

一文读懂 CDN 转载

逛逛GitHub 2021-08-16 14:09:14

©著作权

文章标签 ip地址 服务器 静态资源 dns服务器 域名解析 阅读数 8

前言

最近在了解边缘计算，发现我们经常听说的CDN也是边缘计算里的一部分。那么说到CDN，好像只知道它中文叫做内容分发网络。那么具体CDN的原理是什么？能够为用户在浏览网站时带来什么好处呢？解决这两个问题是本文的目的。

CDN概念

CDN全称叫做“Content Delivery Network”，中文叫内容分发网络。

实际上CDN这个概念是在1996年由美国麻省理工学院的一个研究小组为改善互联网的服务质量而提出的。那么它到底是怎么改善互联网服务质量的呢？

原理分析

我们知道，当我们使用域名访问某一个网站时，实际上就是将请求包（以Http请求为例）通过网络传输给某台服务器，比如访问“www.baidu.com”时：

1. 首先解析出该域名所对应的IP地址(DNS域名解析)
2. 然后将Http请求包通过网络路由到IP地址所对应的服务器

我们通常说“服务器的IP地址”，这其实不太准确，IP地址是和网卡绑定的，一个服务器可以有多个网卡，也就是可能有多IP地址。

我们先来看第一步：域名解析

域名解析

解析域名分为两种：

1. 将一个域名解析为一个IP地址
2. 将一个域名解析为另外一个域名

其实解析思路不难，我们在域名服务商购买了一个域名之后，需要去映射一个IP地址，可以用Map来表示这个关系：{域名: IP}。

同时我们也可以给某个域名取一个别名，比如“www.baidu.com”取一个别名“test.baidu.com”，这种关系也可以用Map来表示：{域名: 别名}。这里的别名专业一点叫做**CNAME**，相信大家对这个词有点眼熟，它就是这个意思。

