

## 1-2 大型网站的特点与设计宗旨

### 大型网站的9大特点

Hello，大家好，本节课我们一起来认知一下大型网站的特点。

说到大型网站，很多人可能都会有一个误区，他们会认为：网站访问量，那就是一个大型网站，其实这只能说对了一半。为什么呢？给大家举个例子就明白了。度旗下的上网导航网站：hao123。这个网站的用户访问量是非常巨大的，那么它是一个大型网站吗？其实不是。为什么呢？

类似 hao123 这样的网站，它们都是提供一个静态网页（可以通过模板技术生成静态页面），用户点击是跳转到其他的网站，它们是根据用户的访问需求提供了网站分类推荐服务。这些网站本身是不具有庞大的数据量的。可以说他们和用户的交互都是单向的，网站通过自身编辑内容对用户访问内容，用户只需要点击相当于 Web1.0 时代的小型网站。

那么我们再来看一下 淘宝 和 京东，这两个网站无疑都是大型网站，首先他们都具有庞大的访问量，不论是 淘宝双11 或是 京东618，用户的请求量都是十分庞大，不仅仅访问量，网站的数据体量也是非常庞大的，拥有海量数据，并且可以支持大数据分析。像每年年底都能通过大数据分析为你提供当年你花了多少钱，预计明年会买什么等等，这些都是大型网站的基本特征。当然，大型互联网网站的特点细分下来有很多，接下来我们一起来看看：

**1. 高并发，流量大：**一个网站从小做到大，业务越来越大，积累的用户也越来越多，那么网站的请求量就势必越来越大。举个例子：

- 百度日均访问达到50亿+（据说已经超过谷歌）
- 腾讯QQ早就突破1亿用户同时在线
- 淘宝双11交易金额达到2000多亿人民币（在此背后，那些工程师付出了多少汗水！？）

**2. 高可用：**大型网站有一个必须达到的标准，就是高可用。什么是高可用，就是假设服务器宕机了，这个时候用户就无法访问，我们必须保证网站稳定地向用户不间断的服务，哪怕有服务器节点宕机了，这个时候备用机也会替换当前不可用的服务器节点，确保服务正常访问。

**3. 大数据：**通过大量服务器以及存储系统管理海量数据，比如阿里在全国各地有上千甚至上万台服务器节点。数据多了，那么就可以做大数据分析了，比如为每个画像，不论是用户购买习惯，还是用户未来预计会购买的商品，都能做到一步步的分析。当然图片数据也是数据，当你上传了图片，网站会通过人工智能来分析，无论是人物图，风景图还是宠物图，都会做到精确的分析。

**4. 敏捷开发，迭代快：**大型网站的用户量必然是会很多的，为了提高用户体验，优化用户各类操作，形形色色的需求都会被产品人员提出，通过头脑风暴、研讨逐出一些非常好用的功能模块，与此同时，开发人员根据需求快速迭代，当然测试与运维也会跟进。一般来说1~2周会快速迭代一次。

**5. 用户体系庞大：**既然一个大型网站的用户量是非常大的，那么这些用户不可能都分布在同一个地区，他们都是分散的个体，分布在全国各地，甚至全球各地。宝和物流，你可以把商品卖到很偏远的农村或者山区，当然这些偏远地也能通过网站把自家的特产，卖向全国各地。

**6. 可持续升级：**任何大型网站都不是一蹴而就的，都是从最初最简单的版本一步一步迭代出来的，很多初创公司都会有这样的误区，一上来就搞一个高大上的架构达到千万。淘宝初期，是从马云老师家里诞生的；苹果公司诞生于车库；谷歌诞生于斯坦福大学宿舍，随后搬入车库。他们刚推出的时候都是非常简单的网站。不会一上来就教大家一个千万级架构的项目，而是会从单体电商项目起步，然后逐步演变，向集群，分布式，微服务发展，当大家真正感受一个大型项目到逐步的从初创不断发展到今天的。

**7. 安全防范：**正所谓树大招风，你越牛叉，就愈会有黑客或者同行来攻击你，所以这对运维来说也有相当大的挑战，我们之前的网站就被黑客攻击勒索比特币，问题是因为服务器漏洞，比如上传图片，没有做到限制文件类型，只在前端判断，却没有在后端判断，导致黑客可以直接攻击Api接口，把PHP或者其他可植入脚本运行，这是非常危险的。

**8. 弹性扩展：**淘宝双11在并发量剧增的时候，可以通过增加服务器来扩容，从而引导剧增的流量；微博里若某明星结婚，流量也会剧增，这个时候服务器都是可以加，来实现动态扩容。不需要的时候适当减少服务器就行。

**9. 吞吐量高，响应速度快：**一个系统数据量一旦大了，那么它的响应速度会比较慢，这会影响用户体验。所以一个大型网站也必须保证用户每次的请求响应是这个网站背后可能有上百上千台服务器，但是对于用户来说是透明的，他是不知道的。

#### 小结

从上面可以看出，大型网站一般具备9个特点：

- 高并发
- 高可用
- 大数据
- 迭代周期短
- 用户量庞大
- 可持续发展
- 安全级别高
- 可弹性、可伸缩