

The logo for TinyURL, featuring the word "TINY" in dark blue and ">URL" in blue, with a black greater-than sign symbol.

硅谷之路37：如何设计TinyURL



Qin Elaine · 14 天前

查看完整视频：<http://www.bittiger.io/classes>

这是一道经典的系统设计面试题，是对SNAKE原则的深度应用，包含了系统设计的方方面面。最初的需求分析发现“长到短”和“短到长”两个基本接口，让我们又一次理解了“读与写是系统设计的基础”。根据日活跃一百万用户进行的QPS估算，让我们理解了什么是“用数字说话”。根据哈希算法，到简单有效的整数累加算法让我们理解了什么是“不要过度设计”。从数字编码，到字母编码，甚至表情编码，让我们看到了短链接的种种变体的来源。最后对存储的估算，让我们看到了貌似复杂的功能，居然占不了什么空间。

做个总结：

日活用户是基础，
插入查找算清楚。
字母编码省空间，
随机算法防冲突。



<http://bittiger.io/classpage/S2gBcdN7C4dbZSN8M>

goo.gl/FHJ2ZL

什么是Tiny URL？就是把长URL变短。这有什么用呢？主要是便于分享。比如微博twitter一共才一百多个字，一个URL就占一大半，还怎么玩儿呢。

我们用SNAKE原则来设计：

首先考虑应用场景（Scenario），应用场景有两个：将长的URL变短并存储；查找短的URL还原。抽象一下就是两个接口。

然后考虑限制性条件（Necessary），假设我们有一百万日活用户，分别计算插入和查找的每日请求次数。假设使用插入功能的日活用户有1%，一个用户插十次，平均到秒，每秒才1.2次，好低的，随便就满足了。一年我们会新生成36,500,000个短URL。但是查找就不同啦，100%要使用，不然怎么知道这个短URL是什么页面呢，算一下每秒35次。

Necessary for 1million daily users

- Daily active users
 - 1,000,000
- Insert
 - Per day: $1,000,000 * 1\%(\text{function usage}) * 10(\text{function frequency}) = 100,000$
 - Per year: $100,000 * 365 = 36,500,000$
 - Per second: $100,000 / 86400 = 1.2$
- Lookup
 - Per day: $1,000,000 * 100\%(\text{function usage}) * 3(\text{function frequency}) = 3,000,000$
 - Per second: $3,000,000 / 86400 = 35$

然后我们就可以考虑算法实现（Algorithm）。做法很简单，使用两个map，一个存储长URL到短URL的映射，一个存储短URL到长URL的映射。

longURL	shortURL
Bittiger.io	---
www.inoreader.com/all_articles	---
blog.csdn.net/longyulu/article/details/9159589	---
---	---

```

class Shortener{
    map<string,string> mLongToShort
    map<string,string> mShortToLong

    string Insert( string longURL )
    {
        if( mLongToShort.find( longURL ) == NULL )
            string shortURL = GenerateShortURL();
            mLongToShort[ longURL ] = shortURL;
            mShortToLong[ shortURL ] = longURL;
        Return mLongToShort[longURL ];
    }
};

```

shortURL	longURL
---	Bittiger.io
---	www.inoreader.com/all_articles
---	blog.csdn.net/longyulu/article/details/9159589
---	---

大家可能比较困惑GenerateShortURL()这个函数怎么实现，有很多种方法，一种就是对长URL哈希；可是还有一种更简单的方法，就是累加，只要返回map的大小就可以。

longURL	shortURL
Bittiger.io	0
www.inoreader.com/all_articles	1
http://blog.csdn.net/longyulu/article/details/9159589	2
---	---

```

string GenerateShortURL()
{
    Return string( mLongToShort.size() );
}

```

shortURL	longURL
0	Bittiger.io
1	www.inoreader.com/all_articles
2	http://blog.csdn.net/longyulu/article/details/9159589
---	---

我们已经计算出一年会生成36,500,000个短URL，如果我们用上面累加的方式来表示（只用数字编码），要8位数，加上前缀可能有点长，怎么变短呢？可以引入a到z和A到Z的字母（相当于62进制），这样长度就缩短到5位。

