

转

一个网页打开的全过程

2017年12月28日 20:41:25

Aliley

阅读数: 12791

1、概要

从用户在浏览器输入域名开始，到web页面加载完毕，这是一个说复杂不复杂，说简单不简单过程，下文暂且把这个过程称作网页加载过程。下面我将依靠自己的经验把这个过程。如有错漏，欢迎指正。

阅读本文需要读者已有一定的计算机知识，了解TCP、DNS等。

2、分析

众所周知，打开一个网页的过程中，浏览器会因页面上的css/js/image等静态资源会多次发起连接请求，所以我们暂且把这个网页加载过程分成两部分：

1. html(jsp/php/aspx) 页面加载(假设存在简单的Nginx负载均衡)
2. css/js/image等 网页静态资源加载(假设使用CDN)

2.1 页面加载

先上一张图，直观明了地让大家了解下基本流程，然后再逐一分析。



2.1.1 DNS解析

什么是DNS解析？当用户输入一个网址并按下回车键的时候，浏览器得到了一个域名。而在实际通信过程中，我们需要的是一个IP地址。因此我们需要先把域名转换成这个IP地址，这个过程称作DNS解析。

1) 浏览器首先搜索浏览器自身缓存的DNS记录。

或许很多人不知道，浏览器自身也带有一层DNS缓存。Chrome 缓存1000条DNS解析结果，缓存时间大概在一分钟左右。

(Chrome浏览器通过输入：chrome://net-internals/#dns 打开DNS缓存页面)

2) 如果浏览器缓存中没有找到需要的记录或记录已经过期，则搜索hosts文件和操作系统缓存。

在Windows操作系统中，可以通过 ipconfig /displaydns 命令查看本机当前的缓存。

Windows系统的hosts文件在%systemroot%\system32\drivers\etc下，linux系统的hosts文件在/etc/hosts下。

3) 如果在hosts文件和操作系统缓存中没有找到需要的记录或记录已经过期，则向域名解析服务器发送解析请求。

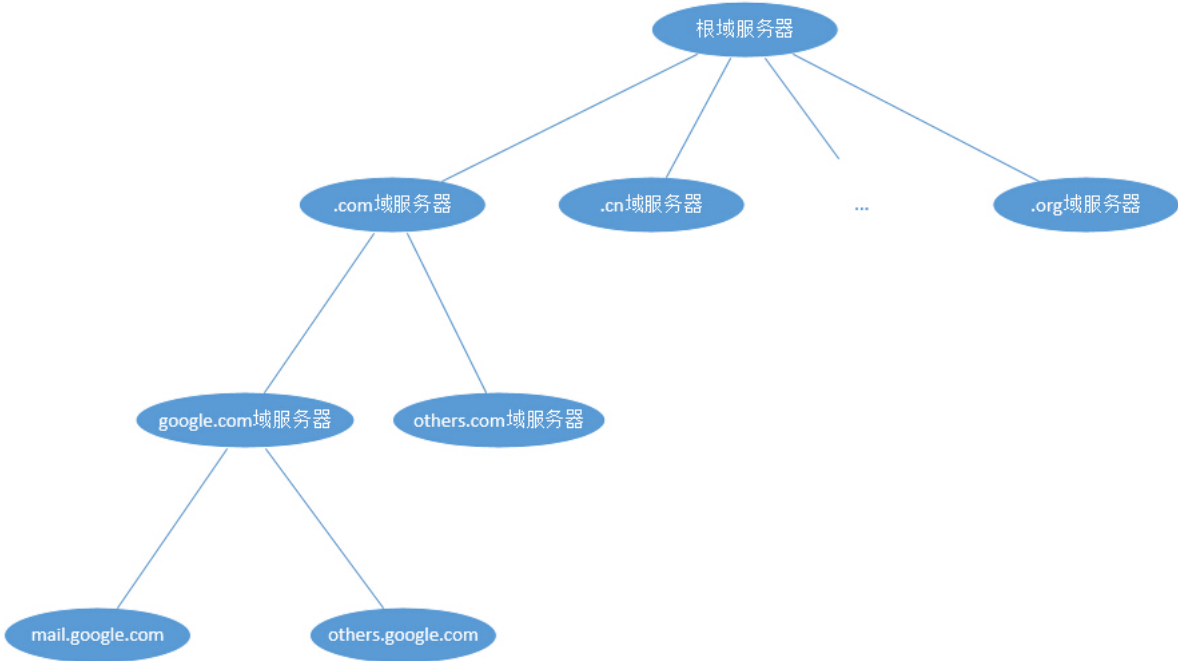
其实第一台被访问的域名解析服务器就是我们平时在设置中填写的DNS服务器一项，当操作系统缓存中也没有命中的时候，系统会向DNS服务器正式发出解析请求。这上开始解析一个未知的域名。

