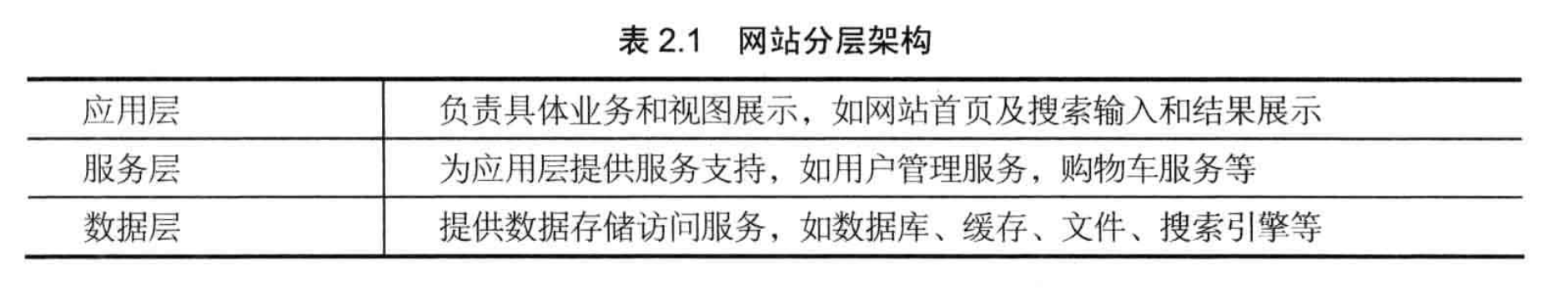
### 章节2-大型网站架构演化

什么是模式

网站问题：高性能、高可用、易伸缩、安全等。

1、分层



好处：把大的软件系统切分成不同部分，便于分工和维护，各层之间具有一定独立性，只要保持接口不变，各层可以独立演化发展。

挑战：合理规划层次边界和接口，禁止跨层调用和逆向调用。

分层架构师逻辑上的，在物理部署层次，三层既可以部署在一个物理机，也可以部署在不同机器。

分层架构对网站支持高并发向分布式方向发展至关重要，因此在项目之处就应该采用分层架构。

2、分割

分层是横向切分，分割是纵向切分。

网站越大，功能越复杂，服务和数据处理的种类也越多，将这些不同的功能和服务分割开，包装成高内聚低耦合的模块单元，一方面有助于软件开发维护，另一方面可以分布式部署，提高并发处理能力和功能扩展能力。

大型网站分割的颗粒度可能会很小，比如将购物、论坛、搜索、广告、消息、日志等分割成不同应用，由不同团队独立负责。同一个应用内部也可以继续分割，比如首页、搜索列表、商品详情等，从逻辑到物理部署都可以独立。

3、分布式

分层和分割主要为了切分模块便于维护和部署。

分布式目的是使用更多的计算机完成同样的功能呢，计算机越多，能够处理的并发量和数据量就越大。

分布式问题：服务调用通过网络，对性能造成严重影响；服务器越多，宕机的概率也越大，需要确保容错，另外，数据在分布式环境中保持事物非常困难，对业务正确性和流程可能造成很大影响。另外，分布式导致依赖错综复杂，开发维护困难。结论：不要为了分布式而分布式。

常用分布式方案：

1）分布式应用和服务：讲分层和分割后的应用和服务分布式部署，除了改善性能和并发性，加快开发和发布速度，还可以复用服务，便于业务功能扩展。

2）分布式静态资源：如JS、CSS、图片等资源独立分布式部署，采用独立域名，即动静分离。静态资源分布式部署可以减轻应用服务器压力，通过使用独立域名可以加快浏览器的并发加载速度。

