**老男孩linux运维实战培训**

**老男孩教育教学核心思想6重：重目标、重思路、重方法、重实践、重习惯、重总结**

**学无止境，老男孩教育成就你人生的起点！**

**版权声明：**

本文作者为《老男孩linux运维实战培训》学生—**你的名字**

本文的所有内容均来自老男孩培训**命令总结**，未经本人及老男孩培训许可，禁止私自转发及使用。

QQ: **你的**XXXX

E-mail: **你的**XXXX

**联系方式:**

|  |  |  |  |
| --- | --- | --- | --- |
| 网站运维交流群： | | | |
| **Linux QQ交流群 385168604** | | **架构师 QQ交流群 390642196** | |
| **Python QQ交流群 29215534** | | **大数据 QQ交流群 421358633** | |
| =================================================================================== | | | |
| **老男孩linux实训联系方式** |  | |  |
| **咨询QQ** | **41117397（丹丹）** | | **70271111（歪歪）** |
|  | **80042789（飞雪）** | | **390320151（小雨）** |
|  | **41117483（冰冰）** | |  |
| **电话：** | **158-1059-0206（丹丹）** | | **189-1171-8229（歪歪）** |
|  | **135-5261-2571（飞雪）** | | **186-0046-2391（小雨）** |
| **网站:** | <http://www.etiantian.org> | | <http://www.oldboyedu.com> |
| **博客:** | http://oldboy.blog.51cto.com | | <http://blog.oldboyedu.com> |

**快捷方式说明:**

**ctrl + 1 一级标题**

**ctrl + 2 二级标题**

**ctrl + 3 三级标题**

**ctrl + 5 程序代码**

**ctrl + 6 正文**

**格式约定：**

蓝色字体：内容注释

目 录

[http协议及www服务应用课程总结 1](#_Toc510989031)

[第1章 HTTP协议 1](#_Toc510989032)

[1.1.1 用户访问网站基本流程 1](#_Toc510989033)

[1.2 DNS系统解析基本流程 2](#_Toc510989034)

[1.2.1 DNS解析流程： 3](#_Toc510989035)

[1.3 http服务重要基础 6](#_Toc510989036)

[1.4 HTTP简介 7](#_Toc510989037)

[1.4.1 http协议请求方法 7](#_Toc510989038)

[1.4.2 http状态码（HTTP Status Code） 7](#_Toc510989039)

[1.4.3 生产场景常见状态代码： 8](#_Toc510989040)

[1.4.4 状态码查看 8](#_Toc510989041)

[1.5 HTTP报文 9](#_Toc510989042)

[1.5.1 请求报文-Request Message 9](#_Toc510989043)

[1.5.2 HTTP响应报文-Response Message 9](#_Toc510989044)

[1.5.3 HTTP资源 12](#_Toc510989045)

[1.5.4 URL介绍 12](#_Toc510989046)

[1.5.5 URI介绍 12](#_Toc510989047)

[1.5.6 静态网页资源： 12](#_Toc510989048)

[1.5.7 动态网页资源 13](#_Toc510989049)

[1.5.8 伪静态网页 13](#_Toc510989050)

[第2章 网站流量度量术语 14](#_Toc510989051)

[2.1 IP-独立IP数 14](#_Toc510989052)

[2.2 PV-访问量-页面浏览 14](#_Toc510989053)

[2.3 UV-独立访客 14](#_Toc510989054)

[2.4 企业网站对IP PV UV的度量 14](#_Toc510989055)

[2.4.1 IP的度量 14](#_Toc510989056)

[2.4.2 PV的度量 14](#_Toc510989057)

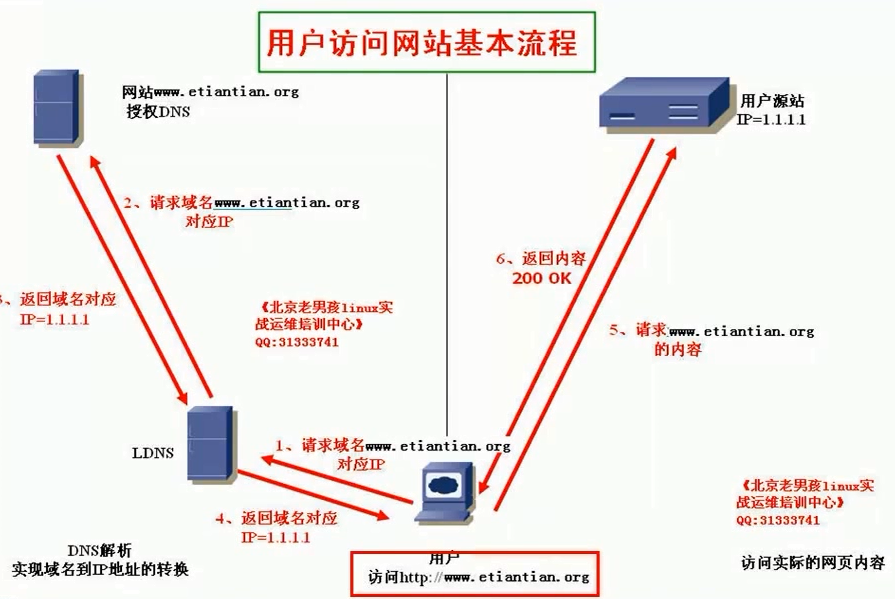
[2.4.3 UV的度量 15](#_Toc510989058)

http协议及www服务应用课程总结

# HTTP协议

DNS解析复习、HTTP访问流程

### 用户访问网站基本流程



1、客户端浏览器输出[www.baidu.com](http://www.baidu.com)后，计算机先查找本机hosts以及DNS缓存记录，有则直接查找对应baidu服务器获取服务；没有的话查找LDNS服务

