Learn How Hyper Text Transfer Protocol works

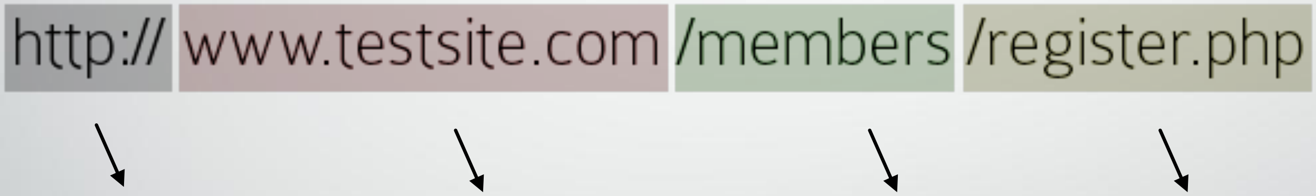
LOVE With HTTP

Author : Gautam Kumawat

Table of Contents

Introduction to HTTP

Well, you daily surf Internet using your favorite Web Browser (Client) to access Web-Sites, Mails, and Social Sites etc.. But how all these stuffs works and communicate with each others ? Every Web-Site has an unique name called **Domain Name** also referred as (Uniform Resource Locator) URL, And every **URL** starts with four letter (string) (http://) **HTTP** (Hyper Text Transfer Protocol).



http:// www.testsite.com /members /register.php

Protocol World Wide Web Domain Directory Path Name

Well Protocol means a set of rule and functions. Sometimes you might see (**https://**) (S) stands for secure : It means the communication is secure, *Sometimes ports are also included in URL (http://localhost:8080/)*

Introduction to HTTP

The HTTP (Hyper Text Transfer Protocol) is an application protocol used to request web server for web pages, applications, scripts, medias and Dynamic pages. HTTP is responsible for responding client's request and to delivers web pages to client's browser. HTTP is simple text based protocol. HTTP is only the protocol for communication between client and server.

