**《网站系统架构》**

|  |  |  |
| --- | --- | --- |
| 安全架构 | 前端架构 | 数据采集与监控 |
| 应用层架构 |
| 服务层架构 |
| 存储层架构 |
| 后台架构 |
| 数据中心机房架构 | | |

目录

[1. 前端架构](#_Toc21086_WPSOffice_Level1) [3](#_Toc21086_WPSOffice_Level1)

[(1) 浏览器](#_Toc7182_WPSOffice_Level2) [3](#_Toc7182_WPSOffice_Level2)

[① 减少HTTP请求](#_Toc7182_WPSOffice_Level3) [3](#_Toc7182_WPSOffice_Level3)

[② 使用浏览器缓存](#_Toc2777_WPSOffice_Level3) [3](#_Toc2777_WPSOffice_Level3)

[③ 启用压缩](#_Toc19570_WPSOffice_Level3) [3](#_Toc19570_WPSOffice_Level3)

[④ CSS放在页面最上面、JavaScript放在页面最下面](#_Toc7144_WPSOffice_Level3) [3](#_Toc7144_WPSOffice_Level3)

[⑤ 减少Cookie传输](#_Toc30255_WPSOffice_Level3) [3](#_Toc30255_WPSOffice_Level3)

[(2) CDN加速](#_Toc2777_WPSOffice_Level2) [3](#_Toc2777_WPSOffice_Level2)

[(3) 反向代理](#_Toc19570_WPSOffice_Level2) [3](#_Toc19570_WPSOffice_Level2)

[(4) DNS](#_Toc7144_WPSOffice_Level2) [4](#_Toc7144_WPSOffice_Level2)

[2. 应用层架构](#_Toc7182_WPSOffice_Level1) [4](#_Toc7182_WPSOffice_Level1)

[(1) 开发框架](#_Toc30255_WPSOffice_Level2) [4](#_Toc30255_WPSOffice_Level2)

[(2) 页面渲染](#_Toc15634_WPSOffice_Level2) [5](#_Toc15634_WPSOffice_Level2)

[(3) 负载均衡](#_Toc6194_WPSOffice_Level2) [5](#_Toc6194_WPSOffice_Level2)

[① 通过负载均衡进行无状态服务的失效转移](#_Toc15634_WPSOffice_Level3) [5](#_Toc15634_WPSOffice_Level3)

[② 负载均衡算法](#_Toc6194_WPSOffice_Level3) [5](#_Toc6194_WPSOffice_Level3)

[ 轮询（Round Robin， RR）](#_Toc9383_WPSOffice_Level3) [5](#_Toc9383_WPSOffice_Level3)

[ 加权轮询（Weighted Round Robin， WRR）](#_Toc24125_WPSOffice_Level3) [5](#_Toc24125_WPSOffice_Level3)

[ 随机（Random）](#_Toc11083_WPSOffice_Level3) [5](#_Toc11083_WPSOffice_Level3)

[ 最少链接（Least Connections）](#_Toc21157_WPSOffice_Level3) [5](#_Toc21157_WPSOffice_Level3)

[ 源地址散列（Source Hashing）](#_Toc32055_WPSOffice_Level3) [5](#_Toc32055_WPSOffice_Level3)

[(4) Session管理](#_Toc9383_WPSOffice_Level2) [5](#_Toc9383_WPSOffice_Level2)

[① Session复制](#_Toc30481_WPSOffice_Level3) [5](#_Toc30481_WPSOffice_Level3)

[② Session绑定](#_Toc19076_WPSOffice_Level3) [6](#_Toc19076_WPSOffice_Level3)

[③ 利用Cookie记录Session](#_Toc17432_WPSOffice_Level3) [6](#_Toc17432_WPSOffice_Level3)

[④ Session服务器](#_Toc5829_WPSOffice_Level3) [7](#_Toc5829_WPSOffice_Level3)

[(5) 动态页面静态化](#_Toc24125_WPSOffice_Level2) [7](#_Toc24125_WPSOffice_Level2)

[(6) 业务拆分](#_Toc11083_WPSOffice_Level2) [7](#_Toc11083_WPSOffice_Level2)