2、LDNS有的话直接返回baidu的IP地址，没有的话去查找根DNS服务器

3、根DNS服务器没有，但只有.com域名地址，将其返回给LDNS

4、LDNS去查找.com服务器, .com服务器有baidu.com的记录，将其返回给LDNS

5、LDNS向baidu.com去获取[www.baidu.com](http://www.baidu.com)服务器IP，授权服务器查到对应的IP后将其反馈给LDNS。

6、LDNS获取到[www.baidu.com对应的IP](http://www.baidu.com对应的IP)后自己先缓存，并将IP反馈给客户端

7、客户端使用获取到的正确IP信息向[www.baidu.com](http://www.baidu.com)请求资源(TCP三次握手-集群架构-负载均衡-nginxWEB-存储-缓存、DB)。百度相应客户端请求

8、此时客户端可以访问到请求的资源页面

至此-整个访问过程完成

Windos 查看DNS缓存记录

ipconfig /displaydns

清除DNS缓存

ipconfig /flushdns

企业真实场景面试题：

1、请描述http协议原理

知识内容：

1、用户访问网站流程框架

2、DNS解析原理

3、tcp/ip三次握手原理

4、http协议原理（www服务的请求过程）细节、报文细节

5、大规模网站集群架构细节

6、http协议原理（www服务响应过程）响应报文细节

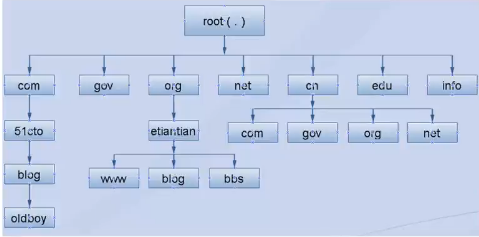
7、tcp/ip四次挥手过程原理

## DNS系统解析基本流程

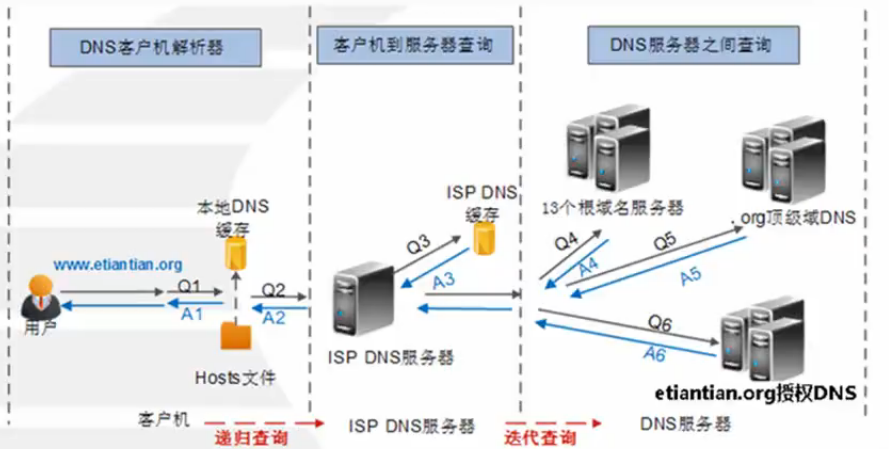
DNS 域名解析系统，网站关键，核心负责把网站域名解析为对应的IP，从域名到IP的过程称为A记录，即address Record.

* 设置CNAME 报名记录，这个功能常备CDN加速服务商应用
* 设置MX邮件记录，这个MX记录功能，常用邮件服务
* 设置PTR记录，反向解析，把IP地址解析为对应的域名，和A记录的解析相反，邮件服务业务中会用到。

DNS倒挂树结构[.]类似linux文件系统结构[/]

