
WEB基础

李 晶 : lijingjing@bupt.edu.cn

Web 概述

- www (World Wide Web) ,万维网
- Internet(互联网)上一个非常重要的信息资源网 , 产生于20世纪90年代初。
- 由许多互相链接的超文本文档组成的系统 , 通过Internet访问。

Web 概述

- 在web中，所有有用的事物都称为 “资源” ；
- 所有资源由一个全局 “统一资源标识符” （ Uniform Resource Identifier,URI ） 标识。
- 用户通过点击链接来获得资源。
- 资源通过超文本传输协议（ HTTP, Hypertext Transfer Protocol ） 传送给用户。

Web 概述

- 超文本传输协议 (HTTP, HyperText Transfer Protocol)
- 客户端浏览器或其他程序与web服务器之间的应用层通信协议，用于实现客户端和服务端的信息传输。

Web 概述

- 统一资源定位符 (URL , Uniform / Universal Resource Locator)
- URL用于完整地描述Internet上网页和其他资源的地址，是实现互联网信息定位统一标识。
- URL是能够唯一标识internet上的计算机、目录和文件位置的命名规则。

Web 概述

- URL主要由三部分组成：协议类型、存放资源的域名或主机IP地址和资源文件名。
- 语法格式如下：

```
protocol://hostname[:port]/path/
```

Web 概述

```
protocol://hostname[:port]/path/
```

- protocol (协议) : 使用的传输协议, 常用的是HTTP协议
- hostname (主机名) : 存放资源的服务器域名或IP地址;
- port (端口号) : 可选项, 省略时使用默认端口, 表示某一服务器在该主机上所使用的TCP端口。如HTTP协议的默认端口是80;
- path (路径) : 由零或多个 “/” 符号隔开的字符串, 一般用来表示服务器主机上提供服务的信息资源所在的一个目录或文件地址。
- 例如 : **http://baike.baidu.com:80/view/560377.htm**

↑ ↑ ↑ ↑ ↑
协议 主机地址 端口 目录 文件名

Web 概述

- 超文本标记语言（ HTML , HyperText Mark-up Language ）
 - 一种描述文档结构的语言。HTML语言通过标记和属性对展示内容进行描述。
- 目前网络上应用最为广泛的页面描述语言，也是构成网页文档的主要语言，其特点如下：
 - 简易性
 - 可扩展性
 - 平台无关性

Web常用术语

- Web页面/网页：
 - 通常在浏览器中所看到的网页，其实是一个单一的文件。
 - 用HTML语言编写的文本文件，包含有文字、表格、图像、链接、声音和视频等。
- 网站：若干网页及网页上所用的资源按一定方式组织在一起，放在服务器上，提供相关信息资源。
- 主页：首页，是网站的第一个页面。通常和一个网址(URL)相对应，引导用户浏览网站。
 - index.asp index.html default.html

Web常用术语

- 超链接 (Hyper Link) :
 - 一个网页的一部分
 - 允许一个网页同其他网页或站点之间进行链接的元素。
 - 一个网页指向另一个目的端的链接，目的端可以是另一个网页或者同一个网页上的不同位置或其它资源

B/S 网络架构

- 应用系统架构主要分为两种：
 - C/S结构（ Client/Server， 客户/服务器模式 ）
 - B/S结构（ Browser/Server， 浏览器/服务器模式 ）

