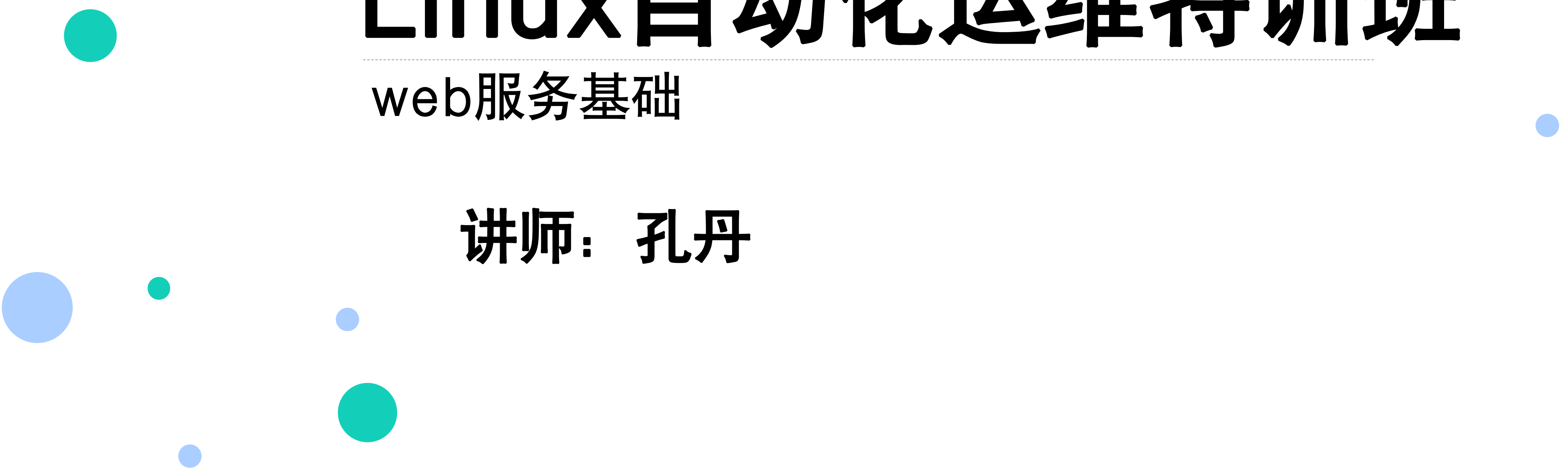




Linux自动化运维特训班

web服务基础

讲师：孔丹



大纲

- Internet与万维网
- Web及其工作原理
- HTTP服务介绍
- HTTP资源

Internet与万维网

□ Internet的产生

阿帕网 (ARPANet) , 萌芽

一个没有中央控制节点的计算机网络, 1969

TCP/IP协议研究成功 , 计算机互联的主要障碍被解决 , 1974

□ 互联网的诞生

ARPANet被分成两部分, 一部分作为军用, 称为MILnet, 另一部分作为民用, ARPANet开始从一个实验型网络向实用型网络转变, 从而成为全球Internet正式诞生的标志。1982

□ 发展阶段

Internet Internet的提出 的实用发展阶段 、研究和试验阶段

Web 商业化发展阶段 2.0

Web3.0, 未来的Web。

Web及其工作原理

□ 什么是Web ?

互联网 = {所有上网的主机}

各种服务器, 如: Web服务, FTP服务器, Email服务器, NNTP服务器, 应用服务器, etc.

客户机

WWW = {Web服务器, Web客户机}

□ Web服务器

网络操作系统

Web服务组件, 如Apache/Tomcat, IIS

□ Web客户机

Web浏览器, 如IE, Chrome, Firefox

Web及其工作原理

□ 构成 ?