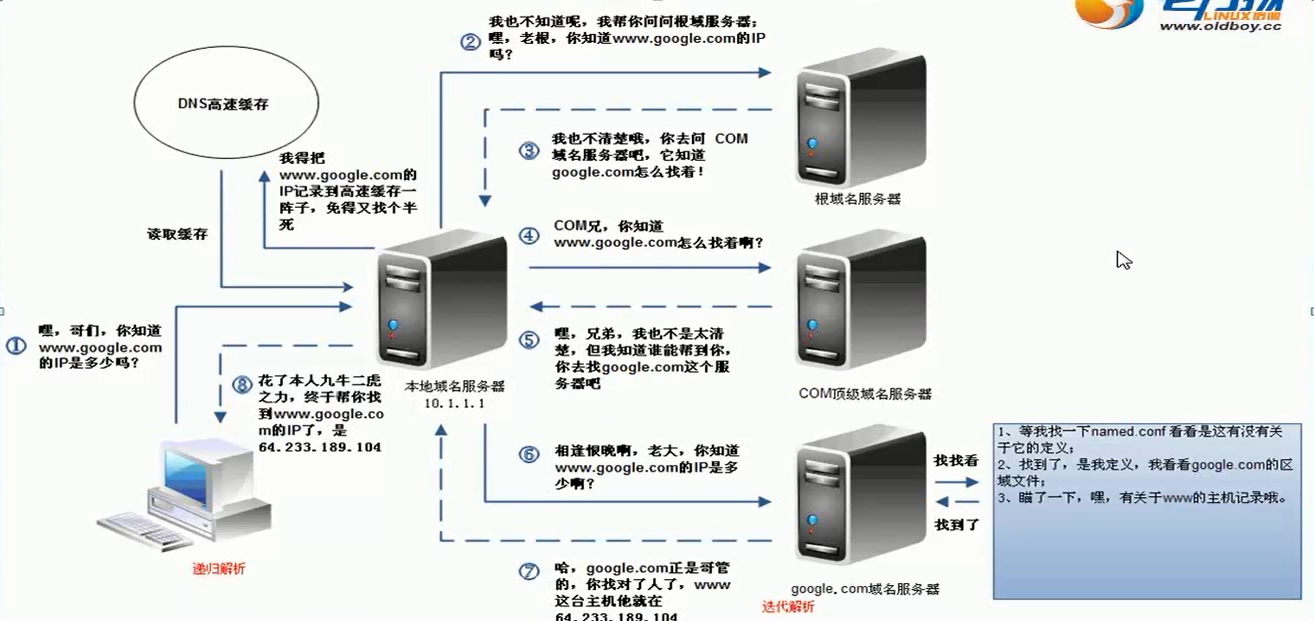


### DNS解析流程：



**递归：是重复调用模块自身实现循环**

**迭代：函数内某段代码实现循环**



***No.1*** 客户端用户在浏览器里输入[www.baidu.com](http://www.baidu.com)网站地址，回车后，系统首先查找系统本地的DNS缓存及host文件信息，确定是否存在www.baidu.com 域名对应的IP解析记录，如果有就直接获取IP地址，然后去访问这个IP地址对应的域名[www.baidu.com](http://www.baidu.com)的服务器。注：一般第一次请求时，DNS缓存是没有解析记录的，而hosts多在内部临时测试使用。

***No.2*** 如果客户端本地DNS缓存及hosts文件没有[www.baidu.com](http://www.baidu.com)域名对应的解析记录，那么，系统会把浏览器的解析请求发送给客户端本地设置的DNS服务器地址（通常称此DNS为LocalDNS，即本地DNS）解析，如果LDNS服务器的本地缓存有对应的解析记录就会直接返回IP地址给客户端，如果没有，则LDNS会继续请求其他（上一层…）DNS服务器。

***No.3***  LDNS从DNS系统的（.）根开始请求对[www.baidu.com域名的解析，根DNS服务器在全球一共有13](http://www.baidu.com域名的解析，根DNS服务器在全球一共有13)台，根服务器下面没有[www.baidu.com](http://www.baidu.com)域名解析记录的，但是根下面有[www.baidu.com对应的顶级域名.com](http://www.baidu.com对应的顶级域名.com)的解析记录，因此，根会把.com对应的DNS服务器地址返回给LDNS。

***No.4*** LDNS服务器获取到.com对应的DNS服务器地址后，就会去.com服务器请求[www.baidu.com域名的解析，而.com](http://www.baidu.com域名的解析，而.com)服务器下面也没有[www.badu.com域名对应的解析记录，但是有baidu.com](http://www.badu.com域名对应的解析记录，但是有baidu.com)域名的解析记录，因此，.com服务器会发baidu.com对应的DNS服务器地址返回给LDNS。

***No.5*** 同理，LDNS获取到baidu.com对应的DNS服务器地址后，就会去baidu.com服务器请求[www.baidu.com](http://www.baidu.com)域名的解析，baidu.com域名对应的DNS服务器是该域名的授权服务器，而这个授权DNS服务器正是企业购买域名时用于管理域名解析的服务器，这个授权服务器会有[www.baidu.com对应的IP](http://www.baidu.com对应的IP)解析记录。如果此时没有，就表示企业的域名管理人员没有为[www.baidu.com](http://www.baidu.com)域名解析设置，即网站没有架设好。

***No.6***  baidu.com域名的授权DNS服务器会把[www.baidu.com对应的最终IP](http://www.baidu.com对应的最终IP)解析记录发给LDNS。

***No.7*** LDNS把来自授权DNS服务器[www.baidu.com对应的IP](http://www.baidu.com对应的IP)解析记录发给客户端浏览器，并且LDNS它会在本地把该域名和IP对应解析缓存起来，以便下一次更快地返回相同的解析请求的记录，至此完成整个DNS的解析流程。

为方便理解dns解析记录可做如下操作：

[root@backup ~]# dig +trace www.baidu.com

; <<>> DiG 9.8.2rc1-RedHat-9.8.2-0.37.rc1.el6 <<>> +trace www.baidu.com

;; global options: +cmd

. 5 IN NS g.root-servers.net.