浏览器访问网页的过程

- 浏览器的工作机制：
- web浏览器与web服务器之间通过HTTP协议进行通信的过程。



第一步：
在浏览器的地址栏
中输入URL，回车

第二步：
解析域名找到主机
IP地址



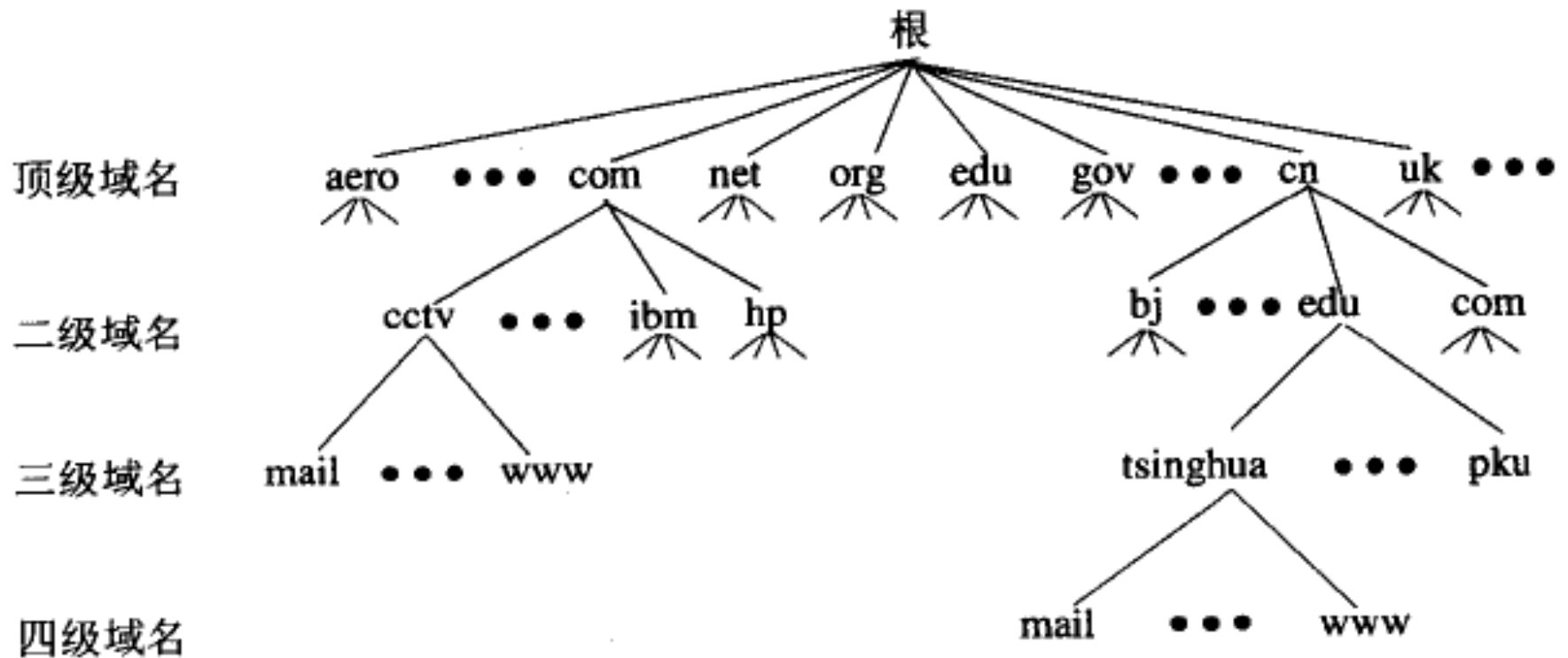
域名系统

- 域名(Domain Name)
 - 因特网命名采用层次树状结构的命名方法。任何一个连接在因特网上的主机或路由器，都有一个唯一的层次结构的名称，即域名(domain name)。
 - 从语法上讲，每一个域名都是由标号(label)序列组成，而各标号之间用点(小数点)隔开。
- 例如：baike.baidu.com
 - 三级域名
 - 二级域名
 - 顶级域名

域名系统

- 域名中的标号由英文和数字组成
- 每一个标号不超过63个字符(为了记忆方便，一般不会超过12个字符)
- 不区分大小写字母
- 除连字符(-)外不能使用其他的标点符号
- 级别最低的域名写在最左边，而级别最高的字符写在最右边。
- 由多个标号组成的完整域名不超过255个字符。