三层结构

客户端Web浏览器

Web服务器

数据库服务器

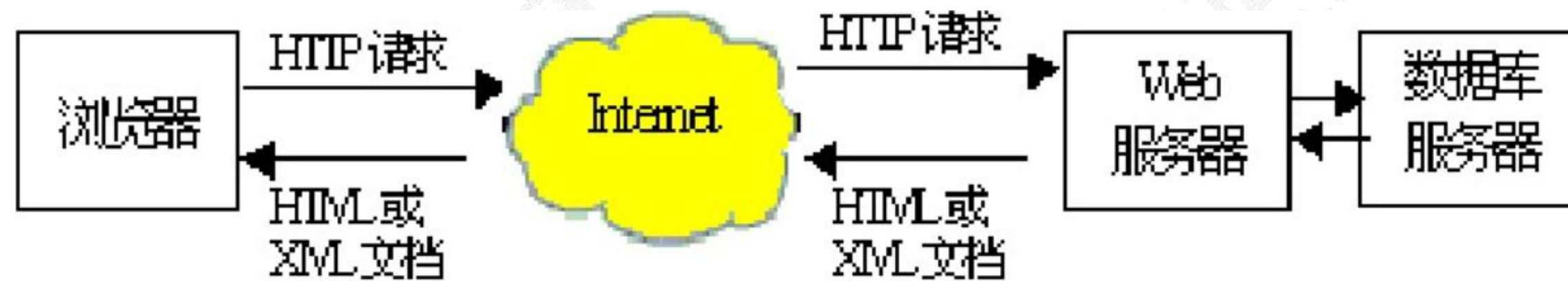
互联网Web服务器

□ 优点?

具有良好的开发性，利用单一的访问点，用户可以在任何地点使用系统，用户可以跨平台以相同的浏览器访问系统
跨越了局域网的限制，将应用分布到互联网

Web及其工作原理

□ web工作原理 ?



- 1) 在浏览器地址栏中，用户输入要访问的网页网址URL
- 2) Web服务器根据URL中指定的网址、路径和网页文件，调出相应的HTML、XML文档或Jsp、Asp文件
- 3) 根据文档类型，Web服务器决定是否执行文档中的服务器脚本程序，还是直接将网页文件传送到客户端。
- 4) 浏览器显示下载的文档，即网页

HTTP服务介绍

□ 用户访问网站流程

a 进行域名信息的DNS解析

dig +trace 获得www.baidu.com ip地址信息

b 进行与网站服务器建立连接, tcp三次握手过程

(syn ack SEQ ACK/状态转换 closed listen syn_sent
syn_rcvd established)

c 和网站服务器数据传输过程 (http协议原理过程-http请求过程)

查看方法curl -v/wget --debug

d 和网站服务器数据传输过程 (http协议原理过程-http响应过程)

e 进行与网站服务器断开连接, tcp四次挥手过程

(fin ack SEQ ACK/状态转换 fin_wait1 fin_wait2 close_wait
last_ack time_wait closing)

HTTP服务介绍

□DNS解析原理

01. 查看本地主机缓存和hosts文件有没有域名与IP对应记录
02. 本地主机向local DNS服务器（网卡上指定配置的DNS服务器）发出请求（递归查询）
03. LOCAL DNS服务器接收到客户端主机请求，查询本地缓存信息
04. LOCAL DNS服务向根域名服务器请求，向顶级域名服务器请求，向二级域名服务器请求（迭代查询）
05. 获取到域名与IP地址的对应关系，是通过授权DNS服务中获取得到（得到的信息称为A记录）
LOCAL DNS服务进行缓存A记录信息，将A记录信息发送给客户端主机
06. 客户端主机接收到A记录信息，进行本地缓存，根据A记录中记录IP地址信息，进行网络访问

HTTP服务介绍

□ 什么是HTTP协议？

HTTP协议，全称HyperText Transfer Protocol,中文名为超文本传输协议，是互联网中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协议最初的目的就是提供一种发布和接收HTML（一种页面标记语言）页面的方法。