. 5 IN NS m.root-servers.net.

. 5 IN NS a.root-servers.net.

. 5 IN NS b.root-servers.net.

. 5 IN NS k.root-servers.net.

. 5 IN NS c.root-servers.net.

. 5 IN NS h.root-servers.net.

. 5 IN NS d.root-servers.net.

. 5 IN NS e.root-servers.net.

. 5 IN NS l.root-servers.net.

. 5 IN NS j.root-servers.net.

. 5 IN NS f.root-servers.net.

. 5 IN NS i.root-servers.net.

;; Received 508 bytes from 10.0.0.2#53(10.0.0.2) in 322 ms

com. 172800 IN NS b.gtld-servers.net.

com. 172800 IN NS d.gtld-servers.net.

com. 172800 IN NS m.gtld-servers.net.

com. 172800 IN NS j.gtld-servers.net.

com. 172800 IN NS e.gtld-servers.net.

com. 172800 IN NS l.gtld-servers.net.

com. 172800 IN NS k.gtld-servers.net.

com. 172800 IN NS g.gtld-servers.net.

com. 172800 IN NS h.gtld-servers.net.

com. 172800 IN NS a.gtld-servers.net.

com. 172800 IN NS i.gtld-servers.net.

com. 172800 IN NS c.gtld-servers.net.

com. 172800 IN NS f.gtld-servers.net.

;; Received 491 bytes from 192.33.4.12#53(192.33.4.12) in 690 ms

baidu.com. 172800 IN NS dns.baidu.com.

baidu.com. 172800 IN NS ns2.baidu.com.

baidu.com. 172800 IN NS ns3.baidu.com.

baidu.com. 172800 IN NS ns4.baidu.com.

baidu.com. 172800 IN NS ns7.baidu.com.

;; Received 201 bytes from 192.33.14.30#53(192.33.14.30) in 138 ms

www.baidu.com. 1200 IN CNAME www.a.shifen.com.

a.shifen.com. 1200 IN NS ns4.a.shifen.com.

a.shifen.com. 1200 IN NS ns3.a.shifen.com.

a.shifen.com. 1200 IN NS ns1.a.shifen.com.

a.shifen.com. 1200 IN NS ns5.a.shifen.com.

a.shifen.com. 1200 IN NS ns2.a.shifen.com.

;; Received 228 bytes from 220.181.37.10#53(220.181.37.10) in 186 ms

## http服务重要基础

HTTP协议，超文本传输协议，是互联网最常用的一种网络协议。HTTP的重要应用之一是WWW服务。设计HTTP协议最初的目的是提供一种发布和接受HTML（页面标记语言）页面的方法

WWW全称Word Wide Web,即web 中文称’万维网’ HTTP协议www服务默认端口80，HTTPS协议为加密www服务端口443



http协议www应用

浏览器（Browser）和服务端（Server）,B/S架构，擅长广域网（互联网）应用

## HTTP简介

**Version1.0** 第一个广泛使用的HTTP版本，在0.9基础上增加了HTTP请求头，支持更多的请求方法，对多媒体进行处理。规定浏览器与服务器只保存短暂的连接，每次请求都需求与服务器建立一个TCP连接，服务器完成请求处理后即断开TCP连接，不跟踪每个客户也不记录过去的请求。

**Version1.1** 在1.0基础上做了扩展性，缓存处理，带宽优化，持久连接，HOST头，错误通知，消息传递。在连接上可传送多个HTTP请求和响应，减少了建立和关闭连接的小号和时间延迟。

HTTP1.1支持长连接（PersistentConnection）和请求的流水线（Pipelining）处理，HTTP/1.1在1.0的基础上加入了一些cache的新特性

### http协议请求方法

在通信中，每个http请求报文都包含一个方法。用于告诉we服务器端需求执行那些具体的动作

获取，提交，删除等

|  |  |
| --- | --- |
| HTTP方法 | 作用描述 |
| GET | 客户端请求指定资源信息，服务器返回指定资源 |
| HEAD | 只请求响应报文中的HTTP头部 |
| POST | 将客户端数据提交给服务器，例如：注册表单 |
| PUT | 用客户端从服务器传送的数据取代指定的文档内容 |
| DELETE | 请求服务器删除Request-URL所表示的资源 |
| MOVE | 请求服务器将指定的页面移至另一个网络地址 |

### http状态码（HTTP Status Code）

表示Web服务器响应http请求状态的数字代码。每当Web客户端向Web服务器发送一个请求时，服务器都会返回一个状态响应代码。就是告诉客户端此处是否成功

|  |  |
| --- | --- |
| 状态码范围 | 作用描述 |
| 100-100 | 指定客户端响应的某些动作 |
| 200-299 | 表示请求成功 |
| 300-399 | 用于已经一点的文件并且常被包含在定位头信息中指定新的地址信息 |
| 400-499 | 指定客户端的错误 |
| 500-599 | 指出服务器的错误 |

