

System Design

Part 2: DNS & CDN

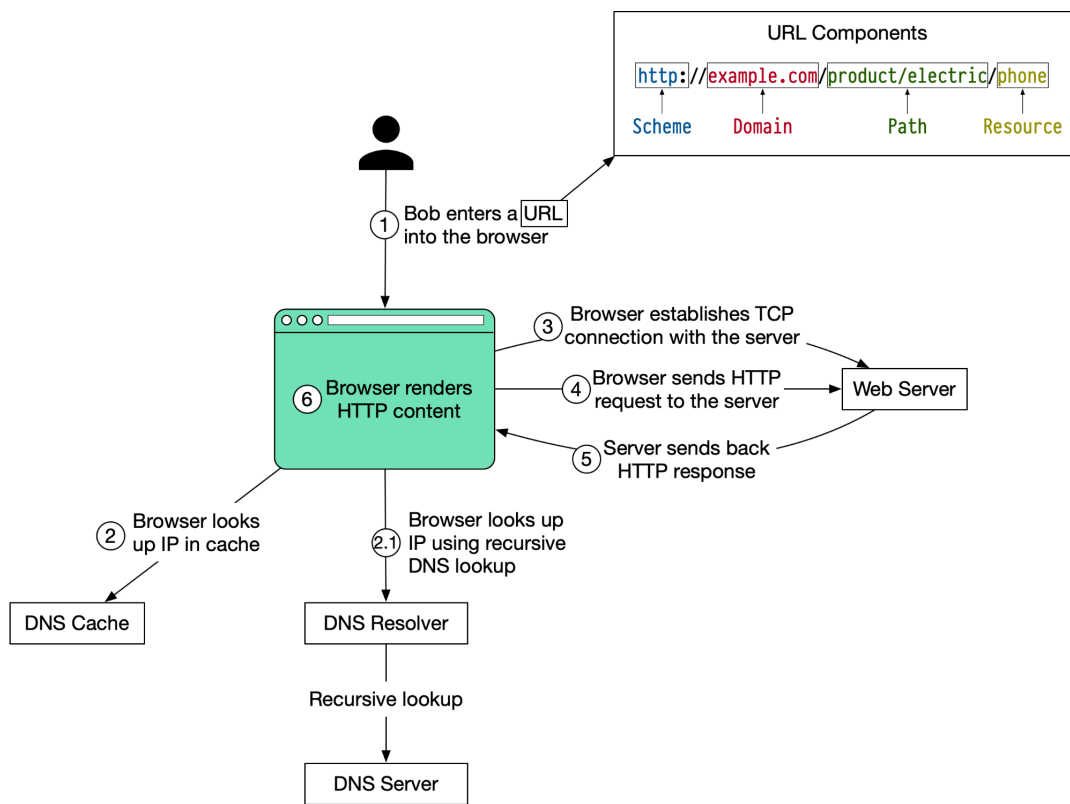
By Alex Xu



Some of you asked to get a PDF version of categorized posts. Here you go. You can subscribe to receive updates or new topics here: blog.bytebytego.com

- ♦ What happens when you type a URL into your browser
- ♦ How does the Domain Name System (DNS) lookup work
- ♦ How does CDN work

