

详解大型网站系统的特点和架构演化发展历程

时间: 2017-03-02 文章来源: 马海祥博客 访问次数: 859

收藏到:

1

随着国内互联网行业的高速发展，我国的大型网站系统的特点也更加鲜明和集中了，而目前，发展最快、用户量最大的莫过于电商网站和生活服务网站了。

大型网站的挑战主要来自庞大的用户，高并发的访问和海量数据，任何简单的业务一旦需要处理数以P计的数据和面对数以亿计的用户，问题就会变得棘手，大型网站架构主要就是解决这类问题。



大型网站不是从无到有一步就搭建好一个大型网站，而是能够伴随小型网站业务的渐进发展，慢慢地演化成一个大型网站。

一、大型网站系统的特点

俗话说的好：“汝欲得之，必先知之”，换言之也是一个道理，想要使用大型网站系统，你得先清楚大型网站系统的特点。



1、高并发，大流量

分类目录

SEO新闻	SEO思维
移动端SEO	SEO问答
医疗SEO	淘宝SEO
企业SEO	站外SEO
网站设计	交互设计
网站策划	网页制作
营销策划	营销案例
竞价技巧	数据分析
写作技巧	微信微博
自媒体	新媒体
内容营销	网站运营
O2O模式	App运营
产品运营	网赚教程
创新思维	电子商务
名人访谈	创业故事

热点推荐



基于眼球追踪技术对用户调研的探讨研究



运营思维

更多>>

[立即访问](#)[立即访问](#)

需要面对高并发用户，大流量访问。Google 日均 PV 35 亿，日IP访问数3亿；腾讯QQ的最大在线用户数1.4亿（2011年数据）。

2、高可用

系统7 x 24小时不间断服务。

3、海量数据

需要存储、管理海量数据，需要使用大量服务器。Facebook每周上传的照片数量接近10亿，百度收录的网页数目有数百亿，Google有近百万台服务器为全球用户提供服务。

4、用户分布广泛，网络情况复杂

许多大型互联网站都是为全球用户提供服务的，用户分布范围广，各地网络情况千差万别。在国内，还有各个运营商网络互通难的问题。

5、安全环境恶劣

由于互联网的开放性，使得互联网站更容易受到攻击，大型网站几乎每天都会被黑客攻击。

6、需求快速变更，发布频繁

和传统软件的版本发布频率不同，互联网产品为快速适应市场，满足用户需求，其产品发布频率极高。一般大型网站的产品每周都有新版本发布上线，中小型网站的发布更频繁，有时候一天会发布几十次。

7、渐进式发展

几乎所有的大型互联网站都是从小网站开始，渐进地发展起来的。Facebook是扎克伯格同学在哈佛大学的宿舍里开发的；Google的第一台服务器部署在斯坦福大学的实验室；阿里巴巴是在马云家的客厅诞生的。

好的互联网产品都是慢慢运营出来的，不是一开始就开发好的，这也正好与网站架构的发展演化过程对应（具体可查看马海祥博客《[一个普通网站发展成大型网站的架构演变历程](#)》的相关介绍）。

二、大型网站架构演化发展历程

大型网站的技术挑战主要来自于庞大的用户，高并发的访问和海量的数据，任何简单的业务一旦需要处理数以P计的数据和面对数以亿计的用户，问题就会变得很棘手。

简单来说，大型网站架构演化发展历程有以下这些阶段：

1、初始阶段的网站架构

大型网站都是从小型网站发展而来，网站架构也是一样，是从小型网站架构逐步演化而来。

小型网站最开始没有太多人访问，只需要一台服务器就绰绰有余，这时的网站架构如下图所示：



自媒体运营的规范准则

立即访问



伪原创文章的方法技巧、等级和作用

立即访问



10个改变未来的科技产品

立即访问



收集客户关系管理数据的策略和需求分析

立即访问



教你写出提高客户转化率的6个文案策略

立即访问



传统企业电商该如何制定网络销售渠道策略

立即访问



社区O2O兴起的本质与未来发展方向

立即访问



从古诗词中来看文章的写作手法

立即访问

互联网

更多>>



互联网思维究竟是一种什么样的思维？

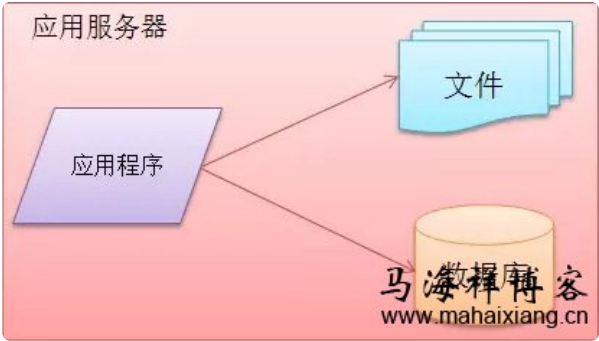
但凡做企业的，不管是创业的还是在互联网冲击下转型升级的传统行业企业家，“互联网思维”已经成为了大家共同.....