[① 分层](#_Toc20209_WPSOffice_Level3) [7](#_Toc20209_WPSOffice_Level3)

[② 分割](#_Toc2685_WPSOffice_Level3) [7](#_Toc2685_WPSOffice_Level3)

[③ 分布式](#_Toc3482_WPSOffice_Level3) [7](#_Toc3482_WPSOffice_Level3)

[④ 集群](#_Toc32049_WPSOffice_Level3) [7](#_Toc32049_WPSOffice_Level3)

[(7) 虚拟化服务器](#_Toc21157_WPSOffice_Level2) [7](#_Toc21157_WPSOffice_Level2)

[3. 服务层架构](#_Toc2777_WPSOffice_Level1) [7](#_Toc2777_WPSOffice_Level1)

[(1) 分布式消息](#_Toc32055_WPSOffice_Level2) [7](#_Toc32055_WPSOffice_Level2)

[(2) 分布式服务](#_Toc30481_WPSOffice_Level2) [8](#_Toc30481_WPSOffice_Level2)

[(3) 分布式缓存](#_Toc19076_WPSOffice_Level2) [8](#_Toc19076_WPSOffice_Level2)

[(4) 分布式配置](#_Toc17432_WPSOffice_Level2) [8](#_Toc17432_WPSOffice_Level2)

[4. 存储层架构](#_Toc19570_WPSOffice_Level1) [8](#_Toc19570_WPSOffice_Level1)

[(1) 分布式文件](#_Toc5829_WPSOffice_Level2) [8](#_Toc5829_WPSOffice_Level2)

[(2) 关系型数据](#_Toc20209_WPSOffice_Level2) [8](#_Toc20209_WPSOffice_Level2)

[(3) NoSQL数据库](#_Toc2685_WPSOffice_Level2) [8](#_Toc2685_WPSOffice_Level2)

[(4) 数据同步](#_Toc3482_WPSOffice_Level2) [8](#_Toc3482_WPSOffice_Level2)

[5. 后台架构](#_Toc7144_WPSOffice_Level1) [8](#_Toc7144_WPSOffice_Level1)

[(1) 搜索引擎](#_Toc32049_WPSOffice_Level2) [8](#_Toc32049_WPSOffice_Level2)

[(2) 数据仓库](#_Toc13587_WPSOffice_Level2) [8](#_Toc13587_WPSOffice_Level2)

[(3) 推荐系统](#_Toc25897_WPSOffice_Level2) [8](#_Toc25897_WPSOffice_Level2)

[6. 数据采集与监控](#_Toc30255_WPSOffice_Level1) [8](#_Toc30255_WPSOffice_Level1)

[(1) 浏览数据采集](#_Toc19110_WPSOffice_Level2) [8](#_Toc19110_WPSOffice_Level2)

[(2) 服务器业务数据采集](#_Toc542_WPSOffice_Level2) [8](#_Toc542_WPSOffice_Level2)

[(3) 系统监控](#_Toc2141_WPSOffice_Level2) [8](#_Toc2141_WPSOffice_Level2)

[(4) 系统报警](#_Toc608_WPSOffice_Level2) [8](#_Toc608_WPSOffice_Level2)

[7. 安全架构](#_Toc15634_WPSOffice_Level1) [8](#_Toc15634_WPSOffice_Level1)

[(1) Web攻击](#_Toc30211_WPSOffice_Level2) [8](#_Toc30211_WPSOffice_Level2)

[① XSS攻击](#_Toc13587_WPSOffice_Level3) [8](#_Toc13587_WPSOffice_Level3)

[② 注入攻击](#_Toc25897_WPSOffice_Level3) [8](#_Toc25897_WPSOffice_Level3)

[③ CSRF攻击](#_Toc19110_WPSOffice_Level3) [8](#_Toc19110_WPSOffice_Level3)

[④ 其他攻击和漏洞](#_Toc542_WPSOffice_Level3) [8](#_Toc542_WPSOffice_Level3)

[⑤ Web应用防火墙](#_Toc2141_WPSOffice_Level3) [8](#_Toc2141_WPSOffice_Level3)