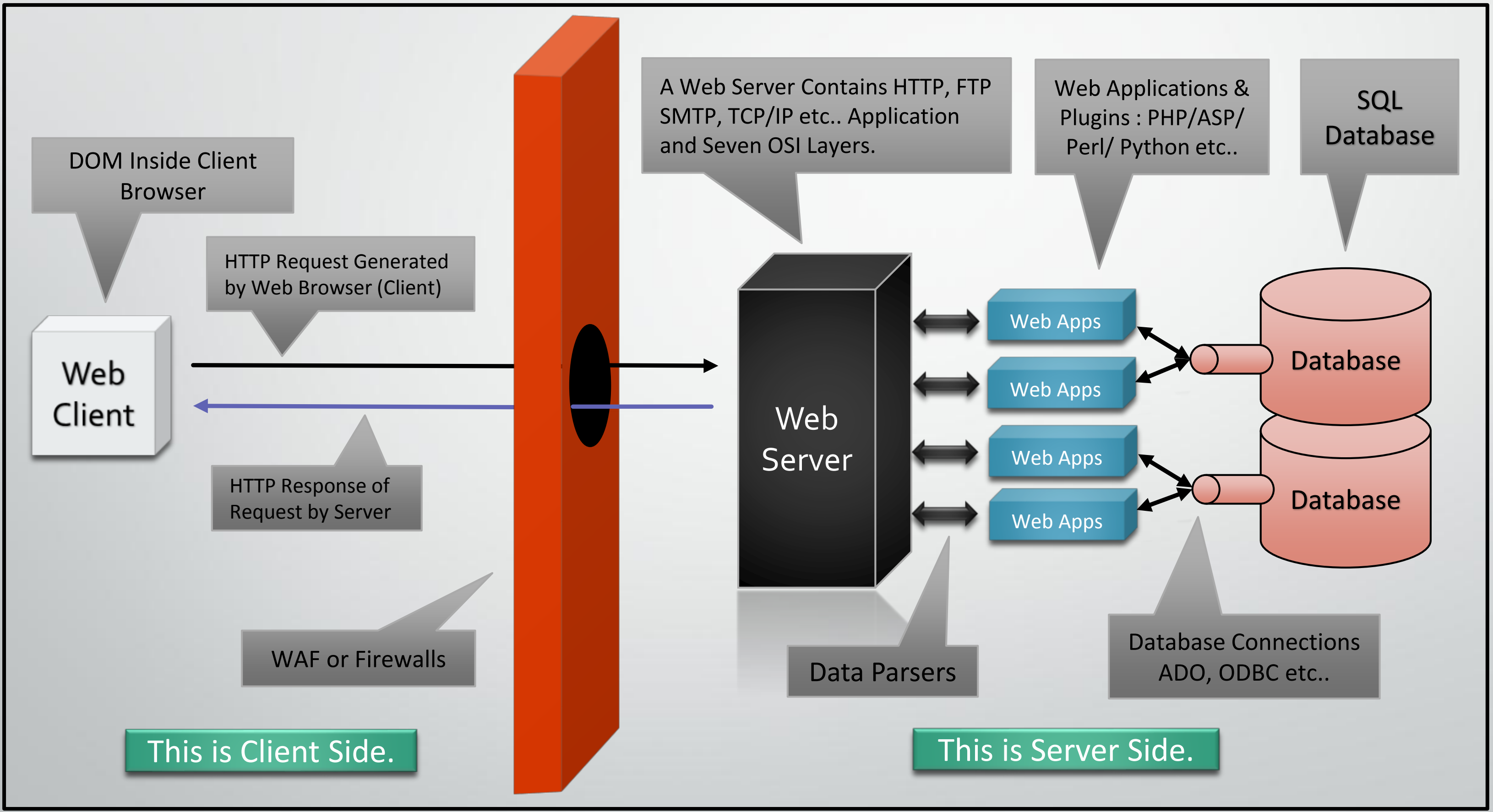
Suppose you want to access Google site on your browser. You type URL in your browser and hit enter, at that moment your browser generates HTTP request including Domain Name, Path, Method, Directory and Requested File and send it to server, The Request goes through the Internet to the server and server responds you back with HTML, JavaScript, Images and CSS files that your browser can understand and finally it display's on your screen.

The Most Important : HTTP is stateless protocol. It doesn't manage State.

History and About

- HTTP is Text based User-Interface Protocol
- HTTP works on Request / Response Mechanism
- Sir. Tim Berners Lee invented HTTP 0.9 in 1989
- Latest version of HTTP is 1.1 (Upcoming HTTP 2.0)
- HTTP was never created for online Banking, Cloud Computing etc.
- HTTP Request Methods (*GET, POST, HEAD, PUT, Trace etc..*)
- HTTP delivers Hypermedia contents to client's browser
- Default port for HTTP is Port 80 and 443 for HTTPS.
- HTTPS is either Protected by SSL or TLS for secure communication
- HTTP is responsible for responding client's request and deliver Web Pages

Web Architecture and Flow



DOM Inside Client Browser

HTTP Request Generated by Web Browser (Client)

HTTP Response of Request by Server

WAF or Firewalls

This is Client Side.

A Web Server Contains HTTP, FTP SMTP, TCP/IP etc.. Application and Seven OSI Layers.

Web Applications & Plugins : PHP/ASP/ Perl/ Python etc..

SQL Database

Data Parsers

Database Connections ADO, ODBC etc..

This is Server Side.

Best viewed on Full Screen

Web Architecture and Flow

Previous page gave you slight knowledge of Web Architecture and Flow. As you saw every communication was carried out by HTTP. The Client requested for a resource on server using URL. Your Browser generated request and sent it to server, And finally server responded with requested resource containing (HTML, CSS and JavaScript) file. Every request/response pass through a Firewall or WAF (Web Application Firewall) between client side and server side.

