如果拼接了随机字符串还是发生哈希冲突那就再拼接一个随机字符串。

并且，我们要将拼接之后得到的字符串和拼接的字符串都存储起来，通过这两者可以获取长链（原始链接）。

**一个长链对应一个短链还是多个短链呢？**

这个还是要看具体的业务需求。个人建议是一个长链可以在不同的条件（比如生成短链的用户不同）下对应上不同的短链。这样的话，我们可以更好地对短链进行相关分析。

🔴 举个例子：通过小码短连接后台，我们可以看到短连接的访问次数、访问人数等信息。



这样的话，我们对长网址取哈希的时候加上对应的条件信息即可（比如生成短链的用户 ID）。

## 短链存储

如果我们使用 MySQL，PostgreSQL 这类关系型数据库存储的话，表结构大概是下面这样：

当然了，也可以使用 Redis 这类 K-V 内存数据库来做，这样性能也会更好！并且，存放在 Redis 中存放的本就是键值对的数据，刚好满足我们的需求。

当我们存放一个长链的时候，我们首先判断一下这个长链是否已经被转换成短链。

如果需要对长链就行区分的话（比如不同的用户使用同一个长链生成的短链不同），我们在判断的时候加上对应的条件即可（比如这个长链对应的用户）。

这里不能直接根据长链哈希之后得到的短链来判断长链是否已经被转换过短链，因为不同的长链生成的短链可能是一样的（哈希冲突，不过，概率很低）。

我个人建议不论是否使用 Redis 数据库，都要将最近比较活跃的短连接存放在缓存中。为了避免缓存过大，我们可以为这些放在缓存中的短连接设置一个过期时间。

👍

7 人点赞

🕒 维时

👤 SnailClimb

🕒 02-18 17:22

📄 引用原文：

个人建议是一个长链可以在不同的条件（比如生成短链的用户不同）下对应...

在互联网营销中，也会针对不同的营销渠道生成不同的短链，然后根据点击或者曝光来看渠道的营销效果

🕒 02-19 09:41

👤 SnailClimb

📄 引用原文：

一般一个完备的短链系统这些都是最基本的。很多短链系统还支持点击用户分析。

🕒 03-21 06:28

👤 许导

📄 引用原文：

锻炼

短链

👤 陈龙

🕒 10-15 15:35

📍 IP 属地广东

感觉像是一个中转站的功能，用户通过短链访问你的系统（其实就是调用你的接口），你在进行相关解密操作，返回长链出去并跳转

👤 注册

或

👤 登录

👤 语雀进行评论