### 生产场景常见状态代码：

|  |  |
| --- | --- |
| 状态码 | 描述 |
| 200~OK | 服务器成功返回 |
| 301-Moved Permanently | 永久跳转，所有亲戚的网页将永久跳转到新的位置 |
| 403-Forbidden | 禁止访问，这个请求是合法的，但是服务器端因为匹配了预先设置的规则拒绝响应客户端的请求，此类问题一般为服务器或服务权限配置不当所致  1.index页面不存在；2.权限设置；3.主配置没指定index路径  http://blog.51cto.com/search/user?uid=2561410&q=403 |
| 404-Not Found | 服务器找不到客户端请求的页面，可能是客户端请求了服务器上不存在的资源 |
| 500-Internal SERVER Error | 内部服务器错误，服务器遇到了意料不到的情况，不能完成客户的请求。SElinux开启，而又没有为HTTP设置规则许可，客户端访问就是500 |
| 502-Bad Gateway | 坏的网关，一般是代理服务器请求后端服务时，后端服务不可用或没有完成响应网关服务器。这通常为反向代理服务器下面的节点出现问题。中介找房子，房东挂了 |
| 503-Service Unavailable | 服务当前不可用，服务器超载或停机维护导致，或反向代理服务器后面没有可提供服务的节点 |
| 504-Gateway Timeout | 网关超时 后端处理超时多时服务器没有在指定时间内返回数据给前端代理服务器 |

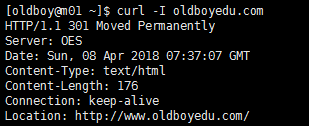
### 状态码查看

[oldboy@m01 ~]$ curl -I baidu.com

HTTP/1.1 200 OK

[oldboy@m01 ~]$ curl -I oldboyedu.com

HTTP/1.1 301 Moved Permanently



## HTTP报文

客户端到服务器的叫请求报文-Request Message

服务器端响应客户端的叫响应报文-Response Message

### 请求报文-Request Message

由请求行、请求头、空行、请求报文主体组成

|  |  |
| --- | --- |
| **报文格式** | **报文信息** |
| **请求行** | **请求方法URL协议版本** |
| **请求头** | **字段名1：值1**  **字段名2：值2** |
| **空行** | **空白无内容** |
| **请求报文主体** | **GET方法没有请求报文主体，POST才有** |

请求行：是请求报文的第一行，说明客户端想要做什么，内容由请求方法字段、URL字段、http协议版本组成

|  |  |  |
| --- | --- | --- |
| **请求方法** | **URL字段** | **HTTP协议版本** |
| **GET** | **/index.html** | **HTTP/1.1** |

请求头：由关键字/值组成，每行一对，关键字和值用英文冒号“:”分隔，其作用是通过客户端吧请求的相关信息高速服务器

|  |  |
| --- | --- |
| **请求头信息** | **说明** |
| **Accept:iage/gif,image/jpeg** | **媒体类型** |
| **Accept-Language:zh-cn** | **语言类型** |
| **Accept-Encoding:gzip,deflate** | **支持压缩** |
| **User-Agent:Mozilla/4.0(compatible;MSIE6.0;Windows NT)** | **客户端类型OS** |
| **Host:www.baidu.com** | **主机名** |

空行：是通过发送回车或换行符，通知Web服务器空行一下不会有请求头部信息

请求报文主体：包含了要发送给Web服务器的数据信息。请求报文主体不会英语于HTTP的GET命令，而是用于POST方法

### HTTP响应报文-Response Message

由起始行、响应头部(header)、空行、响应报文主体组成

|  |  |
| --- | --- |
| **报文格式** | **报文信息** |
| **起始行** | **协议及版本号、数字状态码、状态信息** |
| **响应报头** | **字段名1：值1**  **字段名2：值2** |
| **空行** | **无内容** |
| **响应报文主体** | **< html >**  **< head >< title > oldboy’s blog < /title >< /head >**  **< boby >**  **I am oldnoy,mysql blog is http://blog.51cto.com/oldboy**  **< /boby >**  **< /html >** |

起始行：也叫状态行，用来说明服务器响应客户端请求的状态，一般为协议及版本号、数字状态码、状态情况 例如：HTTP/1.1 200 OK HTTP/1.1 301 Moved Permanently