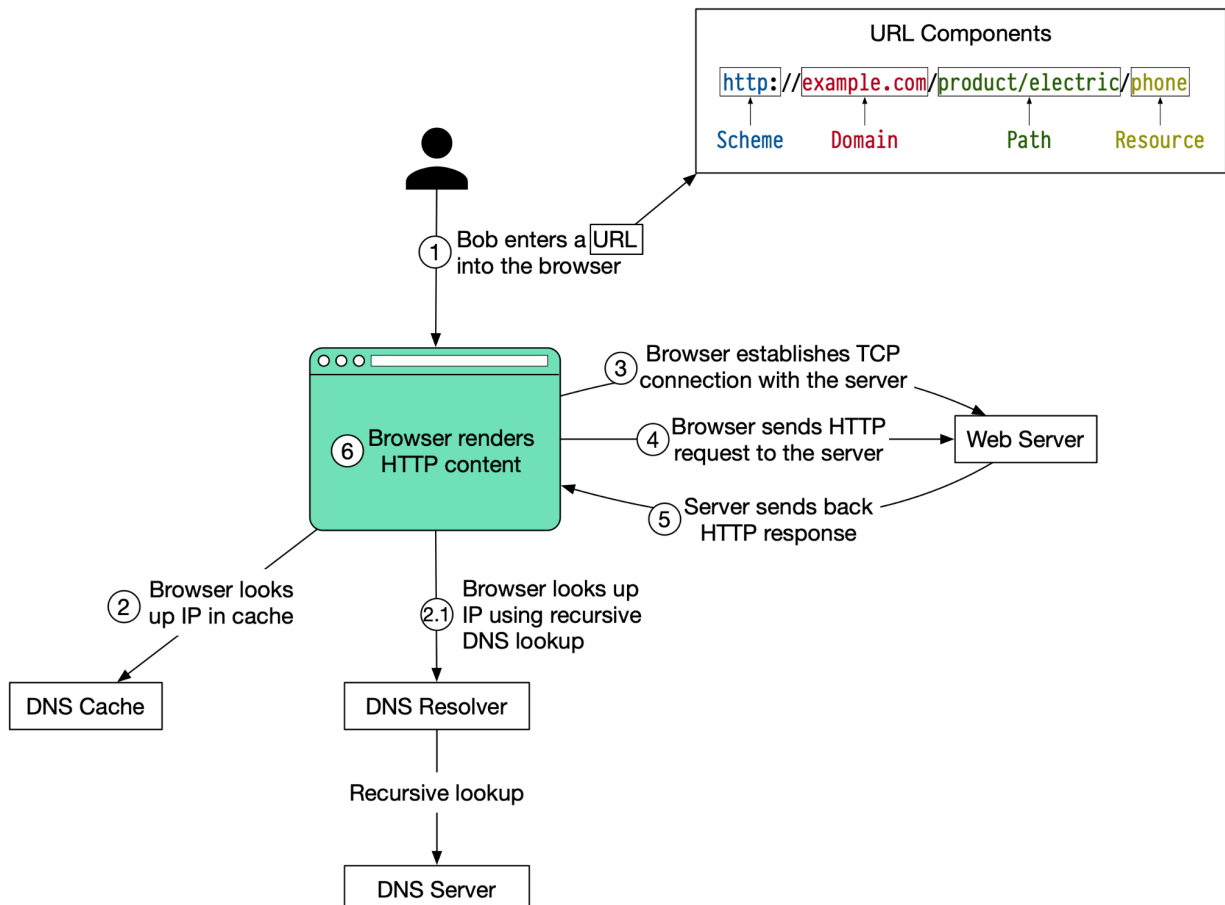
| What happens when you type a URL into your browser?



What happens when you type a URL into your browser?

The diagram below illustrates the steps.

| What happens when you type a URL into your browser?



1. Bob enters a URL into the browser and hits Enter. In this example, the URL is composed of 4 parts:

- ♦ scheme - *https://*. This tells the browser to send a connection to the server using HTTPS.
- ♦ domain - *example.com*. This is the domain name of the site.
- ♦ path - */product/electric/*. It is the path on the server to the requested resource: phone.
- ♦ resource - *phone*. It is the name of the resource Bob wants to visit.

2. The browser looks up the IP address for the domain with a domain name system (DNS) lookup. To make the lookup process fast, data is cached at different layers: browser cache, OS cache, local network cache, and ISP cache.

2.1 If the IP address cannot be found at any of the caches, the browser goes to DNS servers to do a recursive DNS lookup until the IP address is found (this will be covered in another post).

3. Now that we have the IP address of the server, the browser establishes a TCP connection with the server.

4. The browser sends an HTTP request to the server. The request looks like this:

```
GET /iphone HTTP/1.1  
Host: example.com
```

5. The server processes the request and sends back the response. For a successful response (the status code is 200). The HTML response might look like this:

```
HTTP/1.1 200 OK  
Date: Sun, 30 Jan 2022 00:01:01 GMT  
Server: Apache  
Content-Type: text/html; charset=utf-8
```

```
<!DOCTYPE html>  
<html lang="en">  
Hello world  
</html>
```

6. The browser renders the HTML content.

How does the Domain Name System (DNS) lookup work?

DNS acts as an address book. It translates human-readable domain names (google.com) to machine-readable IP addresses (142.251.46.238).