[⑥ 网站安全漏洞扫描](#_Toc608_WPSOffice_Level3) [9](#_Toc608_WPSOffice_Level3)

[(2) 数据保护](#_Toc7749_WPSOffice_Level2) [9](#_Toc7749_WPSOffice_Level2)

[① 信息加密技术及密钥安全管理](#_Toc30211_WPSOffice_Level3) [9](#_Toc30211_WPSOffice_Level3)

[② 信息过滤与反垃圾](#_Toc7749_WPSOffice_Level3) [9](#_Toc7749_WPSOffice_Level3)

[8. 数据中心机房架构](#_Toc6194_WPSOffice_Level1) [9](#_Toc6194_WPSOffice_Level1)

[(1) 机房架构](#_Toc25455_WPSOffice_Level2) [9](#_Toc25455_WPSOffice_Level2)

[(2) 机柜架构](#_Toc17364_WPSOffice_Level2) [9](#_Toc17364_WPSOffice_Level2)

[(3) 服务器架构](#_Toc5609_WPSOffice_Level2) [9](#_Toc5609_WPSOffice_Level2)

1. **前端架构**
2. **浏览器**
3. **减少HTTP请求**

合并CSS、合并JavaScript、合并图片。

1. **使用浏览器缓存**

将CSS、JavaScript、Logo、图片等静态资源文件缓存到浏览器中，更新的文件可以通过修改文件名来清理缓存拉取新文件。

1. **启用压缩**

启用GZip压缩HTML、CSS、JavaScript文件。

1. **CSS放在页面最上面、JavaScript放在页面最下面**

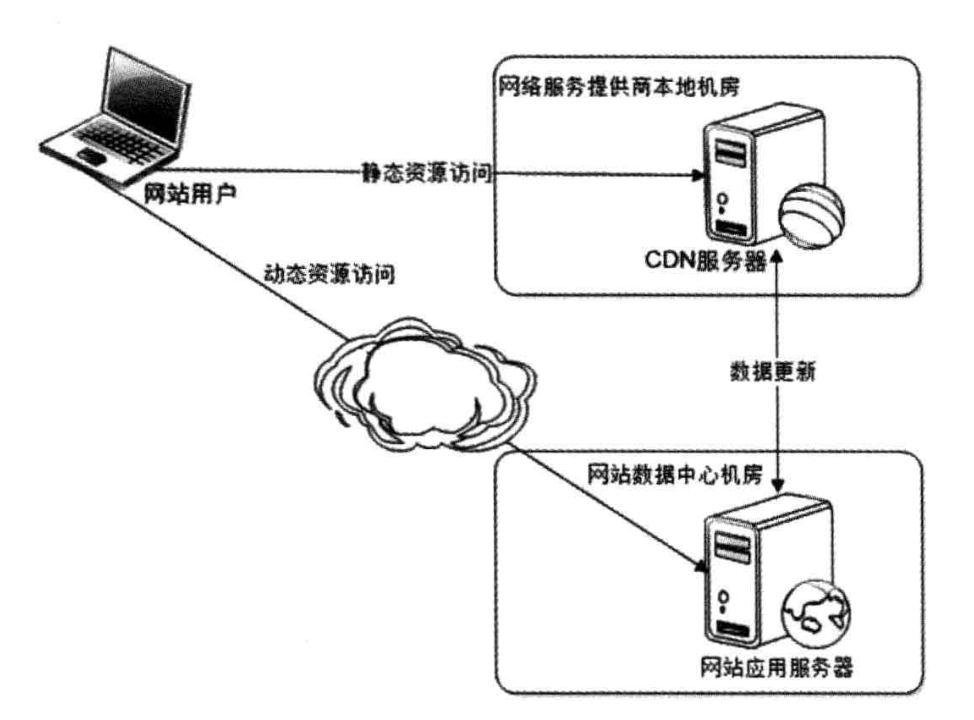
页面先加载css才对整个页面进行渲染，而js是加载后立即执行。

1. **减少Cookie传输**

减少不必要的cookie传输，静态资源不需要传输cookie，考虑静态资源单独名 访问。

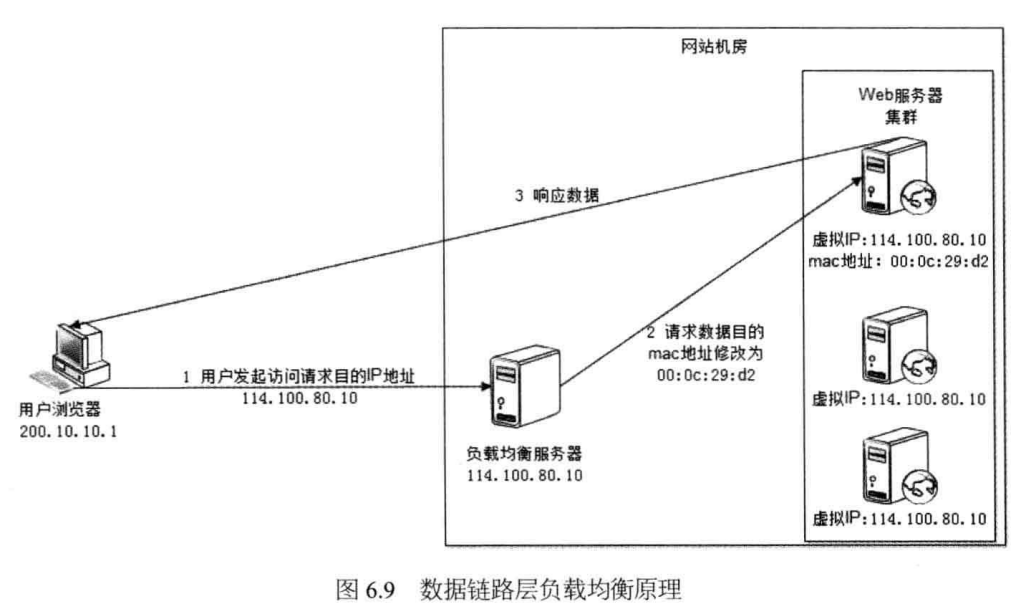
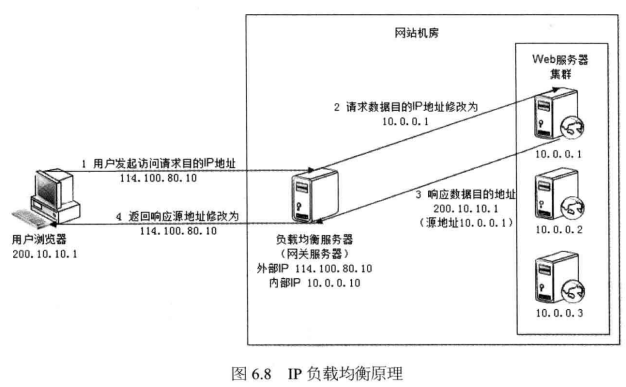
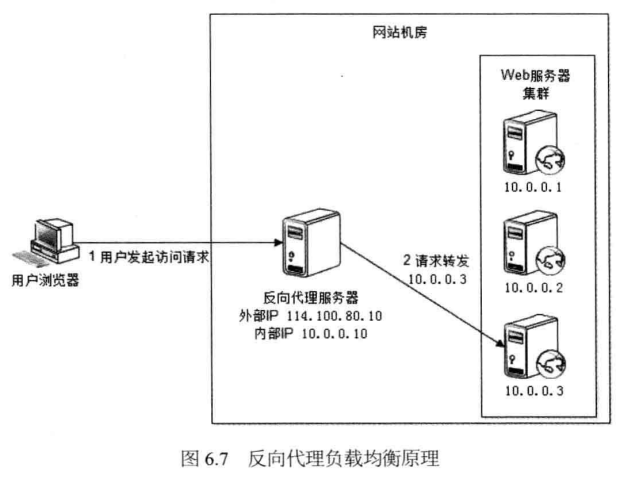
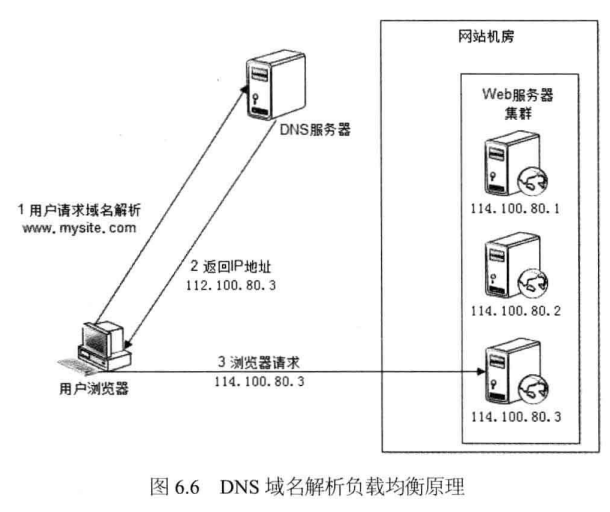
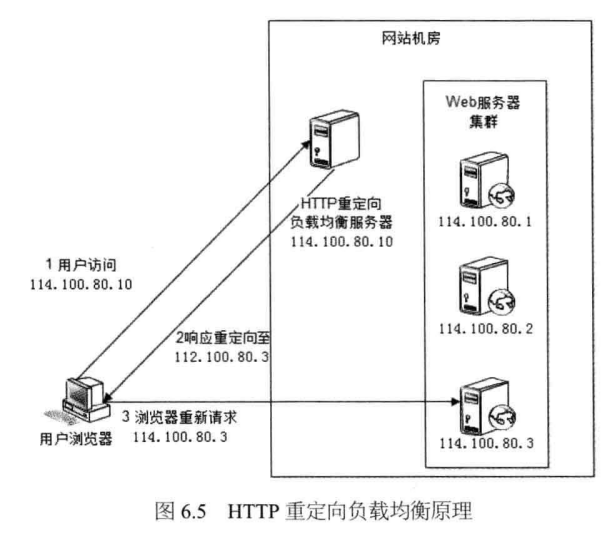
1. **CDN加速**