HTTP协议是互联网上常用的通信协议之一。它有很多的应用，但最流行的就是用于Web浏览器和Web服务器之间的通信，即WWW应用或称Web应用。

www ,全称World Wide Web ,常称为Web ,中文译为"万维网：它是目前互联网上最受用户欢迎的信息服务形式。 HTTP协议的WWW服务应用的默认端口为80,另外一个加密的WWW服务应用https的默认端口为443,主要用于网银、支付等和钱相关的业务。当今，HTTP服务、WWW服务、Web服务三者的概念已经混淆了，都是指当下最常见的网站服务应用。

HTTP服务介绍

□ HTTP协议版本

HTTP/1.0 规定浏览器和服务端只保持短暂的连接（TCP短连接）

HTTP/1.1 支持持久连接，在一个TCP连接上可以传送多个HTTP请求和响应，减少了建立和关闭连接的消耗和时间延迟（TCP长连接）

linux系统中参看报文的方法

curl -v ---显示访问域名请求报文与响应报文信息

curl www.baidu.com -v

> 表示http请求报文 <表示http的响应报文

wget --debug www.baidu.com

HTTP服务介绍

□ 响应http请求方法

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首部
POST	将客户端的数据提交到服务器，例：注册表单
PUT	用从客户端向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-URI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服务介绍

□ 响应http请求方法

HTTP请求报文格式		
请求报文格式	报文信息	说明
请求行	GET / HTTP/1.1	请求的起始行（请求什么信息；协议版本）
请求头	User-Agent: curl/7.19.7 (x86_64-redhat-linux-gnu) libcurl/7.19.7 NSS/3.27.1 zlib/1.2.3 libidn/1.18 libssh2/1.4.2 Host: www.baidu.com Accept: */*	请求的头部信息（客户端有关信息说明）
空行		空行信息（隔离上下文，说明请求结束）
请求报文主体		请求主体内容（只有POST 方法可以提交主体）

HTTP服务介绍

□ 请求头信息说明

请求头内容	说明
Host	接受请求的服务器地址， 可以是IP:端口号， 也可以是域名
User-Agent	发送请求的应用程序名称
Connection	指定与连接相关的属性， 如Connection:Keep-Alive
Accept-Charset	通知服务端可以发送的编码格式
Accept-Encoding	通知服务端可以发送的数据压缩格式

HTTP服务介绍

□ http响应过程

HTTP响应报文格式		
请求报文格式	报文信息	说明
起始行（响应行）	HTTP/1.1 200 OK	响应的起始行（响应状态码 状态信息）
响应头部	Server: bfe/1.0.8.18Date: Mon, 23 Oct 2017 02:50:36 GMTContent-Type: text/htmlContent-Length: 2381Last-Modified: Mon, 23 Jan 2017 13:27:32 GMTConnection: Keep-AliveETag: "588604c4-94d"Cache-Control: private, no-cache, no-store, proxy-revalidate, no-transformPragma: no-cacheSet-Cookie: BDORZ=27315; max-age=86400; domain=.baidu.com; path=/Accept-Ranges: bytes	响应的头部信息（服务器有关信息介绍）
空行		空行信息（隔离上下文，说明响应头部结束）
响应报文主体	<!DOCTYPE html>	响应主体的内容（请求后进行响应的返回的内容 信息）

HTTP服务介绍

□ 响应头信息说明

响应头	说明
Server	服务器应用程序软件的名称和版本
Content-Type	响应正文的类型（是图片还是二进制字符串）
Content-Length	响应正文长度
Content-Charset	响应正文使用的编码
Content-Encoding	响应正文使用的数据压缩格式
Content-Language	响应正文使用的语言

HTTP服务介绍

□ HTTP协议状态码

HTTP状态码 (HTTP Status Code)是用来表示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HTTP请求时，Web服务器都会返回一个状态响应代码。这个状态码是一个三位数字代码，作用是告知Web客户端此次请求是否成功，或者是否要采取其他的动作方式。

状态码范围	作用描述
100 - 199	用于指定客户端相应的某些动作
200 - 299	用于表示请求成功
300 - 399	用于已经移动的文件，并且常被包含在定位头信息中指定新的地址系信息
400 - 499	用于指出客户端的错误
500 - 599	用于指出服务端的错误