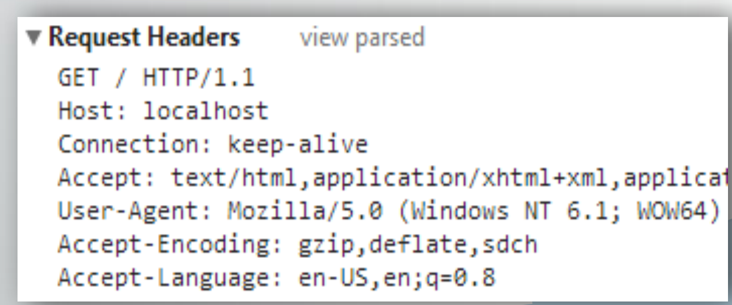
The Client Side : The Client side is also called : User Side. Client's browser saves scripted file (HTML, JS & CSS) inside DOM (Document Object Module) received by Server as requested data and displays on your Screen. Unfortunately Client side validation can be bypassed easily, by tampering HTTP parameter.

The Server Side : The Server side is everything that is responsible for managing Database, Web Applications Flow, Application Plugins (PHP, ASP, Python, C etc..) and Communication between Client and Server. A Web server has different mechanisms for Authorization, Authentication and Protocols for communication like : FTP, SMTP, SSL, HTTP, etc…

Web Servers : Apache, IIS etc.. (Take a good look once again on previous page diagram)

How HTTP Works ?

So, Here comes the main question. Let’s clear this out friends. Your favorite browser is a HTTP Communication tool, Well imagine you’re using Google Chrome to access a site (http://localhost.com) when you type URL and hit enter, it generates request and send it to localhost.com server, Now server reads HTTP request and response requested content. Check out below images in which Request/Response mechanism are shown respectively.

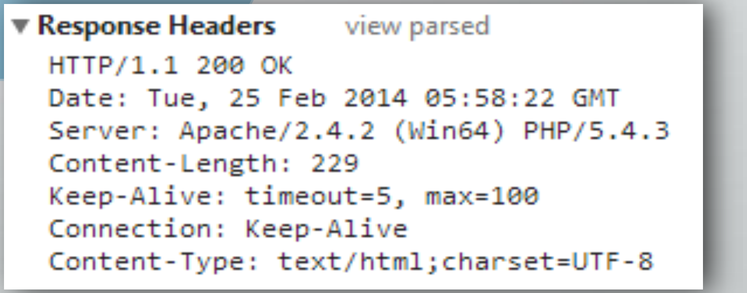


```
GET / HTTP/1.1
Host: localhost
Connection: keep-alive
Accept: text/html,application/xhtml+xml,applicat
User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64)
Accept-Encoding: gzip,deflate,sdch
Accept-Language: en-US,en;q=0.8
```

Request generated by Browser containing : Method, Host, Connection, User-Agent etc..