基于眼球追踪技术对用户调研的探讨...

眼球追踪技术就是当人的眼睛看向不同方向时，眼部会有细微的变化，这些变化会产生可以提取的特征，计算机可以.....





应用程序、数据库、文件等所有资源都在一台服务器上。

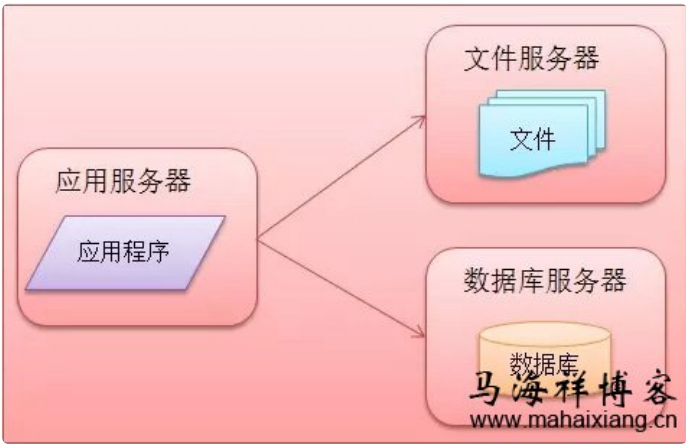
2、应用服务和数据服务分离

随着网站业务的发展，一台服务器逐渐不能满足需求：越来越多的用户访问导致性能越来越差，越来越多的数据导致存储空间不足，这时就需要将应用和数据分离。

应用和数据分离后整个网站使用3台服务器：应用服务器、文件服务器和数据库服务器。这3台服务器对硬件资源的要求各不相同：

- (1)、应用服务器需要处理大量的业务逻辑，因此需要更快更强大的CPU。
- (2)、数据库服务器需要快速磁盘检索和数据缓存，因此需要更快的磁盘和更大的内存。
- (3)、文件服务器需要存储大量用户上传的文件，因此需要更大的硬盘。

此时，网站系统的架构如下图所示：



应用和数据分离后，不同特性的服务器承担不同的服务角色，网站的并发处理能力和数据存储空间得到了很大改善，支持网站业务进一步发展。

但是随着用户逐渐增多，网站又一次面临挑战：数据库压力太大导致访问延迟，进而影响整个网站的性能，用户体验受到影响，这时需要对网站架构进一步优化（具体可查看马海祥博客《[资深架构师如何看待大型网站系统架构](#)》的相关介绍）。

3、使用缓存改善网站性能

网站访问的特点和现实世界的财富分配一样遵循二八定律：80%的业务访问集中在20%的数据上。

既然大部分业务访问集中在一小部分数据上，那么如果把这一小部分数据缓存在内存中，就可以减少数据库的访问压力，提高整个网站的数据访问速度，改善数据库的写入性能了。

如何开启苹果系统的两步验证机制，...

首先，你需要登录至苹果的网页版Apple ID管理系统，你需要点击“管理你的Apple ID”，随后输入帐号密码信息。在登录.....