内容分发网络，本质是一个缓存，将数据缓存在里用户最近的服务器上，以最快速度获取数据，一般为静态资源，如图片、文件、CSS、JavaScript脚本、静态网页等。



1. **反向代理**

反向代理服务器位于网站机房一侧，代理网站web服务器接收HTTP请求。



1. **DNS**

域名服务，将域名解析成IP地址，利用DNS可以实现DNS负载均衡，配置CDN也需要修改DNS，使域名解析后指向CDN服务器。

1. **应用层架构**
2. **开发框架**

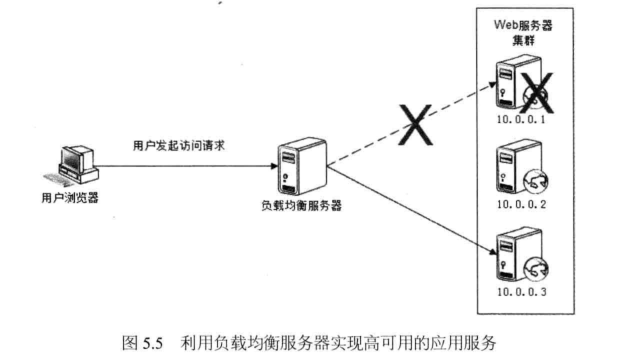
一个好的开发框架应该能分离关注面，使美工、开发工程师可以各司其职，易于协作，同时还应该内置一些安全策略，防护Web应用攻击。

1. **页面渲染**

前后端分离。

1. **负载均衡**
2. **通过负载均衡进行无状态服务的失效转移**

对于应用服务器集群，实现这种服务器可用状态实时监测、自动转移失败任务的机制就是负载均衡，将流量和数据分摊到一个集群组成的多台服务器上，以提高整体的负载处理能力；在网站应用中，当集群中的服务是无状态对等时，负载均衡可以起到事实上高可用的作用，把访问请求映射到正常服务器中，再通过心跳检测机制发现失去响应的服务器把它从列表中删除。