域名系统

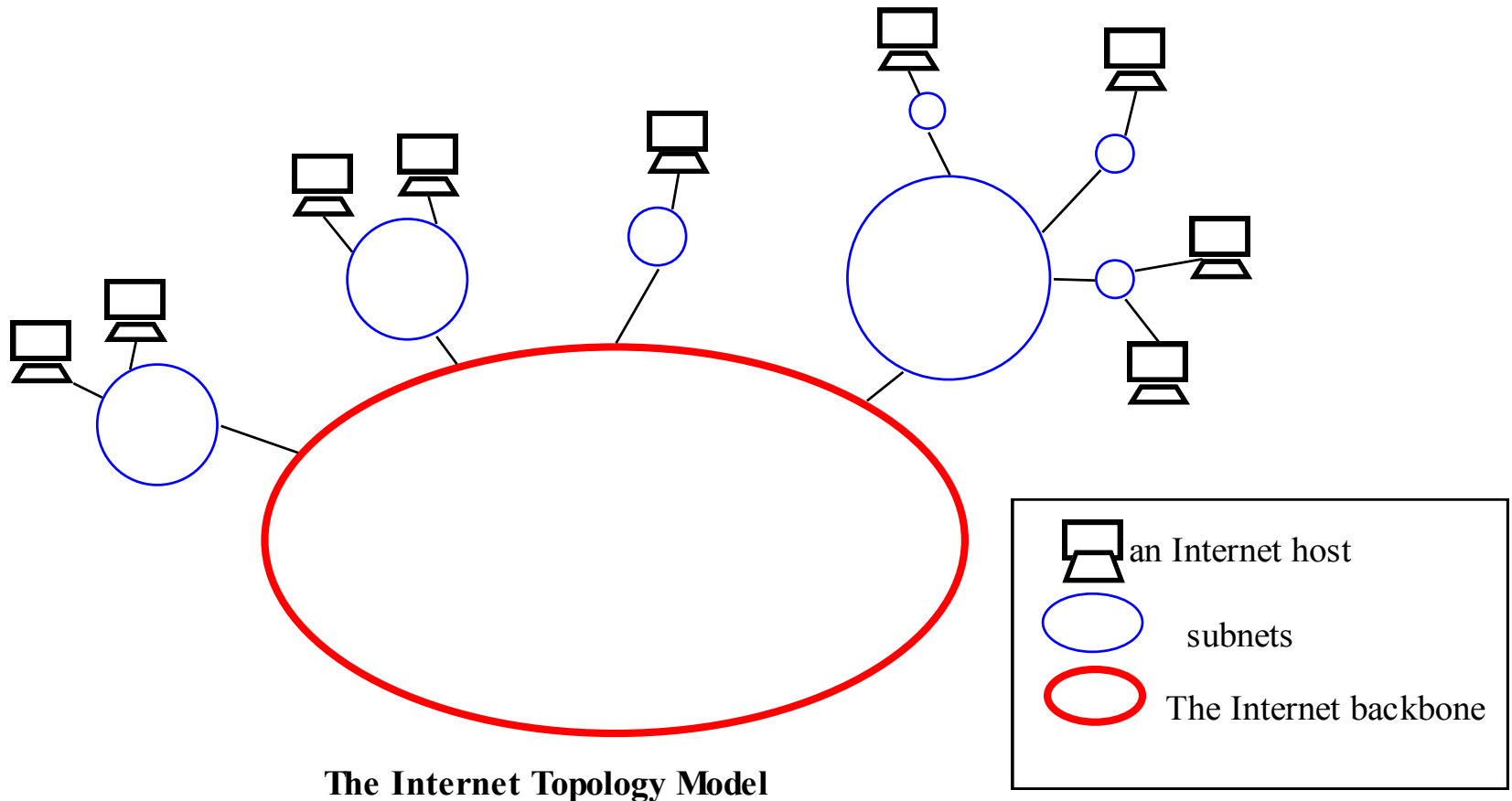


因特网的域名空间

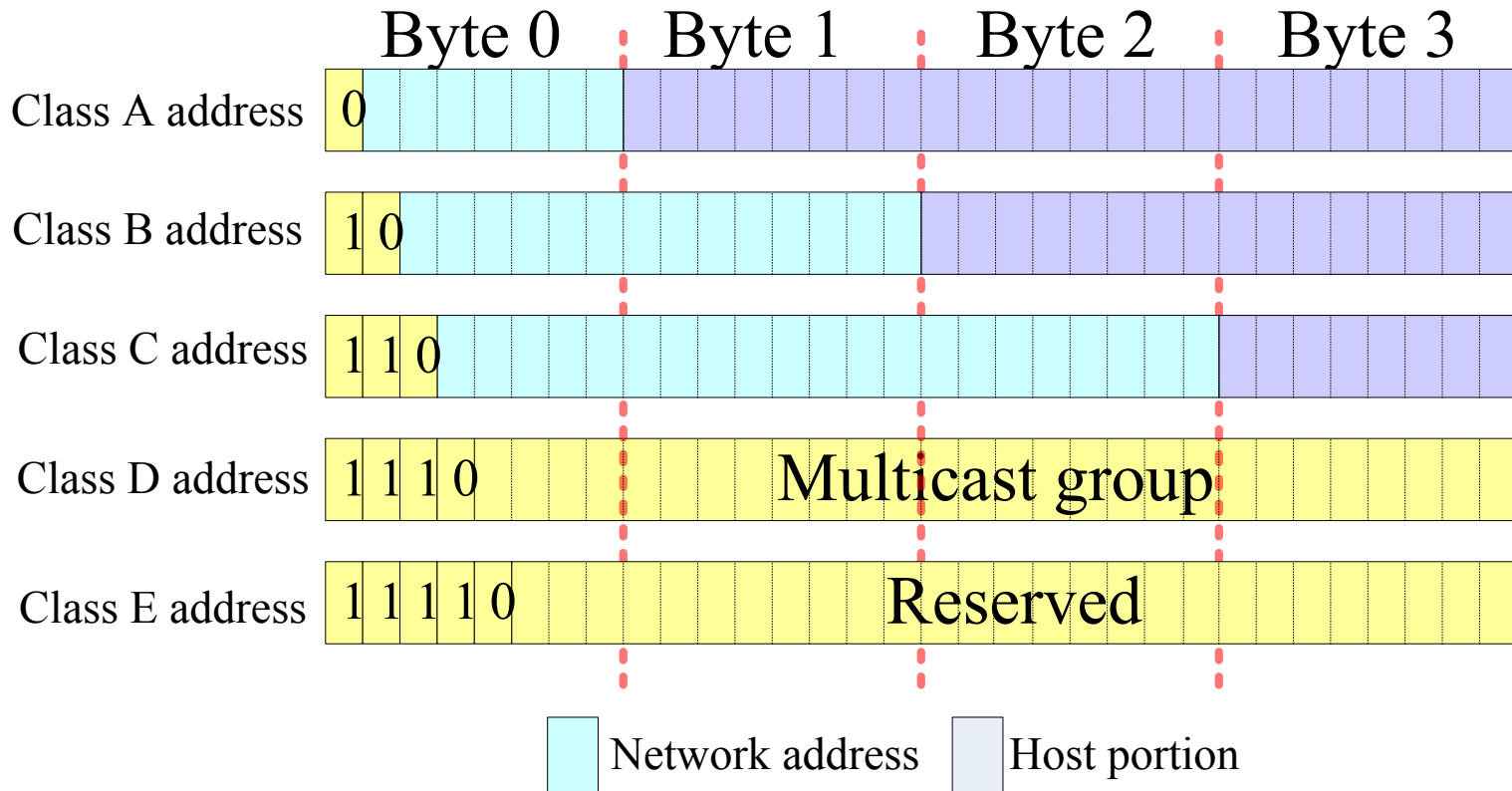
域名系统

- 域名系统DNS(Domain Name System)
 - 因特网使用的命名系统，用来把便于人们使用的机器名字转换成为IP地址
- IP(Internet Protocol)地址
 - 网络之间互连的协议
 - 由32位二进制数字组成
 - 计算机之间通信时使用的识别主机的方法