网络营销

更多>>



如何在行业中打造个人品牌的影响力



腾讯微博为什么会败给新浪微博？



内容营销的方法步骤



图片社交的痛点和定位

网站制作

更多>>



计算机语言的发展简史



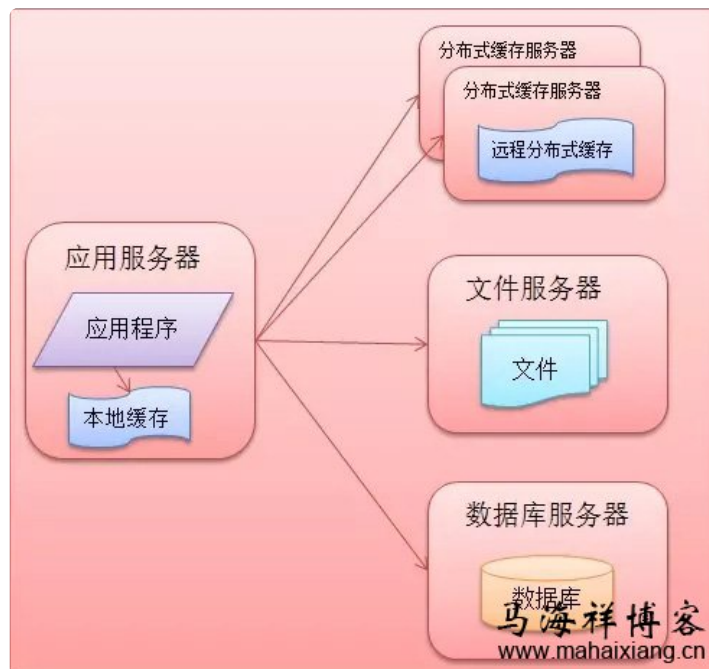
SEO优化

更多>>



百度排名11位现象的判定特征
百度排名11位是指你的站点中流量不错

远程分布式缓存可以使用集群的方式，部署大内存的服务器作为专门的缓存服务器，可以在理论上做到不受内存容量限制的缓存服务。



4、使用应用服务器集群改善网站的并发处理能力

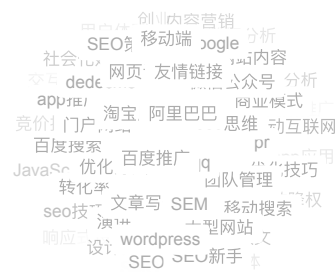
应用服务器实现集群是网站可伸缩架构设计中较为简单成熟的一种，如下图所示：

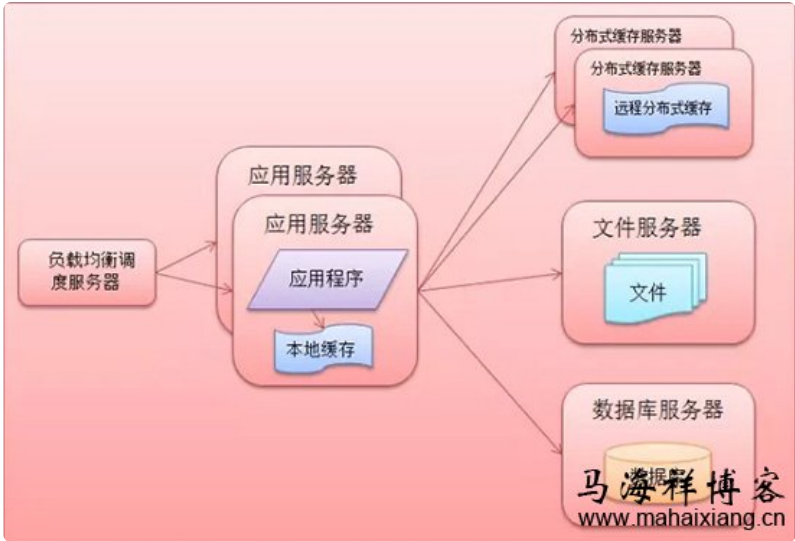
从百度经验的页面代码结构来解析站

本月热门文章

- 1 HTTP、SSL/TLS和HTTPS协议的区...
- 2 关于大型网站架构的负载均衡技术详解
- 3 如何收集和存储服务器运营的数据
- 4 基于眼球追踪技术对用户调研的探讨...
- 5 自然语言处理的单词嵌入及表征方法
- 6 基于贝叶斯推断应用原理的过滤垃圾...
- 7 基于高斯模糊原理的模糊图片的研究
- 8 详解内存数据库中的索引技术
- 9 深入解析互联网协议的原理
- 10 HTTPS建设使用的方案教程解析

标签云



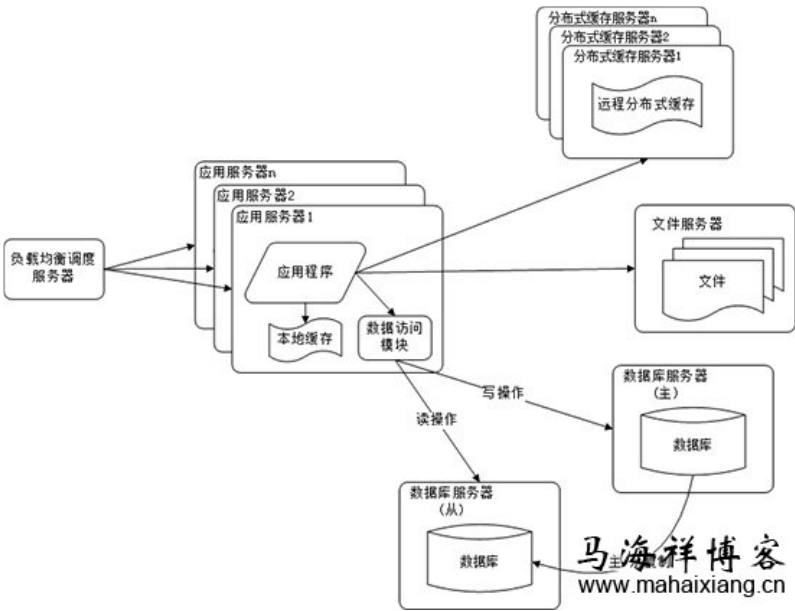


通过负载均衡调度服务器，可以将来自用户浏览器的访问请求分发到应用服务器集群中的任何一台服务器上，如果有更多用户，就在集群中加入更多的应用服务器，使应用服务器的压力不再成为整个网站的瓶颈。