1. **负载均衡算法**
   * + **轮询（Round Robin， RR）**

所有请求依次分发到每台应用服务器上，适合所有服务器硬件都相同的场 景。

* + - **加权轮询（Weighted Round Robin， WRR）**

根据服务器硬件性能的情况，按照权重请求分发到每一台服务器。

* + - **随机（Random）**

请求被随机分配到各个应用服务器。

* + - **最少链接（Least Connections）**

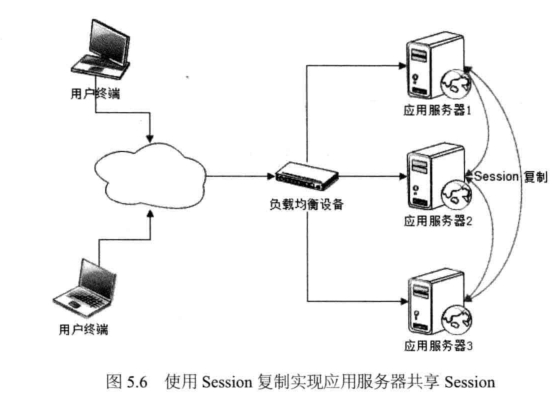
记录每个应用服务器正在处理的连接数，将请求分发到最少链接服务器 上，可以使用nginx的fail智能调度算法。