Internet 拓扑结构

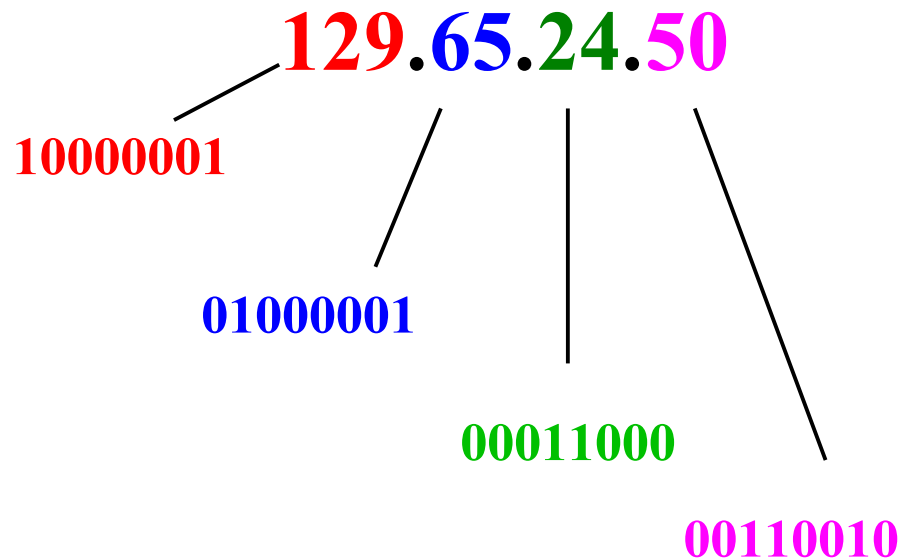


IP 地址



IP 地址

二进制IP地址到点分十进制IP地址的转换



IP 地址

- 点分十进制IP地址到符号化地址的转换
- 依靠DNS（域名服务系统）

解析域名过程

- 第一步：浏览器会缓存DNS一段时间。如果有缓存，直接返回IP，否则下一步。
- 第二步：如缓存中无法找到IP，浏览器会查询操作系统hosts文件。如果找到，直接返回IP，否则下一步。
 - 在Windows中可以通过`c:\windwos\System32\drivers\tec\hosts`文件来设置
 - 在Linux中可以通过`/etc/named.conf`文件来设置
 - 用户可以将任何域名解析到任何能够访问的IP地址

解析域名过程

- 第三步：路由器一般都会有自己的DNS缓存，ISP服务商DNS缓存，如果有缓存，直接返回IP，否则下一步。
 - 在网络配置中有“DNS服务器地址”这一项。
 - 在Windows下可以通过ipconfig查询这个地址

```
以太网适配器 本地连接:

    连接特定的 DNS 后缀 . . . . . : taobao.ali.com
    本地连接 IPv6 地址. . . . . : fe80::5045:aa87:2899:3d14%11
    IPv4 地址 . . . . . : 10.13.35.2
    子网掩码 . . . . . : 255.255.255.0
    默认网关. . . . . : 10.13.35.250
```