响应头部：和请求报文类似，起始行后面一般有若干个头部字段，每个字段都包含一个名字和一个值，两者之间用冒号分隔，头部结尾也是以一个空行结束常见头部如下：

空行：是通过发送回车或换行符，通知Web服务器空行一下不会有请求头部信息

响应报文主体：装载了要返回给客户端的数据，这些数据可以是文本，也可以是二进制的图片、视频，下面是响应报文主体的html格式文本示例：

< html >

< head >< title > oldboy’s blog < /title >< /head >

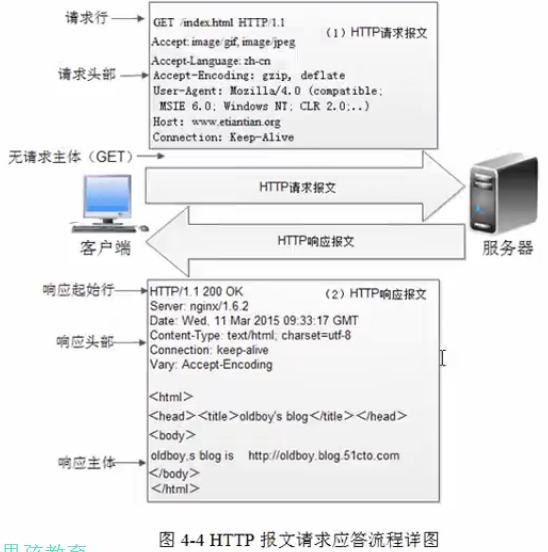
< boby >

I am oldnoy,mysql blog is http://blog.51cto.com/oldboy

< /boby >

< /html >

***请求以及响应报文图例***



HTTP协议原理及重点分析

HTTP属于OSI中的第七层应用层协议，在客户端向服务器端发送请求报文前，先要通过TCP/IP协议在Web客户端和服务器之间建立一个TCP/IP连接（三次握手）。这个HTTP协议请求工作流程如下：

1. 终端在浏览器中输入[www.baidu.com](http://www.baidu.com)
2. Web浏览器请求DNS服务器把域名转换成Web服务器的IP地址【DNS解析知识】
3. Web浏览器将端口号80从访问地址URL解析出来
4. Web浏览器通过解析后的IP地址及端口号与Web服务器之间建立一条TCP连接（三次握手）【TCP/IP三次握手知识】
5. 建立TCP连接后，Web浏览器向服务器发送一条HTTP请求报文【http请求报文知识】
6. Web服务器响应并读取浏览器的请求，后返回一条HTTP响应报文【http响应报文知识】
7. Web服务器关闭HTTP连接，关闭TCP连接（四次挥手），Web浏览器显示访问的网站内容【TCP/IP四次挥手知识】

***到达HTTP服务后端集群节点流程Nginx-fastcgi-PHP-数据库，存储***

### HTTP资源

媒体类型-常见媒体文件格式等..

### URL介绍

统一资源定位符，也称网页地址如同名牌一样，是因特网上标准的资源唯一的地址。

url组成：

第一部分是协议：http https

第二部分主机资源服务器IP地址或域名 [www.baidu.com](http://www.baidu.com)

第三部分主机资源具体地址、如目录和文件名 [www.baidu.com/index.html](http://www.baidu.com/index.html)

第一部分和第二部分之间“://”分隔，第二部分个第三部分用“/”分隔

### URI介绍

统一资源标识符。标识某互联网资源名称的字符串，URL是URI的子集。

### 静态网页资源：

实质纯粹HTML格式的网页可以包含图片、视频、JS、CSS通常被称为静态网页，个人简单理解就是没有后台数据库、不含程序如PHP、JSP、ASP、不可交换的网页。

#### 静态网页的特点：

* 纯文件类程序或文件.html、.html、.xml、js、.css等
* 图片类文件或数据文档.jpg、.gif、.png、.bmp、.doc
* 视频类流媒体文件.mp4、.swf、.avi、.wmv、.fiv

#### 重要特征：

1. 页面URL固定后缀.htm、.html、，且地址中不包含“？”“&”
2. 实实在在存在服务器上的文件实体，每个网页都是一个独立的文件
3. 网页内容是固定不变的，因此容易被搜索引擎收录
4. 网页没有数据库的支持
5. 网页交互性差
6. 网页程序在用户浏览器解析，程序解析效率高。

#### 静态网页语言

常见静态语言有huml、js、css等

1. 程序在客户端浏览器解析，不读取数据库，因此性能和效率很高
2. 因为后端没有数据，所有和用户的交互性较差，功能实现少库

#### 静态网页的架构思想

在高并发，高访问量场景下，涉及关键环节就是把动态网页转成静态网页，而不直接请求数据库和动态服务器，并且可以把静态内容推送到前端缓存中提供服务，这样可以提升用户体验，以及节约了服务器和维护成本。

### 动态网页资源

相对静态而言，在网页网址汇总有标志性的符号“?”“&”,在大多数情况下后端都需要数据库的支持。

#### 动态网页资源特点

1. 网页以数据库技术为基础
2. 动态网页中有“?”“&”但存在风险
3. 在服务器端解析

#### 动态网页的架构思想

一般来讲，静态网页的性能效率是动态网页的10~30倍。且动态网站效率很差，并发能力也很低，在高并发场景中，应尽可能转换成静态网页提供服务。动态转静态几乎是所有高并发网站必备的架构思路，也是高级架构师的职责。动态转静态页要根据业务需求设计

### 伪静态网页

#### 伪静态网页介绍

伪静态就是通过某些技术(rewrite)，把动态网页的URL地址伪装成静态的URL,实质上还是动态网页，只不过是符合静态网页地址特征罢了。

#### 伪静态特点

实质是通过rewrite规则来实现URL地址重写 。

#### 伪静态作用

1. 让搜索引擎收录；
2. 提升用户访问体验

#### 伪静态缺点

伪静态网页并不能提升网站的访问效率，理论上讲还降低了网站性能

需求提出方：产品、需求

**伪静态网页小结：**

* 利用rewrite技术将动态网页伪装成静态网页（URL改写）
* 便于搜索引擎搜录，提升用户访问量及体验
* 访问性能没有提升，并且转换成伪静态会消耗资源，因此性能反而会下降
* 尽可能地将动态网页转换成真正的静态页面
* 并发量不是很大或动态更新过于频繁时，用rewrite实现伪静态也可
* 伪静态网页的实现过程，一般由产品运营提出需求，开发和运维共同实现