* + - **源地址散列（Source Hashing）**

根据请求IP进行Hash计算，得到应用服务器，来自同一个IP地址的请 求总在同一个服务器上处理，实现在一个会话周期内重复使用，不必会 话复制。

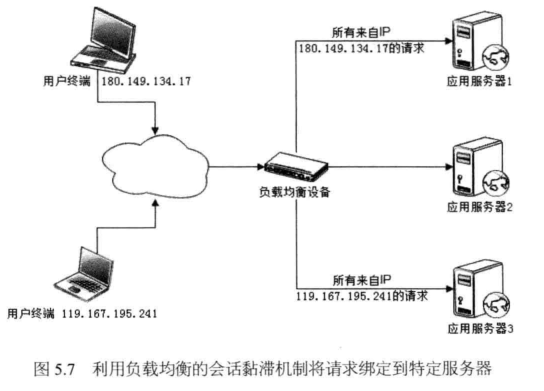
1. **Session管理**
2. **Session复制**

每次session更新，都在集群中的几台服务器之间复制同步session。



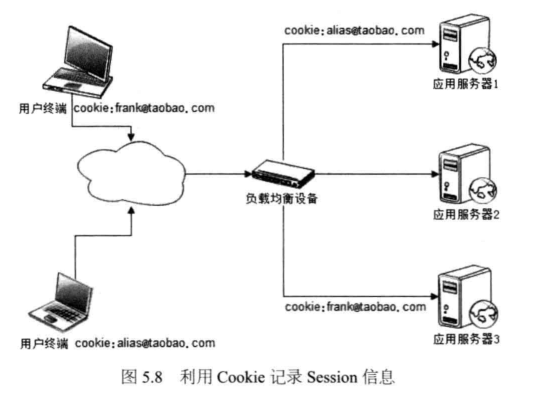
1. **Session绑定**

可以利用源地址IP的Hash算法实现。



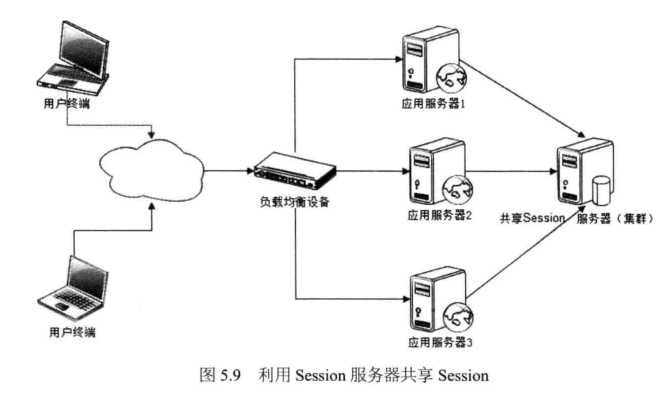
1. **利用Cookie记录Session**

C/S架构中，将session记录在客户端。



1. **Session服务器**

利用独立部署的session服务器（集群）统一管理session。



1. **动态页面静态化**

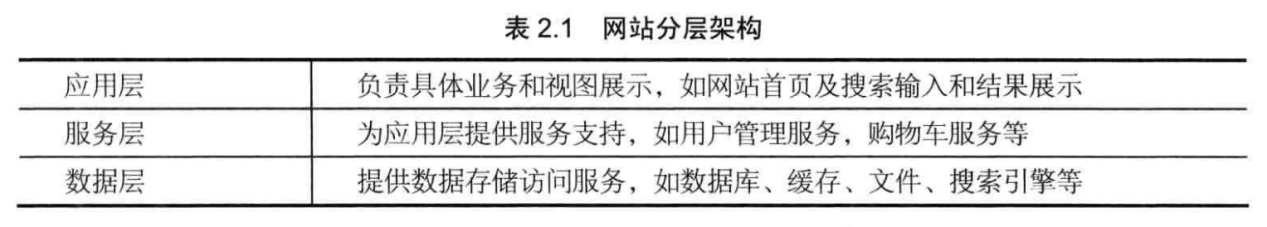
对于访问量特别大而更新又不很频繁的动态页面，可以将其静态化，即生成一个静 态页面，利用静态页面的优化手段加速用户访问，如反向代理、CDN、浏览器缓存 等。

1. **业务拆分**

根据不同业务拆分为多个规模较小的产品，独立开发、部署、维护。

1. **分层**

可分为应用层、服务层、数据层。



1. **分割**

分割就是在纵向方面对软件进行切分，如分为用户应用、商品应用（餐饮应用、 美妆应用、电子应用）、订单应用。

1. **分布式**

分层和分割的目的就是为了便于分布式部署，即将不同模块部署在不同服务器 上，通过远程调用协同工作，使用更多计算机完成同样的功能。

1. **集群**

每个应用模块在单服务器下处理请求会很吃力，所以需要多台服务器部署相同 应用构成一个集群，通过负载均衡设备共同对外提供服务。

1. **虚拟化服务器**

将一台物理服务器虚拟化成多台虚拟服务器，对于并发访问较低的业务，更容易用 较少的资源构建高可用的应用服务器集群。

1. **服务层架构**
2. **分布式消息**

利用消息队列机制，实现业务和业务、业务和服务之间的异步消息发送及低耦合的 业务关系。

1. **分布式服务**

提供高性能、低耦合、易复用、易管理的分布式服务，在网站实现面向服务架构 （SOA）。

1. **分布式缓存**

通过可伸缩的服务集群提供大规模热点数据的缓存服务，是网站性能优化的重要手 段。

1. **分布式配置**

分布式配置在系统运行期提供配置动态推送服务，将配置修改实时推送到应用系统 无需重启服务器。

1. **存储层架构**
2. **分布式文件**

一些较小但数量非常庞大的文件，进行伸缩性设计分布式文件系统。

1. **关系型数据**
2. **NoSQL数据库**
3. **数据同步**
4. **后台架构**
5. **搜索引擎**

全文搜索引擎，加快搜索速度。

1. **数据仓库**

提供数据分析与数据挖掘服务。

1. **推荐系统**

通过挖掘人与人之间的关系，人和商品之间的关系，发掘潜在人际关系和购物星期， 为用户提供个性化推荐服务。

1. **数据采集与监控**
2. **浏览数据采集**

通过在网站页面中嵌入JS脚本采集用户浏览器环境与操作记录，分析用户行为。

1. **服务器业务数据采集**

服务器业务数据包括两种，一种是采集在服务端记录的用户请求操作日志；一种是 采集应用程序运行期业务数据，比如待处理消息数目等。

1. **系统监控**

将采集的数据以图表的方式展示，以便运营和运维人员监控网站运行情况，做进一 步系统监视，更先进的做法是根据采集的数据进行自动化运维，自动处理异常，自 动化控制。

1. **系统报警**

对采集的数据进行预设阈值，比如系统负载过高，就通过邮件、短信、语音电话等 方式发出报警信号，等待工程师干预。

1. **安全架构**
2. **Web攻击**
3. **XSS攻击**
4. **注入攻击**
5. **CSRF攻击**
6. **其他攻击和漏洞**
7. **Web应用防火墙**
8. **网站安全漏洞扫描**
9. **数据保护**
10. **信息加密技术及密钥安全管理**
11. **信息过滤与反垃圾**
12. **数据中心机房架构**
13. **机房架构**