	Before	After
Yearly URL	36,500,000	36,500,000
Usable characters	[0-9] = 10	[0-9a-zA-Z] = 62
Encoding length	$\log_{10}(36,500,000) = 7.6 = 8$	$\log_{62}(36,500,000) = 4.2 = 5$
Example	goo.gl/36500000	goo.gl/2t9jG

```
string GenerateShortURL( string longURL )
    Return ConvertTo62( LongToShort.size() );

string ConvertTo62(int number)
    char Encode[62] = {'0',..., '9', 'a', ..., 'z', ..., 'A',..., 'Z'};
    string ret = "";
    While(number)
        ret = Encode[ number%62 ] + ret;
        number/=62;
    Return ret;
```

然后考虑数据存储（Kilobit），百万用户听起来很吓人，然而数据真的很多吗？来算一算：假设长URL的长度为100B，短URL使用我们上面的算法可以用int来表示，就是4B，假设我们还需要存储state（比如超时过期等）也用int型。这样每天有10.8M的数据，一年就是4GB。才4GB而已哎，两个map也才8GB，可以全放在内存里，并没有很吓人。

Average size of longURL = 100 bytes
Average size of shortURL = 4 bytes (int)
State = 4 byte
Daily new URL = 100,000 * 108 = 10.8MB
Yearly new URL = 10.8 * 365 = 4GB

URL table (MySQL or NoSQL)

longURL (varchar)	shortURL (int)	state (int)
Bittiger.io	0	0
www.inoreader.com/all_articles	1	0
http://blog.csdn.net/longyulu/article/details/9159589	2	1
...	...	0

以上我们就用SNAKE原则设计了一个Tiny URL的服务，棒不棒。

本文整理作者：Mengying Tian, [查看完整视频：http://www.bittiger.io/classes](http://www.bittiger.io/classes)

更多内容，请访问：BitTiger.io，扫描下面二维码，关注微信公众账号“论码农的自我修养”

weixin.qq.com/r/v0MnP17... (二维码自动识别)

程序员面试

系统设计

👍 140

📄 分享 🚩 举报



文章被以下专栏收录



论码农的自我修养

[进入专栏](#)

“有趣，有用，有效”。刷项目，做实战，捅破技术的...

3 条评论



写下你的评论



CLKBKB

我之前的做法是用md5，然后再截取部分作为tinyurl

13 天前



尹天仇 回复 **CLKBKB**

[查看对话](#)

如何做到防止重复呢

13 天前

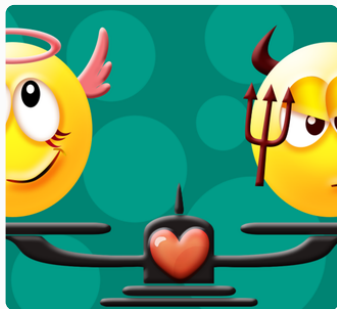


魏通

SNAKE原则最后一个字母E似乎没有体现？

12 天前

推荐阅读



一款有情绪的输入法——Kika Keyboard

“Kika Kika” 像是敲击键盘时的机械音效，也是Kika Keyboard手机键盘输入法公司名字的来源。不同于一般创业公司开扩国内市场或是打入近郊东... [查看全文 >](#)

Qin Elaine · 12 天前

发表于 论码农的自我修养



Java LinkedHashSet工作原理及实现

1. 概述Hash table and linked list implementation of the Set interface , with predictable iteration order. This implementation differs... [查看全文 >](#)

Qin Elaine · 15 天前

发表于 论码农的自我修养



关于51Talk的思考

1, 51talk计划在美国IPO的事在周四终于坐实了传闻。招股书最大的亮点莫过于公司名字 China Online Education Group。这名字太牛了 , 基本关闭... [查看全文 >](#)

潘欣 · 2 个月前 · 编辑精选

发表于 独立思考



暴雪游戏开发趣闻 (若干则)

这是 (Youtube) Blizzcon 2015 Engineering Community Amphitheater Discussion 的部分内容。挑了重点, 简单记录了一下。设计和工程风暴英雄... [查看全文 >](#)

顾露 · 6 天前 · 编辑精选

发表于 游戏人间