3）分布式数据和存储：传统关系型数据库进行分布式部署，NoSQL产品几乎都是分布式的。

4）分布式计算：狭义上的计算，比如搜索引擎的索引构建，数据分析统计，常用Hadoop、MapReduce分布式计算框架。

4、群集

负载均衡,如首页，每一个独立应用至少两台服务器完成。

5、缓存

缓存是改善软件性能的第一手段。

缓存方式：

CDN，缓存静态资源或静态热点资源；

反向代理：静态资源放反向代理服务器；

本地缓存：应用服务器本地数据（如果常访问的热点）；

分布式缓存：大型网站的数据量非常庞大；

使用缓存的两个前提条件：1）、数据访问热点不均衡。2）、数据在某个时间段内不会很快过期，避免脏读。

缓存不仅加快数据访问速度，还可以减轻后端应用和数据存储的负载，这点至关重要，网站数据库几乎都是按照有缓存的前提进行负载能力设计的。

6、异步

系统解耦的方式除了分层、分割和分布，还有异步，业务消息传递不是同步，而是将一个业务操作分成多个阶段，各阶段通过共享数据方式异步执行。

优点：

1）提高系统可用性，队列服务出故障不影响前端业务进行；

2）加快网站相应速度，任务写入队列后可以立刻返回；

3）消除并发高峰堵塞，不会突然加重服务器负荷；

注意点：

异步处理业务可能会对用户体验、业务流程造成影响，需要根据业务需求使用。

7、冗余

宕机是必然，宕机时保证服务和数据都正常运行。冷备、热备、灾备。

8、自动化

发布方向：自动化代码管理、代码版本控制、代码分支创建合并、自动化测试、自动反馈测试结果、自动化安全检查（静态代码安全扫描和测试环境进行安全攻击测试）、自动化部署。

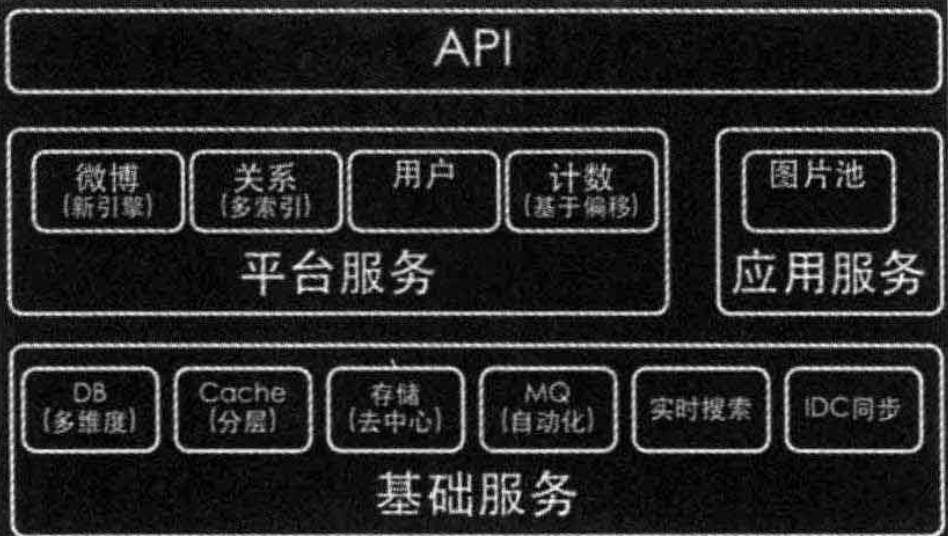
运维方向：生产环境自动化监控，服务器心跳、性能指标和应用程序关键指标，超出预定阈（yu）值自动报警。检测到故障发生后，自动化失效转移，故障消除后自动化失效恢复（RPC导向、数据一致性恢复）。自动化分配资源。

9、安全

密码和手机校验码身份认证；验证码验证；通信加密；数据库中敏感信息加密； 垃圾信息、敏感信息过滤；交易付款等重要操作根据交易模式进行风险控制。

架构模式在新浪微博的应用

开始的LAMP，短时间几次重构，现在架构：



基础服务、平台&应用服务、客户端&及生态。

业务模块和基础模块分布式部署，每个模块都部署在独立的集群上。

早期架构，微博发布使用同步推模式，用户发表微博系统会立刻将微博插入到数据库所有粉丝的订阅列表中，当明星发布微博时引起大量数据库写操作，性能急剧下降，后来改用异步推拉结合模式，发布微博将微博写入队列立刻返回，消息队列将微博推送给所有在线粉丝订阅列表，非在线登录后拉取微博。

多级缓存：热门微博和明星微博缓存在所有微博服务器上，在线用户微博和近期微博缓存在分布式缓存集群，微博中最常见的“刷微博”，几乎全部都是缓存访问。

多数据中心：即提供就近访问，又提供灾备。

工具：自动化监控、发布、故障修复等。

安全：垃圾内容、僵尸粉、攻击。

问题：缓存除了加快访问速度，另一个重要作用是什么？