一般一台域名解析服务器会被地理位置临近的大量用户使用（特别是ISP的DNS），一般常见的网站域名解析都能在这里命中。

4) 如果域名解析服务器也没有该域名的记录，则开始递归+迭代解析。

这里我们举个例子，如果我们要解析的是mail.google.com。

首先我们的域名解析服务器会向根域服务器（全球只有13台）发出请求。显然，仅凭13台服务器不可能把全球所有IP都记录下来。所以根域服务器记录的是com域服务服务器的IP、org域服务器的IP.....。如果我们要查找.com结尾的域名，那么我们可以到com域服务器去进一步解析。所以其实这部分的域名解析过程是一个树形的搜索过程



根域服务器告诉我们com域服务器的IP。

接着我们的域名解析服务器会向com域服务器发出请求。根域服务器并没有mail.google.com的IP，但是却有google.com域服务器的IP。

接着我们的域名解析服务器会向google.com域服务器发出请求。...

如此重复，直到获得mail.google.com的IP地址。

为什么是递归：问题由一开始的本机要解析mail.google.com变成域名解析服务器要解析mail.google.com，这是递归。

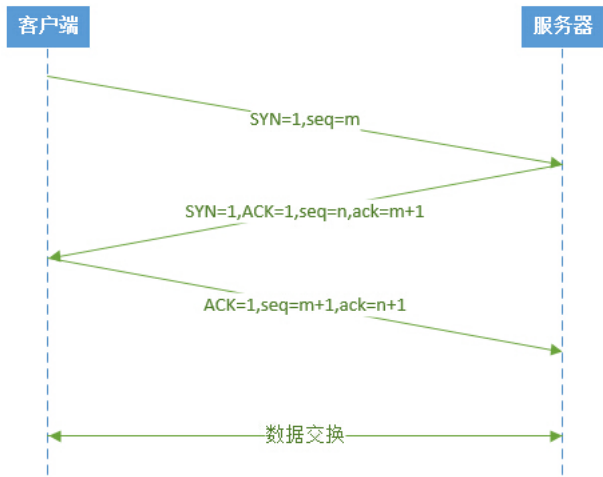
为什么是迭代：问题由向根域服务器发出请求变成向com域服务器发出请求再变成向google.com域发出请求，这是迭代。

5) 获取域名对应的IP后，一步步向上返回，直到返回给浏览器。

2.1.2 发起TCP请求

浏览器会选择一个大于1024的本机端口向目标IP地址的80端口发起TCP连接请求。经过标准的TCP握手流程，建立TCP连接。

关于TCP协议的细节，这里就不再阐述。这里只是简单地用一张图说明一下TCP的握手过程。如果不了解TCP，可以选择跳过此段，不影响本文其他部分的浏览。



2.1.3 发起HTTP请求

其本质是在建立起的TCP连接中，按照HTTP协议标准发送一个索要网页的请求。

2.1.4 负载均衡

什么是负载均衡？当一台服务器无法支持大量的用户访问时，将用户分摊到两个或多个服务器上的方法叫负载均衡。

什么是Nginx？Nginx是一款面向性能设计的HTTP服务器，相较于Apache、lighttpd具有占有内存少，稳定性高等优势。

负载均衡的方法很多，Nginx负载均衡、LVS-NAT、LVS-DR等。这里，我们以简单的Nginx负载均衡为例。关于负载均衡的多种方法详情大家可以Google一下。

Nginx有4种类型的模块：core、handlers、filters、load-balancers。

我们这里讨论其中的2种，分别是负责负载均衡的模块load-balancers和负责执行一系列过滤操作的filters模块。

1) 一般，如果我们的平台配备了负载均衡的话，前一步DNS解析获得的IP地址应该是我们Nginx负载均衡服务器的IP地址。所以，我们的浏览器将我们的网页请求发送到负载均衡服务器上。

2) Nginx根据我们设定的分配算法和规则，选择一台后端的真实Web服务器，与之建立TCP连接、并转发我们浏览器发出去的网页请求。

Nginx默认支持 RR轮转法 和 ip_hash法 这2种分配算法。

前者会从头到尾一个个轮询所有Web服务器，而后者则对源IP使用hash函数确定应该转发到哪个Web服务器上，也能保证同一个IP的请求能发送到同一个Web服务器上，实现会话保持（session stickiness）。