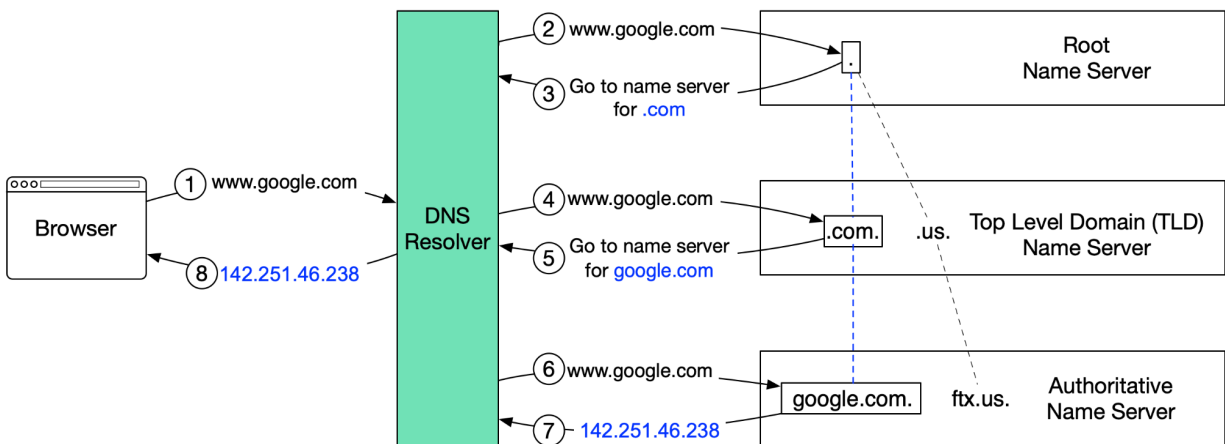
To achieve better scalability, the DNS servers are organized in a hierarchical tree structure.

There are 3 basic levels of DNS servers:

1. Root name server (.). It stores the IP addresses of Top Level Domain (TLD) name servers. There are 13 logical root name servers globally.
2. TLD name server. It stores the IP addresses of authoritative name servers. There are several types of TLD names. For example, generic TLD (.com, .org), country code TLD (.us), test TLD (.test).
3. Authoritative name server. It provides actual answers to the DNS query. You can register authoritative name servers with domain name registrar such as GoDaddy, Namecheap, etc.

The diagram below illustrates how DNS lookup works under the hood:

How does DNS resolve IP



1. google.com is typed into the browser, and the browser sends the domain name to the DNS resolver.

2. The resolver queries a DNS root name server.
3. The root server responds to the resolver with the address of a TLD DNS server. In this case, it is .com.
4. The resolver then makes a request to the .com TLD.
5. The TLD server responds with the IP address of the domain's name server, google.com (authoritative name server).
6. The DNS resolver sends a query to the domain's nameserver.
7. The IP address for google.com is then returned to the resolver from the nameserver.
8. The DNS resolver responds to the web browser with the IP address (142.251.46.238) of the domain requested initially.

DNS lookups on average take between 20-120 milliseconds to complete (according to YSlow).

