https://mp.weixin.qq.com/s/JxJC4m8vhoWRCVhVACz4aQ

**关于 DNS，你想聊的可能更多**

DNS（Domain Name System）， 也叫网域名称系统，是互联网的一项服务。它实质上是一个 域名 和 IP 相互映射的分布式数据库，有了它，我们就可以通过域名更方便的访问互联网。

DNS 有以下特点：

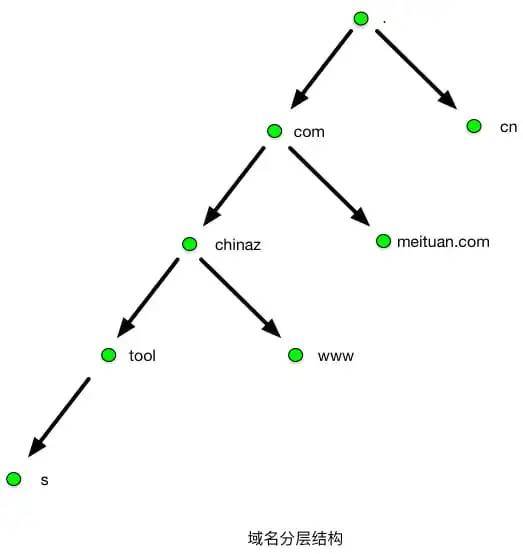
* 分布式的
* 协议支持 TCP 和 UDP，常用端口是 53
* 每一级域名的长度限制是 63
* 域名总长度限制是 253

**那么，什么情况下使用 TCP，什么情况下使用 UDP 呢?**

最早的时候，DNS 的 UDP 报文上限大小是 512 字节， 所以当某个 response 大小超过512 (返回信息太多)，DNS 服务就会使用 TCP 协议来传输。后来 DNS 协议扩展了自己的UDP 协议，DNS client 发出查询请求时，可以指定自己能接收超过512字节的 UDP 包， 这种情况下，DNS 还是会使用 UDP 协议。

### 分层的数据库结构

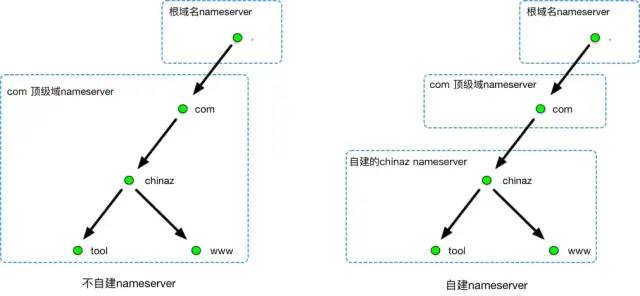
DNS 的结构跟 Linux 文件系统很相似，像一棵倒立的树。下面用站长之家的域名举例：



最上面的.是根域名，接着是顶级域名com，再下来是站长之家域名 chinaz 依次类推。使用域名时，从下而上。s.tool.chinaz.com. 就是一个完整的域名，www.chinaz.com. 也是。

之所以设计这样复杂的树形结构， 是为了防止名称冲突。这样一棵树结构，当然可以存储在一台机器上，但现实世界中完整的域名非常多，并且每天都在新增、删除大量的域名，存在一台机器上，对单机器的存储性能就是不小的挑战。另外，集中管理还有一个缺点就是管理不够灵活。可以想象一下，每次新增、删除域名都需要向中央数据库申请是多么麻烦。所以现实中的 DNS 都是分布式存储的。

根域名服务器只管理顶级域，同时把每个顶级域的管理委派给各个顶级域，所以当你想要申请com下的二级域名时，找 com 域名注册中心就好了。例如你申请了上图的 chinaz.com 二级域名，chinaz.com 再向下的域名就归你管理了。当你管理 chinaz.com 的子域名时，你可以搭建自己的 nameserver，在 .com 注册中心把 chinaz.com 的管理权委派给自己搭建的nameserver。自建nameserver 和不自建的结构图如下:



一般情况下，能不自建就不要自建，因为维护一个高可用的 DNS 也并非容易。据我所知，有两种情况需要搭建自己的 nameserver：

* 搭建对内的 DNS。公司内部机器众多，通过 IP 相互访问太过凌乱，这时可以搭建对内的 nameserver，允许内部服务器通过域名互通
* 公司对域名厂商提供的 nameserver 性能不满意。虽然顶级域名注册商都有自己的nameserver，但注册商提供的 nameserver 并不专业，在性能和稳定性上无法满足企业需求，这时就需要企业搭建自己的高性能 nameserver，比如增加智能解析功能，让不同地域的用户访问最近的 IP，以此来提高服务质量

概括一下 DNS 的分布式管理， 当把一个域委派给一个nameserver后，这个域下的管理权都交由此nameserver处理。这种设计一方面解决了存储压力，另一方面提高了域名管理的灵活性 (这种结构像极了Linux File System, 可以把任何一个子目录挂载到另一个磁盘，还可以把它下面的子目录继续挂载出去)