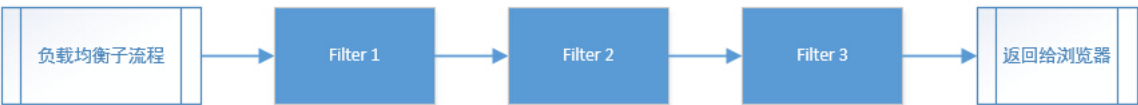
也有其他扩展分配算法，如：

fair：这种算法会选择相应时间最短的Web服务器

url_hash：这种算法会使得相同的url发送到同一个Web服务器

3) Web服务器收到请求，产生响应，并将网页发送给Nginx负载均衡服务器。

4) Nginx负载均衡服务器将网页传递给filters链处理，之后发回给我们的浏览器。



而Filter的功能可以理解成先把前一步生成的结果处理一遍，再返回给浏览器。比如可以将前面没有压缩的网页用gzip压缩后再返回给浏览器。

- 1) 浏览器根据页面内容，生成DOM Tree。根据CSS内容，生成CSS Rule Tree(规则树)。调用JS执行引擎执行JS代码。
- 2) 根据DOM Tree和CSS Rule Tree生成Render Tree(呈现树)
- 3) 根据Render Tree渲染网页

但是在浏览器解析页面内容的时候，会发现页面引用了其他未加载的image、css文件、js文件等静态内容，因此开始了第二部分。

2.2 网页静态资源加载

以阿里巴巴的淘宝网首页的logo为例，其url地址为

我们清楚地看到了url中有cdn字样。

什么是CDN？如果我在广州访问杭州的淘宝网，跨省的通信必然造成延迟。如果淘宝网能在广东建立一个服务器，静态资源我可以直接从就近的广东服务器获取，必然的打开速度，这就是CDN。CDN叫内容分发网络，是依靠部署在各地的边缘服务器，使用户就近获取所需内容，降低网络拥塞，提高用户访问响应速度。

接下来的流程就是浏览器根据url加载该url下的图片内容。本质上是浏览器重新开始第一部分的流程，所以这里不再重复阐述。区别只是负责均衡服务器后端的服务器器，而是提供静态资源的服务器。

10

2 条评论

收藏

分享

CSDN APP 程序员必备

目录

下一篇



想对作者说点什么



Hermione Granger：这个写的很棒（1周前 #2楼）



ka1iwang：写得很详细，感谢！（6个月前 #1楼）

- 访问一个网页的全过程

阅读量 2万+

引言打开浏览器，在地址栏输入URL，回车，出现网页内容。整个过程发生了什么？其中的原理是什么... 博文 来自： [u012862311...](#)
- 访问一个网页的全过程详解！

阅读量 327

访问一个网页的全过程详解！打开一个网页的过程中，浏览器会因页面上的css/js/image等静... 博文 来自： [魏波](#)
- 打开一个网站的详细过程

06-14

本文档详细的说明了，当我们在浏览器中输入网址后，打开网站的详细过程，其中对一些相关的概念做了详细说明，大家可以参考一下。 下载
- 浏览器打开一个网站可能经历那些步骤

阅读量 4804

此话题，可能被很多大神都写过很多次，但我依旧还要再写一次，因为这是我写的。当前社会下，相信... 博文 来自： [l821398100的...](#)
- 浅谈一个网页打开的全过程（涉及DNS、CDN、Nginx负载均衡等）

阅读量 8810

s 博文 来自： [king_a_123的...](#)
- 用户访问一个网站的过程描述

阅读量 2864

描述从浏览器打开http://www.baidu.com地址回车发送请求到看到页面的过程？打开浏览器输入网址... 博文 来自： [qq_26460507...](#)
- 当我们打开一个网页，浏览器做了什么？

阅读量 2358

当我们打开一个网页，浏览器是如何将一堆代码渲染成我们看到的页面的呢？这篇小博客简单的介绍了... 博文 来自： [博勋的博客](#)
- 当你打开一个网页时，都发生了什么事情？

阅读量 754

我们很多人每天都要通过网页来查询和获取自己所需要的信息，一般情况下，对于用户来说，从在地址... 博文 来自： [artadonis 的小...](#)
- 浏览器访问一个网站全过程

阅读量 516

网上关于这个话题的文章颇多，这里只是做个总结，理清思路！解析出主机名和端口号查询浏览器或者... 博文 来自： [zhangyonguu...](#)
- 电脑上访问一个网页的整个过程

阅读量 3175

电脑上访问一个网页，整个过程是怎么样的（url：www.sina.com）：其中DNS、HTTP、TCP、OSPF... 博文 来自： [匪鉴](#)
- 打开一个网页的时候，到底经历了哪些步骤？

阅读量 572

本文介绍从用户在浏览器输入域名开始，到web页面加载完毕的过程。一、页面加载（一般使用Nginx... 博文 来自： [sinat_2206577...](#)