

HTTP协议

协议: 就是两个计算机之间为了能够流畅的进行沟通而设置的一个君子协定. 常见的协议有TCP/IP. SOAP协议, HTTP协议, SMTP协议等等.....

HTTP协议, Hyper Text Transfer Protocol (超文本传输协议) 的缩写,是用于从万维网 (WWW:World Wide Web) 服务器传输超文本到本地浏览器的传送协议. 直白点儿, 就是浏览器和服务器的数据交互遵守的就是HTTP协议.

HTTP协议把一条消息分为三大块内容. 无论是请求还是响应都是三块内容

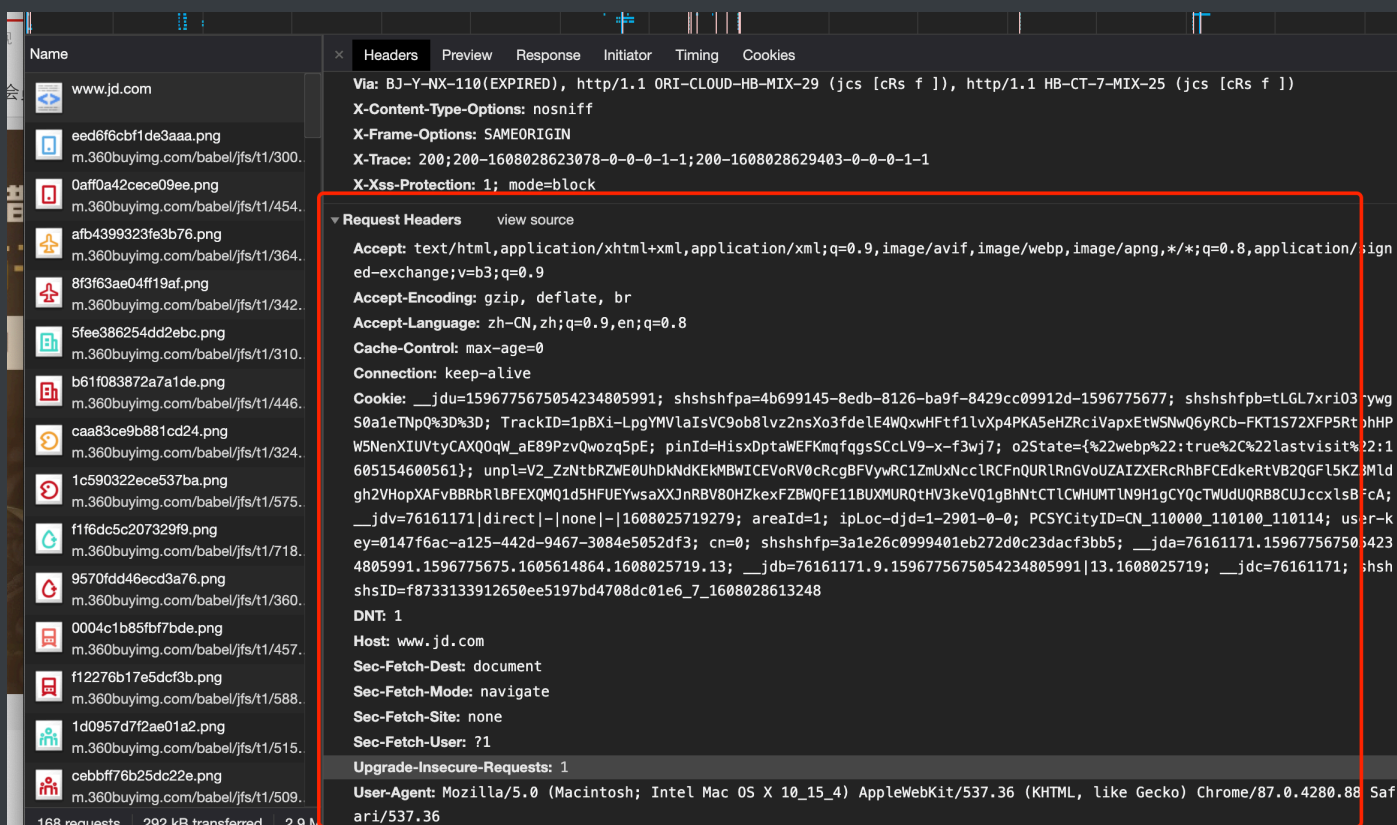
请求:

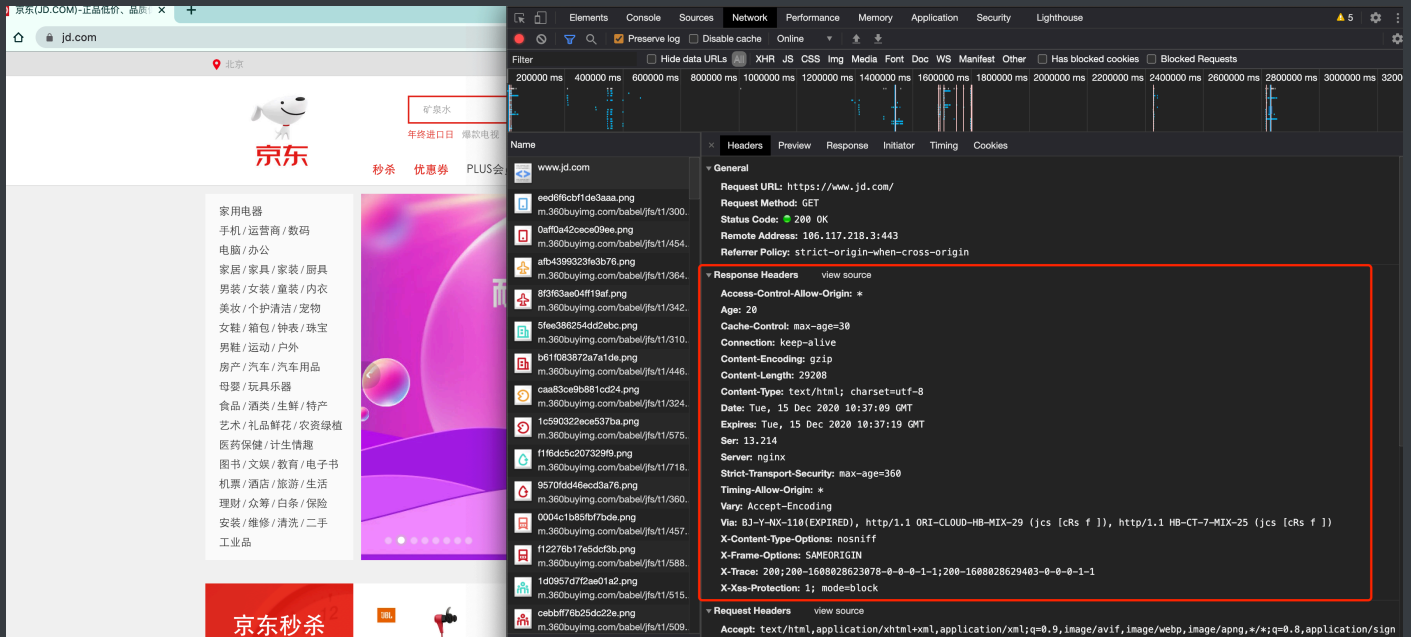
- 1 请求行 -> 请求方式(get/post) 请求url地址 协议
- 2 请求头 -> 放一些服务器要使用的附加信息
- 3
- 4 请求体 -> 一般放一些请求参数

响应:

- 1 状态行 -> 协议 状态码
- 2 响应头 -> 放一些客户端要使用的一些附加信息
- 3
- 4 响应体 -> 服务器返回的真正客户端要用的内容(HTML, json)等

在后面我们写爬虫的时候要格外注意请求头和响应头. 这两个地方一般都隐含着一些比较重要的内容





请求头中最常见的一些重要内容(爬虫需要):

1. User-Agent : 请求载体的身份标识(用啥发送的请求)
2. Referer: 防盗链(这次请求是从哪个页面来的? 反爬会用到)
3. cookie: 本地字符串数据信息(用户登录信息, 反爬的token)

响应头中一些重要的内容:

1. cookie: 本地字符串数据信息(用户登录信息, 反爬的token)
2. 各种神奇的莫名其妙的字符串(这个需要经验了, 一般都是token 字样, 防止各种攻击和反爬)

请求方式:

GET: 显示提交