顶级域名

像 com 这样的顶级域名，由 ICANN 严格控制，是不允许随便创建的。顶级域名分两类:

* 通用顶级域名
* 国家顶级域名

通用顶级域名常见的如.com、.org、.edu等， 国家顶级域名如我国的.cn， 美国的.us。一般公司申请公网域名时，如果是跨国产品，应该选择通用顶级域名；如果没有跨国业务，看自己喜好（可以对比各家顶级域的服务、稳定性等再做选择）。这里说一下几个比较热的顶级域，完整的顶级域参见维基百科。

#### me

me顶级域其实是国家域名， 是黑山共和国的国家域名，只不过它对个人开发申请，所以很多个人博主就用它作为自己的博客域名（本博客也是这么来的~)

#### io

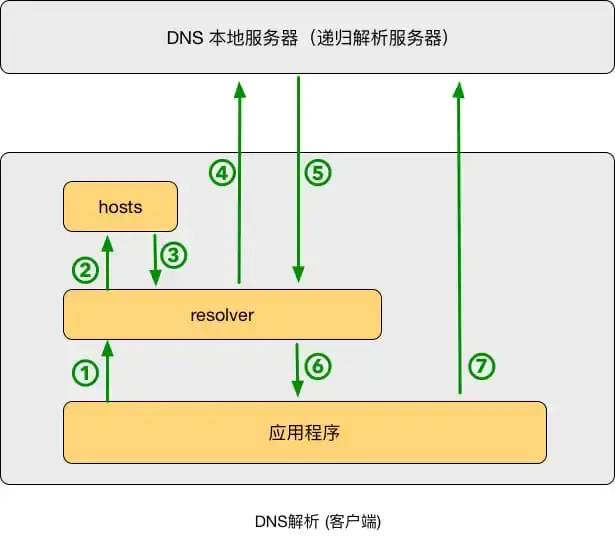
很多开源项目常用io做顶级域名，它也是国家域名。因为io 与计算机中的 input/output 缩写相同，和计算机的二机制10也很像，给人一种geek的感觉。相较于.com域名，.io下的资源很多，更多选择。

### DNS 解析流程

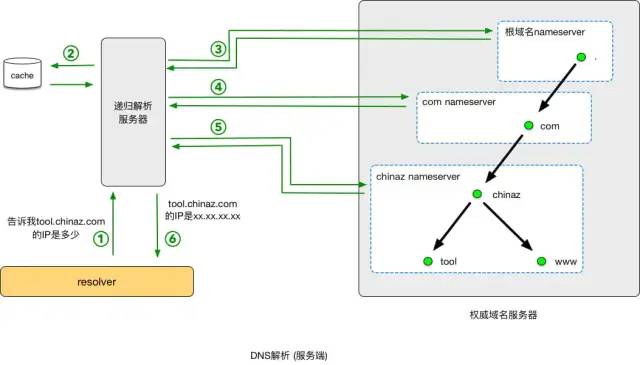
聊完了 DNS 的基本概念，我们再来聊一聊 DNS 的解析流程。当我们通过浏览器或者应用程序访问互联网时，都会先执行一遍 DNS 解析流程。标准 glibc 提供了 libresolv.so.2 动态库，我们的应用程序就是用它进行域名解析（也叫 resolving）的， 它还提供了一个配置文件/etc/nsswitch.conf 来控制 resolving 行为，配置文件中最关键的是这行：



它决定了 resolving 的顺序，默认是先查找 hosts 文件，如果没有匹配到，再进行 DNS 解析。默认的解析流程如下图：



上图主要描述了 client 端的解析流程，我们可以看到最主要的是第四步请求本地 DNS 服务器去执行 resolving，它会根据本地 DNS 服务器配置，发送解析请求到递归解析服务器（稍后介绍什么是递归解析服务器)， 本地 DNS 服务器在 /etc/resolv.conf 中配置。下面我们再来看看服务端的 resolving 流程：



我们分析一下解析流程：

* 客户端向本地DNS服务器(递归解析服务器) 发出解析tool.chinaz.com域名的请求
* 本地dns服务器查看缓存，是否有缓存过tool.chinaz.com域名，如果有直接返回给客户端；
* 如果没有执行下一步
* 本地dns服务器向根域名服务器发送请求，查询com顶级域的nameserver 地址
* 拿到com域名的IP后，再向com nameserver发送请求，获取chinaz域名的nameserver地址
* 继续请求 chinaz 的nameserver，获取 tool 域名的地址，最终得到了tool.chinaz.com 的 IP，本地 dns 服务器把这个结果缓存起来，以供下次查询快速返回
* 本地dns服务器把把结果返回给客户端