INTERNET

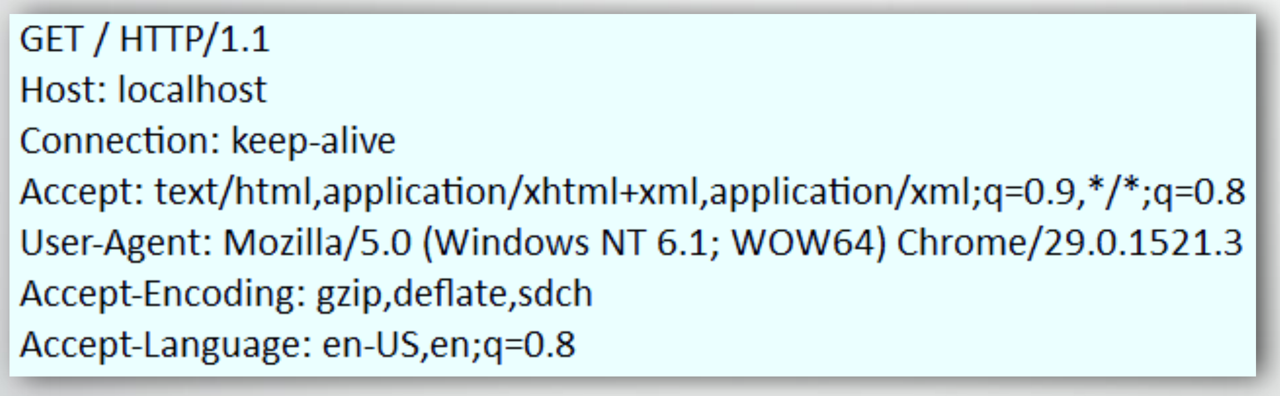
Response send by Server. Containing : Response Code, Date, Server Info, Content and Connection, Type etc..



```
HTTP/1.1 200 OK
Date: Tue, 25 Feb 2014 05:58:22 GMT
Server: Apache/2.4.2 (Win64) PHP/5.4.3
Content-Length: 229
Keep-Alive: timeout=5, max=100
Connection: Keep-Alive
Content-Type: text/html;charset=UTF-8
```

How HTTP Works ?

As you saw in previous page diagram - flow of HTTP, Now it’s time to explore HTTP request headers. Method, Host, User-Agent, Connection etc.. Are all headers containing information about client and request.

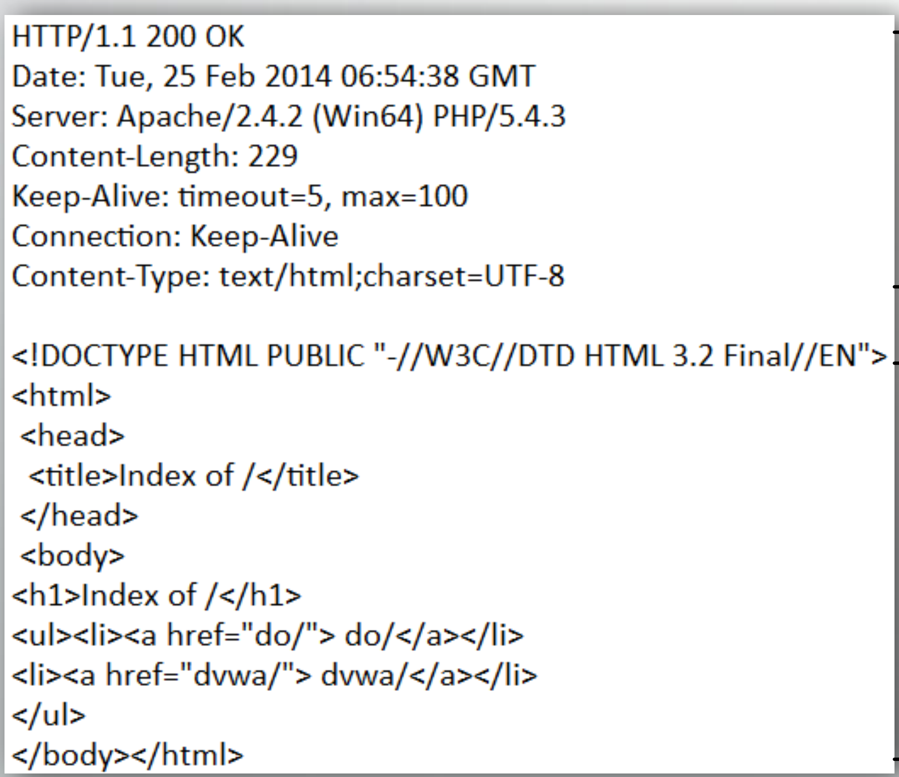


```
GET / HTTP/1.1
Host: localhost
Connection: keep-alive
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64) Chrome/29.0.1521.3
Accept-Encoding: gzip,deflate,sdch
Accept-Language: en-US,en;q=0.8
```

The **Method** header always comes first, **GET** and **POST** are common methods used to request web server. Just after request method, It shows HTTP version. The current version of HTTP is 1.1. The Second line contain **Host or Domain**. It is also called as the “Request receiver”. Third Line : Connection should be alive. Fourth line describes what language the client’s browser can read and accept. User-Agent is client’s browser name. And the last two lines : Define what language and encoding can be accepted. Except first two headers, all are default by browser, and after response it add Cookies.