而**域名解析**，实际上就是解析出指定域名所对应的IP地址，或者该域名的一个CNAME。

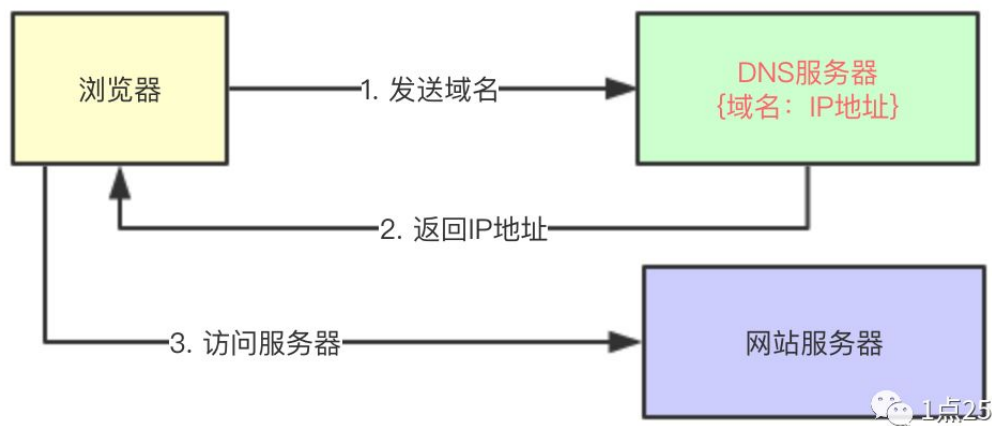
而域名解析是由DNS系统来负责的，DNS服务接受外部请求，从请求里提取域名，

- 如果这个域名对应的是IP地址，则返回这个IP地址，
- 如果这个域名对应的是CNAME，则继续查找CNAME域名的IP地址，然后将该地址返回给请求发送者。

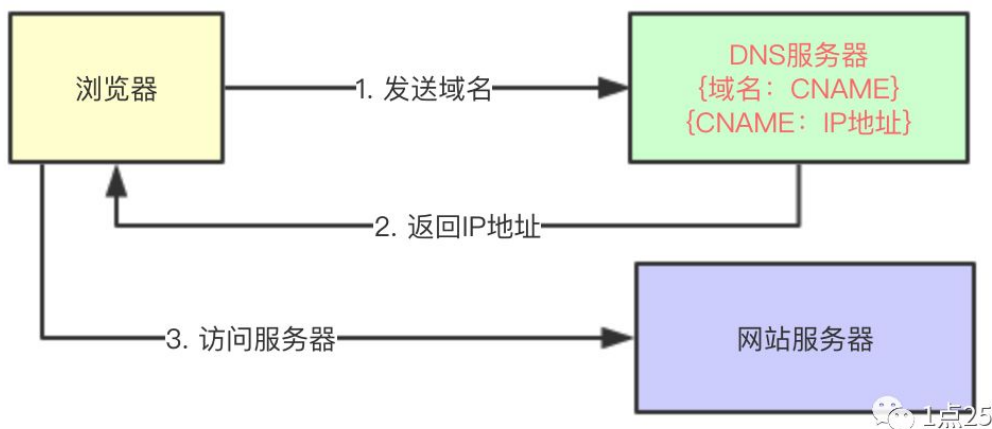
请求发送者拿到IP地址之后，完成真正的请求调用。

实际上DNS系统是非常庞大的，这里不去多讲，大家把它当作一个黑盒子，这个盒子的作用就是上文所描述的，这里用一个简单的图来表示一下。

没有CNAME的情况：



有CNAME的情况：



特别注意：在有CNAME的情况下，我们可以发现，CNAME实际上在域名解析的过程中承担了中间人（或者说代理）的角色，这是CDN实现的关键。

CDN原理

首先CDN是为了改善互联网的服务质量的。通俗一点说其实就是提高访问速度。
假设百度网站现在只有一台服务器，现在有一个人在上海访问百度，如果该服务器也在上海，那么通常来说访问比较快，如果该服务器在拉萨，那么相对而言访问就比较慢了。那么这个问题的根本原因是网络传输是依赖于网线的，网线越长，那么时间肯定就越久。

那么怎么解决这个问题呢？其实思路很简单，百度在全国各地都部署一模一样的服务器就行了，专业一点叫冗余。思路很简单，但实现还是比较麻烦的，服务器上的资源分为两种：**静态资源与动态资源**。

- 静态资源：这种资源通常是很少变动的，比如图片，视频，css，javascript等等
- 动态资源：这种资源不同用户不同时刻访问通常是不一样的，比如html，jsp等等。

那么如果百度要在全国各地都部署服务器，如果说每个服务器上都有相同的动态资源，那么可能还需要配置相应的数据库，因为动态资源所记录的信息通常会存储在数据库中，那么这就涉及到了数据同步等等问题，这会导致成本很高，这种做法专业一点其实就是集群，而目前来说集群架构最多是三地五中心，不是说全国多地集群不可能，主要是成本太高。

那么有没有成本比较低的方式呢，有，就是在每个服务器上只部署静态资源，静态资源通常不涉及到数据库，所以成本也比较低，而且也能提高用户的访问速度。

到这里，介绍了CDN想要达到的目的，那么怎么达到这个目的呢？

现在如果要比较CDN系统，我们可以考虑两点：

1. CDN系统中存储静态资源服务器的性能以及网速怎么样。
2. CDN系统中全国甚至全球范围内服务器节点的数量以及部署情况。

第一点很好理解，第二点大家应该也能理解了，如果静态资源的服务器节点很多，能够让每个用户在访问这些静态资源时都不用“跑很远的路程”才能获取到，那么自然这是CDN系统的优点。

有公司看到了这种需求，所以现在其实有很多CDN供应商，比如阿里，腾讯等等都有自己的CDN服务。只要你自己的系统接入了这些大厂所提供的CDN服务，你把自己的静态资源传给CDN服务，那么这些静态资源将自动的分布到全世界各地去。

好，那么现在的问题是，用户在访问静态资源时也是通过域名来访问的，域名会被解析成某一个IP地址，关键的问题就是，DNS系统怎么在做域名解析时，解析出来一个离用户最近的一个IP地址呢。