5、数据库读写分离

网站在使用缓存后，使对大部分数据读操作访问都可以不通过数据库就能完成，但是仍有一部分读操作（缓存访问不命中、缓存过期）和全部的写操作都需要访问数据库，在网站的用户达到一定规模后，数据库因为负载压力过高而成为网站的瓶颈。

目前大部分的主流数据库都提供主从热备功能，通过配置两台数据库主从关系，可以将一台数据库服务器的数据更新同步到另一台服务器上。网站利用数据库的这一功能，实现数据库读写分离，从而改善数据库负载压力。如下图所示：



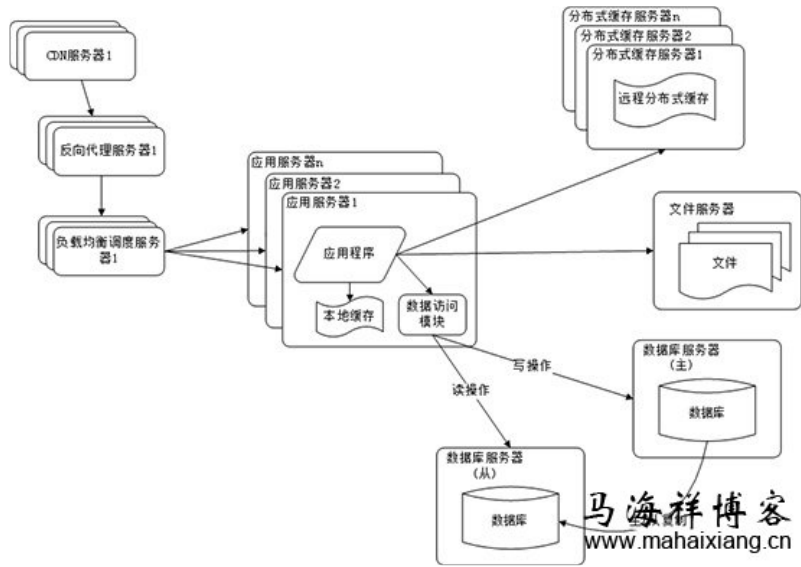
应用服务器在写数据的时候，访问主数据库，主数据库通过主从复制机制将数据更新同步到从数据库，这样当应用服务器读数据的时候，就可以通过从数据库获得数据。

为了便于应用程序访问读写分离后的数据库，通常在应用服务器端使用专门的数据访问模块，使数据库读写分离对应用透明。

6、使用反向代理和CDN加速网站响应

随着网站业务不断发展，用户规模越来越大，由于中国复杂的网络环境，不同地区的用户访问网站时，速度差别也极大。

有研究表明，网站访问延迟和用户流失率正相关，网站访问越慢，用户越容易失去耐心而离开。为了提供更好的用户体验，留住用户，网站需要加速网站访问速度，主要手段有使用CDN和方向代理。如下图所示：



CDN和反向代理的基本原理都是缓存。

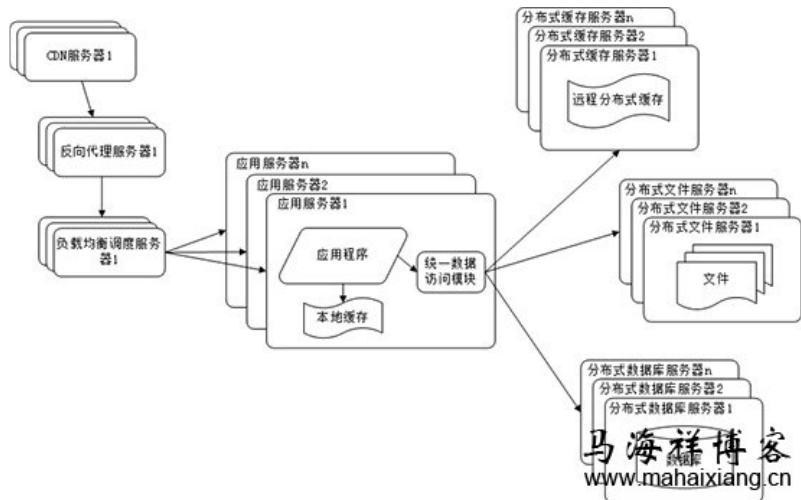
(1)、CDN部署在网络提供商的机房，使用户在请求网站服务时，可以从距离自己最近的网络提供商机房获取数据。