How HTTP Works ?

There is nothing hard about learning how HTTP works, all you need to put little interested. So now we'll explore and learn Response header. After receiving request from client, Server always responds with various Response Code & other headers. The response header contains most important part called : Body. The Body part contains HTML, JavaScript, CSS and Other medias that your browser can read and display.



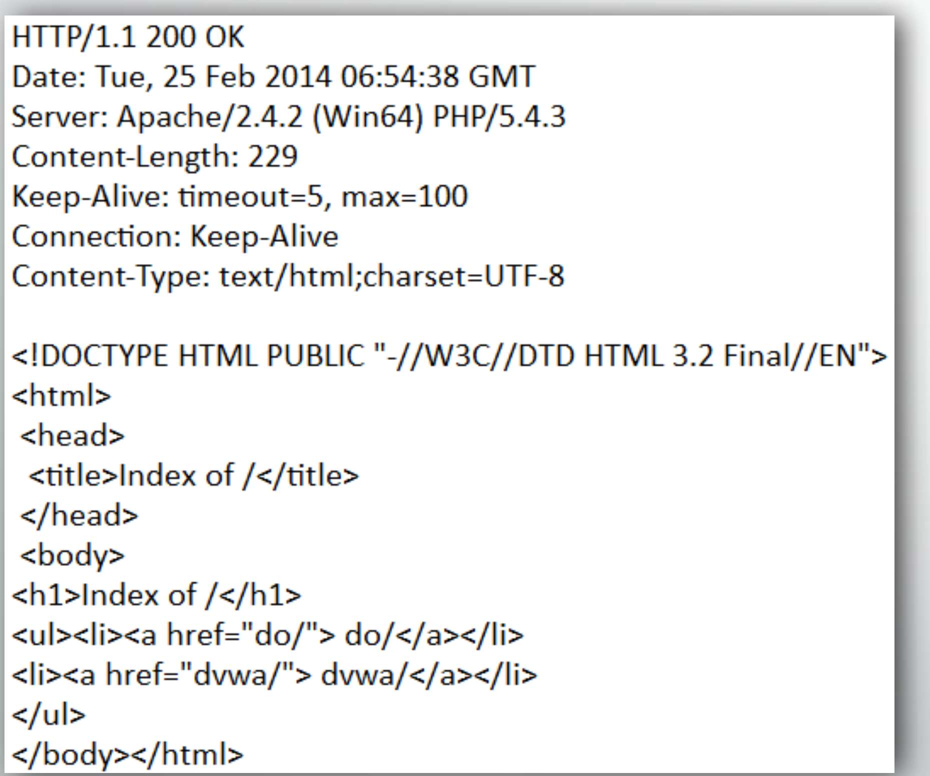
```
HTTP/1.1 200 OK
Date: Tue, 25 Feb 2014 06:54:38 GMT
Server: Apache/2.4.2 (Win64) PHP/5.4.3
Content-Length: 229
Keep-Alive: timeout=5, max=100
Connection: Keep-Alive
Content-Type: text/html;charset=UTF-8

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<html>
 <head>
  <title>Index of /</title>
 </head>
 <body>
<h1>Index of /</h1>
<ul><li><a href="do/"> do/</a></li>
<li><a href="dvwa/"> dvwa/</a></li>
</ul>
</body></html>
```

HTTP Response Header Containing : Response Code, Date, and Server Info etc...