#### 递归解析服务器 vs 权威域名服务器

我们在解析流程中发现两类 DNS 服务器，客户端直接访问的是 递归解析服务器， 它在整个解析过程中也最忙。它的查询步骤是递归的，从根域名服务器开始，一直询问到目标域名。

递归解析服务器通过请求一级一级的权威域名服务器，获得下一目标的地址，直到找到目标域名的权威域名服务器

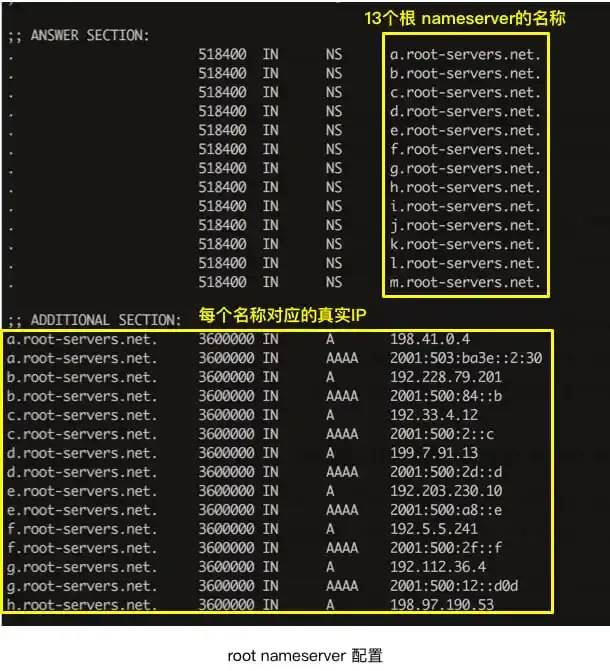
简单来说：递归解析服务器是负责解析域名的，权威域名服务器是负责存储域名记录的

递归解析服务器一般由 ISP 提供，除此之外也有一些比较出名的公共递归解析服务器， 如谷歌的 8.8.8.8，联通的 114，BAT 也都有推出公共递归解析服务器，但性能最好的应该还是你的ISP提供的，只是可能会有 DNS劫持的问题

#### 缓存

由于整个解析过程非常复杂，所以 DNS 通过缓存技术来实现服务的鲁棒性。当递归nameserver 解析过 tool.chianaz.com 域名后，再次收到 tool.chinaz.com 查询时，它不会再走一遍递归解析流程，而是把上一次解析结果的缓存直接返回。并且它是分级缓存的，也就是说，当下次收到的是 www.chinaz.com 的查询时， 由于这台递归解析服务器已经知道 chinaz.com 的权威 nameserver，所以它只需要再向 chinaz.com nameserver 发送一个查询 www 的请求就可以了。

根域名服务器递归解析服务器是怎么知道根域名服务器的地址的呢？根域名服务器的地址是固定的，目前全球有13个根域名解析服务器，这13条记录持久化在递归解析服务器中：



为什么只有 13 个根域名服务器呢，不是应该越多越好来做负载均衡吗？之前说过 DNS 协议使用了 UDP 查询， 由于 UDP 查询中能保证性能的最大长度是 512 字节，要让所有根域名服务器数据能包含在512字节的UDP包中， 根服务器只能限制在13个， 而且每个服务器要使用字母表中单字母名

#### 智能解析

智能解析，就是当一个域名对应多个 IP 时，当你查询这个域名的 IP，会返回离你最近的 IP。

由于国内不同运营商之间的带宽很低，所以电信用户访问联通的IP就是一个灾难，而智能 DNS 解析就能解决这个问题。

智能解析依赖 EDNS 协议，这是 google 起草的 DNS 扩展协议， 修改比较简单，就是在 DNS 包里面添加 origin client IP, 这样 nameserver 就能根据 client IP 返回距离 client 比较近的 server IP 了

国内最新支持 EDNS 的就是 DNSPod 了，DNSPod 是国内比较流行的域名解析厂商，很多公司会把域名利用DNSPod 加速， 它已经被鹅厂收购

### 域名注册商

一般我们要注册域名，都要需要找域名注册商，比如说我想注册 hello.com，那么我需要找com域名注册商注册hello域名。com的域名注册商不止一家， 这些域名注册商也是从ICANN 拿到的注册权， 参见如何申请成为.com域名注册商

那么，域名注册商 和 权威域名解析服务器  有什么关系呢？

域名注册商都会自建权威域名解析服务器，比如你在狗爹上申请一个.com下的二级域名，你并不需要搭建nameserver， 直接在godaddy控制中心里管理你的域名指向就可以了， 原因就是你新域名的权威域名服务器默认由域名注册商提供。当然你也可以更换，比如从godaddy申请的境外域名，把权威域名服务器改成DNSPod，一方面加快国内解析速度，另一方面还能享受DNSPod 提供的智能解析功能