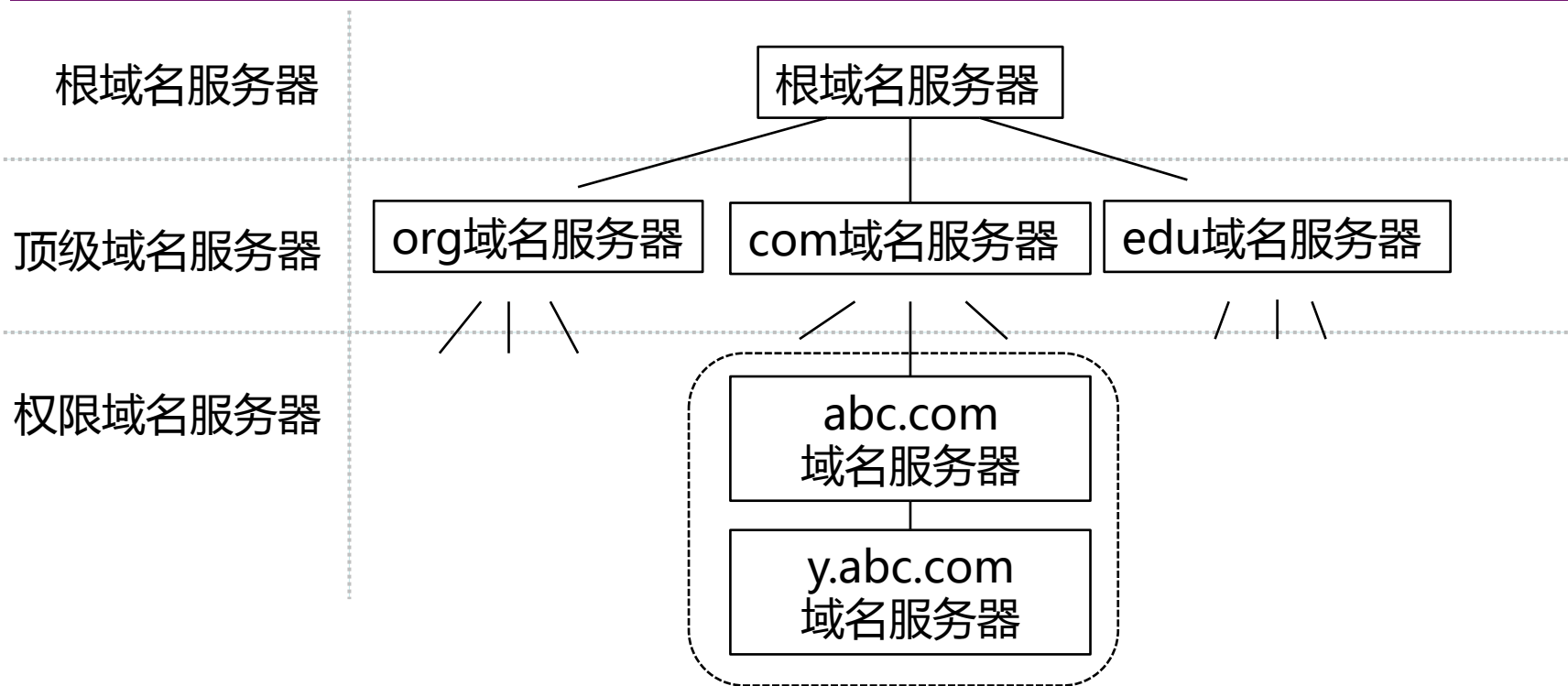
- 在Linux下可以通过如下方式查询配置的DNS Server

```
[junshan@v101055.sqa.cm4 admin]$ cat /etc/resolv.conf
search tbsite.net
nameserver 10.232.2.254
nameserver 172.24.102.254
nameserver 10.246.124.254
```