This is Response Body part, Containing HTML, JS, CSS & Other Media Contents like : Video, Images Etc..

HTTP Response Message

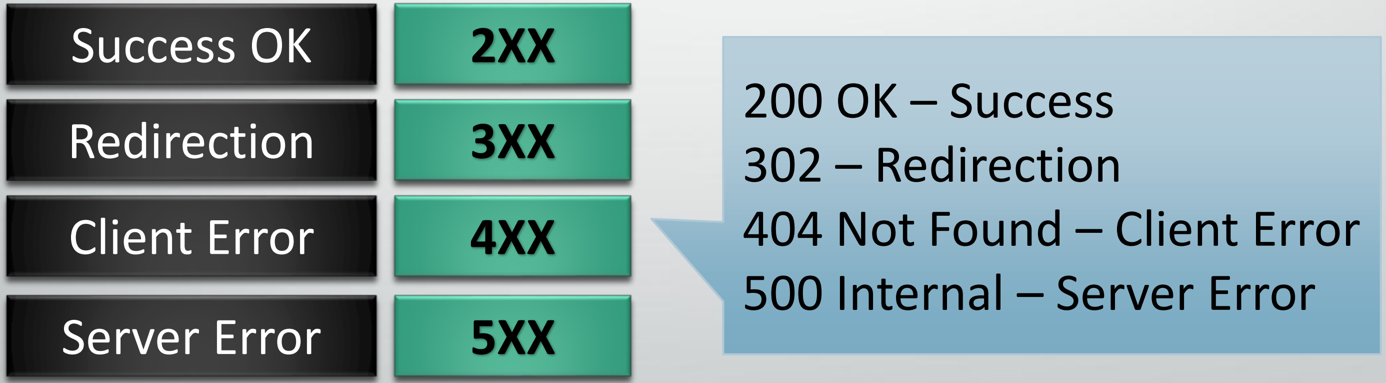
In Response Message : There are two part. The Above part is HTTP Response Headers and Below part is called The Body Part and the blank line Separating both content is called CRLF. **HTTP Response Header** : The First line represents HTTP version and Response code! (We'll learn a lot about Response code in next page.) After that it shows Date and Server Information. The fourth line counts content-length. Keep-Alive and Connection are default by server. Content-Type represents what type of content – Server delivered to Client. Like text/html and charset = UTF-8.



```
HTTP/1.1 200 OK
Date: Tue, 25 Feb 2014 06:54:38 GMT
Server: Apache/2.4.2 (Win64) PHP/5.4.3
Content-Length: 229
Keep-Alive: timeout=5, max=100
Connection: Keep-Alive
Content-Type: text/html;charset=UTF-8

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<html>
 <head>
  <title>Index of /</title>
 </head>
 <body>
<h1>Index of /</h1>
<ul><li><a href="do/"> do/</a></li>
<li><a href="dvwa/"> dvwa/</a></li>
</ul>
</body></html>
```

HTTP Response Code

HTTP Response Code represents the reaction and response of requested URL/Directory. Mostly you'll get 200 OK (It means your request has been completed successfully and you can receive requested file in your Browser) I guess you all are familiar with one of the most common and popular HTTP Response code : 404 Not Found. Assume you type www.facebook.com you'll get 200 OK Response code, but if you try to access someone else private pictures or locked data you'll get client error 404 Not Found.



Success OK	2XX
Redirection	3XX
Client Error	4XX
Server Error	5XX

200 OK – Success
302 – Redirection
404 Not Found – Client Error
500 Internal – Server Error

HTTP Methods

Well, you know the two most common HTTP Methods **GET** and **POST**. But do you know that there are some more Methods. Let's get hands on it.

Request to get a resource from the server	GET
Send data (in the body) to the server	POST
Request to Get only Header not Body	HEAD
Request to Get final request. (for Proxies)	TRACE
store request body on server	PUT
Delete a resource on the server	DELETE
Request to get a list of methods supported by server	OPTION