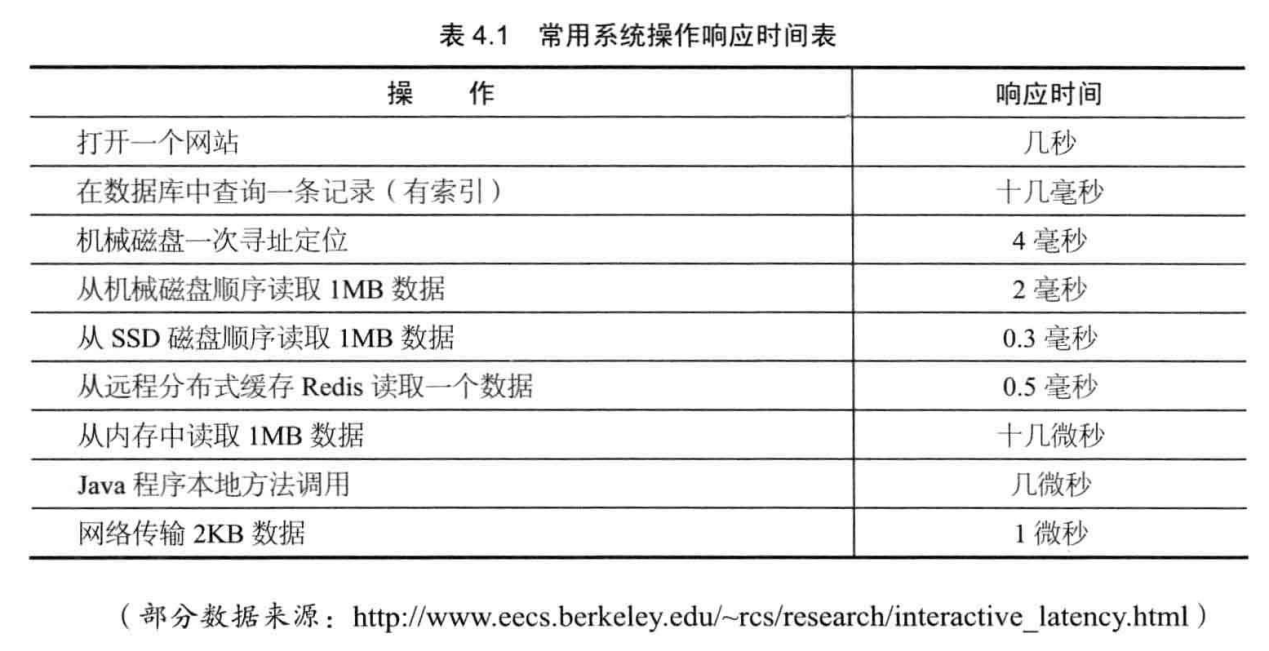
机房配置，机器设点、机房电费等问题处理。

1. **机柜架构**

包括机柜大小，网线布局、指示灯规格、不间断电源、电压规格等一系列问题。

1. **服务器架构**

定制化服务器，定制硬盘、内存、CPU等硬件。



架构总结：

1. 应用层：让处理请求更快响应，减少文件请求，减少服务器请求；
2. 服务层：应用集群化，服务颗粒细分化，达到高并发的性能；
3. 存储层：数据分布式存储，使数据能快速响应读取能力，集群的同时处理数据一致性问题；
4. 数据监控：对整个系统群数据进行采集监控，预报警提示，自动化处理，达到自动化运维的效果；
5. 分布式管理：分布式比单机版更难管理维护，对分布式配置、分布式测试、分布式发布等操作进行统一管理操作；
6. 安全性：防范Web攻击，数据安全；
7. 搜索引擎：提高网站对商品等数据的模糊搜索速度。