HTTP服务介绍

□ HTTP协议状态码

常见的状态码

代码	代码说明
301 Moved Permanently (永久跳转)	请求的网页已永久跳转到新位置
403 Forbidden (禁止访问)	服务器拒绝请求
404 - Not Found,	服务器找不到请求的页面
500	内部服务器错误
502	坏的网关，一般是网关服务器请求后端服务时，后端服务没有按照http协议正确返回结果
503 - Service Unavailable (服务当前不可用)	可能因为超载或停机维护
504 - Gateway Timeout (网关超时)	一般是网关服务器请求后端服务时，后端服务没有在特定的时间内完成服务。

HTTP资源

□ 媒体资源

互联网上的数据有很多不同的类型，Web服务器会把通过Web传输的每个对象都打上MIME类型（即MIME type）的数据格式标签。最初设计MIME(Multipurpose Internet Mail Extension,多用途因特网邮件扩展)是为了解决在不同的电子邮件系统之间搬移报文时存在的问题.MIME在电子邮件系统中工作得非常好。后来，HTTP也支持了这个功能，用它来描述数据并标记不同的数据内容类型。

当Web服务器响应HTTP请求时，会为每一个HTTP对象数据加一个MIME类型，当Web浏览器获取到服务器返回的对象时，会去查看相关的MIME类型，并进行相应处理。

MIME类型存在与HTTP响应报文的响应头部信息里，它是一种文本标记，表示一种主要的对象类型和一个特定的子类型，中间由一条斜杠来分割。

HTTP资源

□ 媒体资源

MIME媒体资源

web服务可以处理的资源类型，根据扩展名进行识别 (html css mp4)

url: URL, 全称为Uniform Resource Location, 中文翻译为统一资源定位符

uri: URI, 全称为Uniform Resource Identifier, 中文翻译为统一资源标识符

URN:统一资源名称 (Uniform Resource Name, URN), 是URI两种形式之一。

例如

网址: <http://reg.jd.com/reg/person?ReturnUrl=https%3A//www.jd.com/>

<http://reg.jd.com>

--- 称为url

</reg/person?ReturnUrl=https%3A//www.jd.com/>

--- 称为uri

HTTP资源

□ 网页资源种类

> 静态网页资源

请求响应信息，发给客户端进行处理，由浏览器进行解析，显示的页面称为静态页面

在网站设计中，纯粹HTML格式的网页（可以包含图片、视频JS（前端功能实现）、CSS（样式）等）通常被称为"静态网页"

> 动态网页资源

所谓的动态网页是与静态网页相对而言的，也就是说，动态网页的URL后缀不是.html .htm、.xml、.shtml、.js css 等静态网页的常见扩展名形式，而是.asp、.aspx、.php、.js、.do、.cgi等形式

请求响应信息，发给事务端进行处理，由服务端处理完成，将信息返回给客户端，生成的页面称为动态页面

> 伪静态网页资源

将动态页面伪装成静态页面，便于被搜索引擎收录

将动态页面转换为静态页面的方式(rewrite 正则表达式)

将动态页面内容转换为静态页面，去掉动态页面uri地址中的? &字符，伪装成静态页面uri信息

HTTP资源

□ 网站流量度量术语

- > 独立IP数度量值
独立IP数是指不同IP地址的计算机访问网站时被计算的总次数
- > 页面浏览次数PV
页面浏览量或点击量
- > 独立访客数UV
根据http请求报文：浏览器版本， OS
根据http响应报文： cookie(id)
指当天有多少台电脑访问，一般以COOKIE来统计。

□ 网站并发连接说明

网站服务器在单位时间内能够处理的最大连接数。

HTTP资源

□ web服务软件分类

> 静态

apache、nginx、lighttpd

> 动态

php、asp、jsp

IBM WebSphere、Oracle WebLogic、Tomcat

总结

- Internet与万维网
- Web及其工作原理
- HTTP服务介绍
- HTTP资源



谢谢观看

更多好课，请关注[万门大学APP](#)