普通的DNS系统是做不到的，需要一个特殊的DNS服务器，这个特殊DNS需要知道

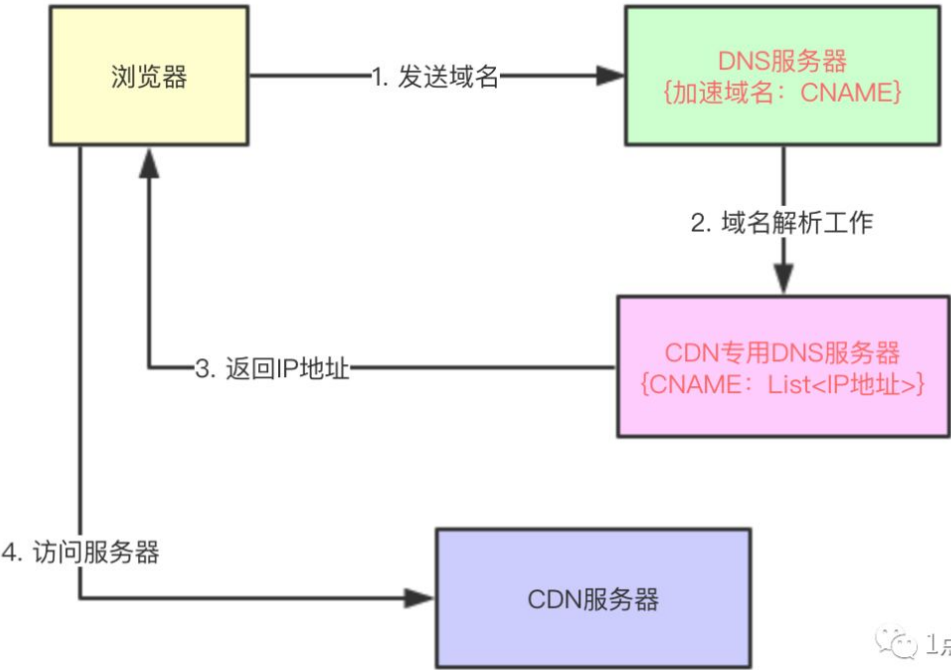
- 1. 用户当前位置
- 2. 还需要知道用户现在访问的这个域名对应哪些IP地址，以及这个IP地址分别在哪？

对于第一个问题好解决，直接从用户请求里提取出用户的ip地址，比如这个ip地址被解析为北京电信、上海移动等等。

第二个问题由谁来解决，我们现在考虑的是CDN，CDN提供商肯定知道他们公司在哪些地方部署了机器以及它们的IP地址，所以这个问题只能有CDN提供商来解决，CDN提供商会提供这个特殊的DNS服务器，我们叫做 CDN专用DNS服务器。

这样的话，只要用户在使用某个域名访问静态资源时，如果用户直接配置自己电脑的DNS地址为CDN专用DNS服务器。那么自然解决了问题，但是我们需要考虑的时，我们不能要求世界上所有的用户都去修改自己电脑的DNS地址。所以这个时候就要利用DNS中的CNAME了。

用户使用某个域名来访问静态资源时（这个域名在阿里CDN服务中叫做“加速域名”），比如这个域名为“image.baidu.com”，它对应一个CNAME，叫做“cdn.ali.com”，那么普通DNS服务器（区别CDN专用DNS服务器）在解析“image.baidu.com”时，会先解析成“cdn.ali.com”，普通DNS服务器发现该域名对应的也是一个DNS服务器，那么会将域名解析工作转交给该DNS服务器，该DNS服务器就是CDN专用DNS服务器。CDN专用DNS服务器对“cdn.ali.com”进行解析，然后依据服务器上记录的所有CDN服务器地址信息，选出一个离用户最近的一个CDN服务器地址，并返回给用户，用户即可访问离自己最近的一台CDN服务器了。



总结

通过上面的文章我们可以发现，CDN的实现原理依赖于DNS，因为本人不是专门搞网络的，所以文中如果有不准确的地方，还请各位大佬指出。