CDN

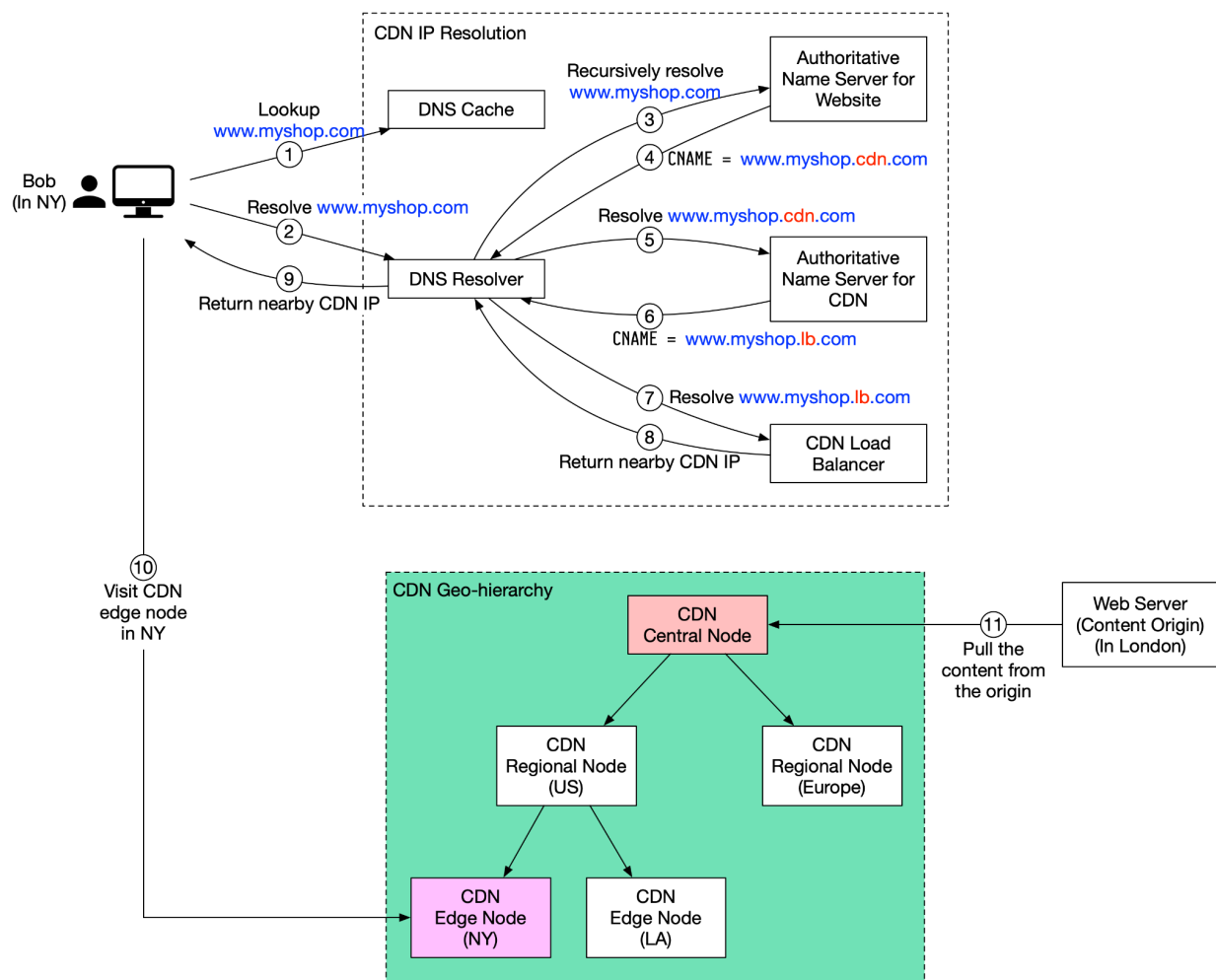
A content delivery network (CDN) refers to geographically distributed servers (also called edge servers) that provide fast delivery of static and dynamic content. Let's take a look at how it works.

Suppose Bob who lives in New York wants to visit an eCommerce website that is deployed in London. If the request goes to servers located in London, the response will be quite slow. So we deploy CDN servers close to where Bob lives, and the content will be loaded from the nearby CDN server.

The diagram below illustrates the process:

How does CDN work

ByteByteGo



1. Bob types in `www.myshop.com` in the browser. The browser looks up the domain name in the local DNS cache.
2. If the domain name does not exist in the local DNS cache, the browser goes to the DNS resolver to resolve the name. The DNS resolver usually sits in the Internet Service Provider (ISP).
3. The DNS resolver recursively resolves the domain name (see my previous post for details). Finally, it asks the authoritative name server to resolve the domain name.
4. If we don't use CDN, the authoritative name server returns the IP address for `www.myshop.com`. But with CDN, the authoritative name server has an alias pointing to `www.myshop.cdn.com` (the domain name of the CDN server).
5. The DNS resolver asks the authoritative name server to resolve `www.myshop.cdn.com`.
6. The authoritative name server returns the domain name for the load balancer of CDN `www.myshop.lb.com`.
7. The DNS resolver asks the CDN load balancer to resolve `www.myshop.lb.com`. The load balancer chooses an optimal CDN edge server based on the user's IP address, user's ISP, the content requested, and the server load.
8. The CDN load balancer returns the CDN edge server's IP address for `www.myshop.lb.com`.
9. Now we finally get the actual IP address to visit. The DNS resolver returns the IP address to the browser.
10. The browser visits the CDN edge server to load the content. There are two types of contents cached on the CDN servers: static contents and dynamic contents. The former contains static pages, pictures, and videos; the latter includes results of edge computing.
11. If the edge CDN server cache doesn't contain the content, it goes upward to the regional CDN server. If the content is still not found, it will go upward to the central CDN server, or even go to the origin - the London web server. This is called the CDN distribution network, where the servers are deployed geographically.