# 网站流量度量术语

## IP-独立IP数

同一个客户端一天（24小时）访问N次也只被计数一次,因为客户端访问的IP地址没变。

## PV-访问量-页面浏览

页面浏览或点击量（刷新也算）。点击浏览十次就计算十次

## UV-独立访客

同一个客户端（PC）访问网站被计算为一个访客。一天（24小时内）访问多次也只记为一位访客，以客户端Cookie等技术作为统计依据，存在偏差。

## 企业网站对IP PV UV的度量

### IP的度量

分析所有web服务器的访问日志信息，对IP地址段去重后技术。

在网站的每一个页面结尾，嵌入JS等统计程序代码，待用户加载网页后，IP即传给统计IP的服务器，这种方法一般被第三方统计公司开发日志分析程序时使用

第三方大家比较信任的统计工具，例如谷歌的GA

### PV的度量

分析web服务的访问日志，只计算HTML，PHP等页面数量

在网站的每一个页面结尾，嵌入JS统计代码，带用户加载后，访问数量即传给统计PV的服务器。

用第三方大家比较信任的统计工具 谷歌的GA

### UV的度量

#### 通过客户端HTTP请求报文分析

一个客户端会多次请求网站服务器，每次HTTP请求都会携带客户端自身的大量信息，比如IP地址，请求发出时间，浏览器版本，操作系统版本。共同特征的定义是由服务器决定的

#### 通过Cookie鉴别

当客户端第一次访问某个网站服务器时，网站服务器会给客户端的计算机发出一个Cookie,通常放在这个客户端计算机的C盘中，Cookie中会分配一个独一无二的编号，其中记录一些访问服务器的信息，访问时间，访问那些页面等等。当你下一次再访问这个服务器时，服务器可以直接从计算机中找到上次放进去的Cookie文件，并且对其进行更新，但那个独一无二的编号不会改变。如果在一定时间内，服务器发现2个访者对应编号是同一个编号，那么就可以认为是同一个访者。使用Cookie的方法要比分析客户端HTTP请求头部信息更为准确些。

#### IP PV UV的区别

简单理解：

比如：网吧10个人访问[www.baidu.com网站，每个人平均访问5](http://www.baidu.com网站，每个人平均访问5)次 。

计算结果为：（独立IP）IP 1 （浏览量）PV50 （独立访客）UV 10

结论：一个网站的独立IP数量要比网站实际访问的PV数量小得多，通常情况下网站的UV数也会大于独立IP数。

PV数高说明访问的页面数多，但是不一定就代表来访者多；但PV数一定与来访者的数量成正比，不过，PV并不直接决定页面的真实来访者。比如在访问某网站时，一个人也可以通过不断地刷新页面，制造出非常高的PV数。PV数多，用户访问网站页面的总数量多，通常服务器的压力会大一些。

## 并发连接

网站并发连接：

简单理解是：网站服务器在单位时间内能够处理的最大连接数

其他并发连接：

Qps每秒查询率

用于衡量一个特定查询服务器在规定时间内所处理流量多少的标准。运维工作中，DNS系统以数据库服务的查询性能经常用每秒查询率来衡量。

IOPS每秒读写操I/O作的次数

多用于数据库场合，衡量随机访问的性能，存储端的IOPS性能和主机端的I/O是不同的，IOPS指存储每秒可接受多少次主机发出的访问，主机一次I/O需要多少次访问存储才可以完成。例如：主机写入一个小数据块，也要经过“发送写入请求、写入数据、收到写入确认”三个步骤，也就是3个存储端访问。

<http://alexa.chinaz.com/www.51cto.com>

* CDN占90%
* 带宽计算：先计算出一个用户访问业务需要多少带宽
* 服务器量计算：先压力测试除服务器可以承受多少用户

#### 常见面试题：

1. 请问你如何理解网站并发？

解答：网站服务器在单位时间内能够处理的最大连接数量

你们公司网站访问量是多少？是怎么计算？

解答：思路

CDN占90%

带宽计算：先计算出一个用户访问业务需要多少带宽

服务器量计算：先压力测试除服务器可以承受多少用户

关于网站访问指标计算：

运维部门的日志分析

开发在页面嵌入JS程序（用于统计、收集、分析）

运营市场通过第三方公司提供的工具统计例如谷歌的GA

## www服务软件

### 常用静态服务软件：

Apche

Nginx

Lighttpd

### 常用动态服务软件：

Resin

Tomcat

PHP(FastCGI) 动态网页语音PHP程序的解析容器 -FastCGI-5.5

IIS

### 考试题：

1. 请描述DNS系统的解析原理

1）客户端浏览在浏览器中输入[www.baidu.com](http://www.baidu.com)后回车，浏览器先会查找本机的DNS缓存和Hosts文件有没有域名对应的IP信息。[有的话使用IP访问百度服务器]

2) 客户端会向ISP的DNS服务器查询[www.baidu.com的IP](http://www.baidu.com的IP)信息，LDNS有的话直接将IP给客户端浏览进行访问