解析域名过程

- 第四步：DNS解析。DNS服务器开始从根域名服务器搜索，到要查找的目标域名服务器。

域名系统



解析域名过程

- 第五步：返回IP地址给客户。





第三步：
浏览器凭借得到的IP与服务器建立连接，并向服务器发起查询请求。



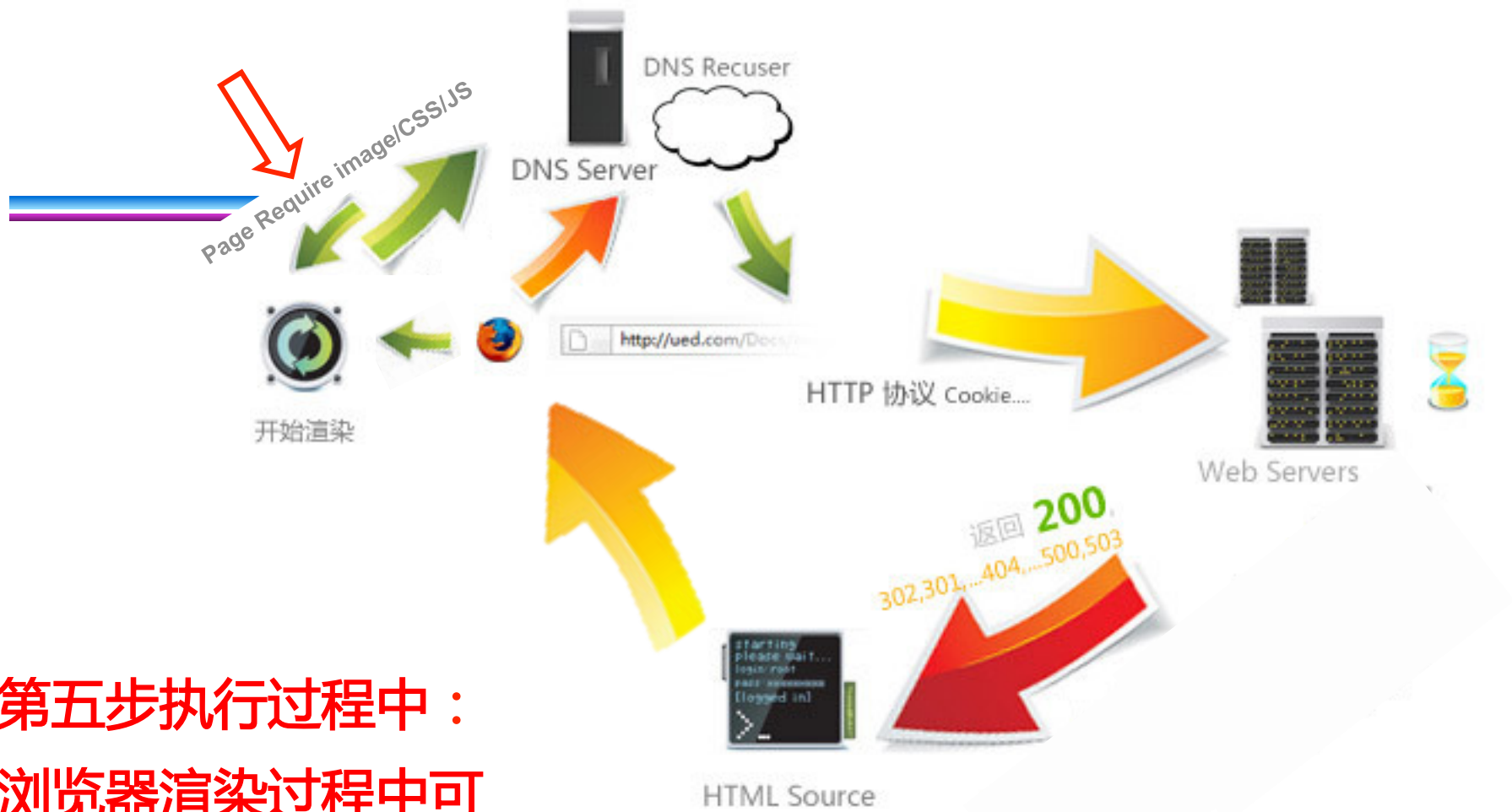
第四步：
服务器返回浏览器请求的页面的源代码。



第四步之后：
页面的源代码被浏览器接收到。



第五步：
浏览器开始渲染接收到的页面源代码。



第五步执行过程中：
浏览器渲染过程中可能会再次向DNS发起查询请求，以获得页面渲染过程中需要的图片、脚本和样式表等外部资源。

WEB服务器

- Microsoft IIS(Internet Information Services)
- WebSphere
- BEA WebLogic Server
- APACHE
- Tomcat