3）LDNS服务器会向(13台)根DNS服务器查询[www.baidu.com对应的IP](http://www.baidu.com对应的IP)信息，根DNS服务器中的.com顶级DNS服务器查找到有关baidu.com的记录。并反馈给LDNS(缓存)

4）LDNS根据.com顶级DNS提供的信息找到baidu.com授权服务器（购买，配置好的），baidu.com授权服务器将查找到的IP信息反馈给LDNS（缓存信息以便下次访问使用）

5）LDNS将[www.baidu.com的IP](http://www.baidu.com的IP)信息发送给客户端机器

6）客户端浏览器通过域名解析到的IP以及默认端口信息，首先与[www.baidu.com服务器建立一个TCP](http://www.baidu.com服务器建立一个TCP)连接（经过三次握手）后并发送一个HTTP请求报文（请求报文知识点）

7）请求报文经过负载均衡器后，传送到Web服务器，Web服务器响应报文并根据读取报文请求内容，向后查找[NFS-缓存-存储-Mysql]到请求的内容。

8）Web服务器向客户端返回一条HTTP响应报文[请求的内容数据]

9）客户端浏览收到响应报文-页面呈现完成后，Web服务器关闭HTTP连接，断开TCP连接（四次挥手）

1. 请描述HTTP协议的工作原理

HTTP协议通信时，客户端与服务端之间先建立TCP连接，也就是三次握手的过程。之后客户端向服务器发送一条HTTP请求报文[请求报文知识点]；服务器端收到请求报文，响应读取并返回一条HTTP响应报文；服务器断开HTTP连接、TCP连接（四次挥手）。

1. 请问你公司的网站访问量是多少？
2. 请说出HTTP状态码代表的意义 200 301 403 404 500 502 504

200 服务器返回的状态码代表业务响应正常

301 永久跳转代码，网站更改后可配置访问原地址时自动跳转到新的URL

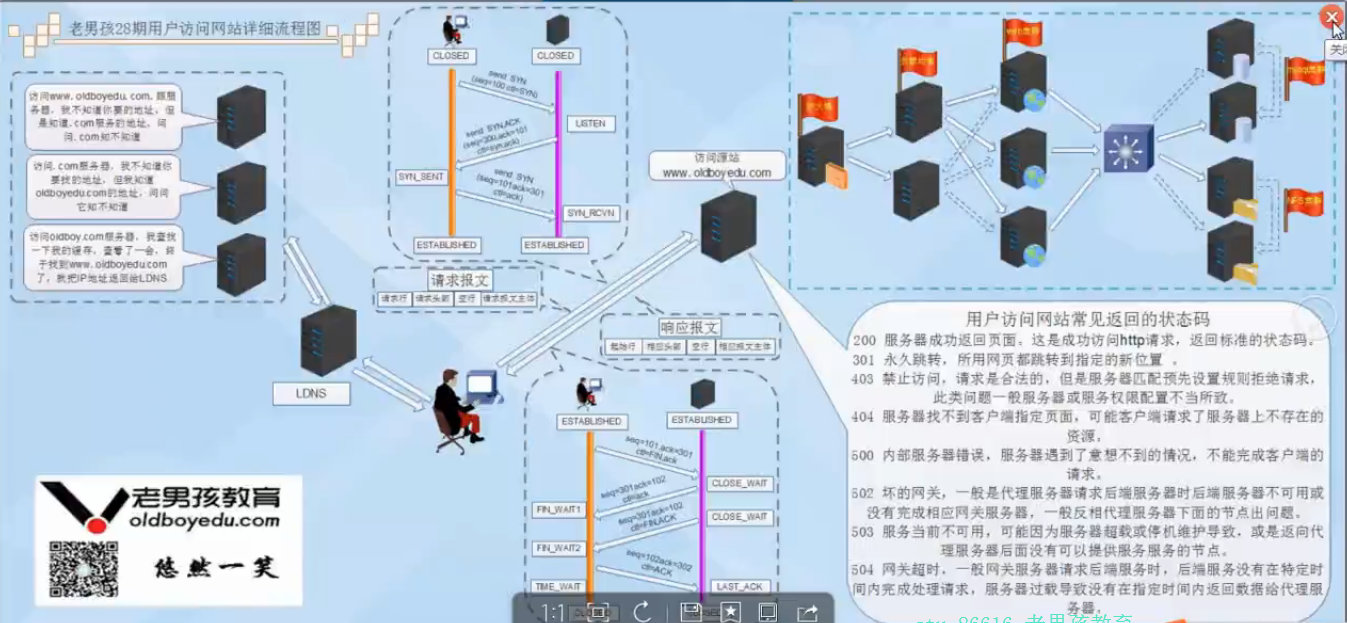
403 禁止访问，匹配到预先设置的规则拒绝提供服务，一般是权限设置造成

404 用户访问的页面文件不存在，被移动或权限更迭所致

500 内部服务错误设置所致

502 坏的网关 代理后端服务不可用或没有响应代理的请求所致

504 网关超时 后端服务器没有在特定时间内处理完请求



<https://edu.csdn.net/course/play/2278/35535>

需要巩固项：

集群架构、TCP三次握手、四次挥手、11种连接状态