(2)、反向代理则部署在网站的中心机房，当用户请求到达中心机房后，首先访问的服务器是反向代理服务器，如果反向代理服务器中缓存着用户请求的资源，就将其直接返回给用户。

使用CDN和反向代理的目的都是尽早返回数据给用户，一方面加快用户访问速度，另一方面也减轻后端服务器的负载压力（具体可查看马海祥博客《[关于大型网站架构的负载均衡技术详解](#)》的相关介绍）。

7、使用分布式文件系统和分布式数据库系统

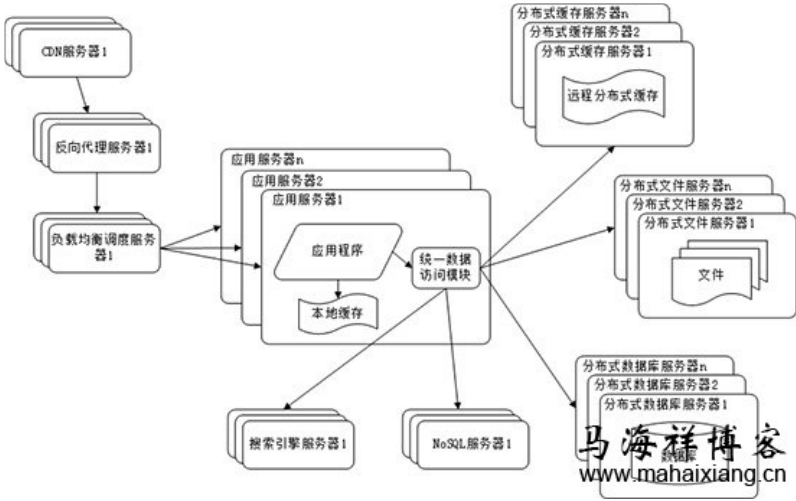
任何强大的单一服务器都满足不了大型网站持续增长的业务需求。数据库经过读写分离后，从一台服务器拆分成两台服务器，但是随着网站业务的发展依然不能满足需求，这时需要使用分布式数据库。文件系统也一样，需要使用分布式文件系统。如下图所示：



分布式数据库是网站数据库拆分的最后手段，只有在单表数据规模非常庞大的时候才使用。不到不得已时，网站更常用的数据库拆分手段是业务分库，将不同业务的数据部署在不同的物理服务器上。

8、使用NoSQL和搜索引擎

随着网站业务越来越复杂，对数据存储和检索的需求也越来越复杂，网站需要采用一些非关系数据库技术如NoSQL和非数据库查询技术如搜索引擎。如下图所示：



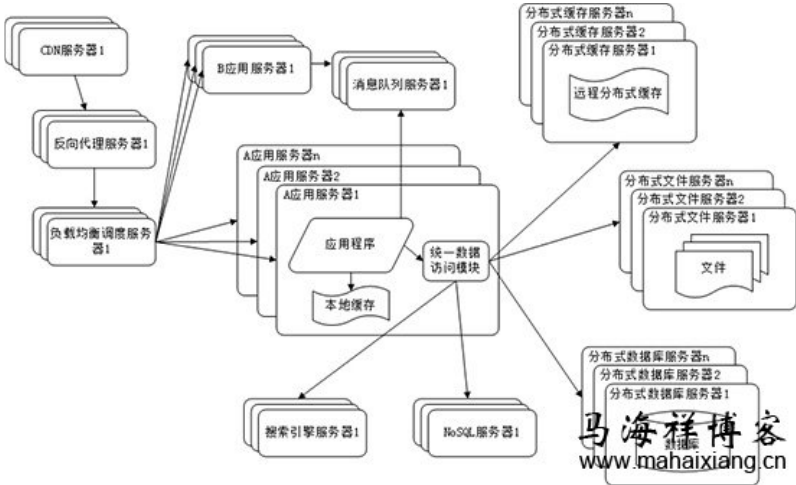
NoSQL和搜索引擎都是源自互联网的技术手段，对可伸缩的分布式特性具有更好的支持。

应用服务器则通过一个统一数据访问模块访问各种数据，减轻应用程序管理诸多数据源的麻烦。

9、业务拆分

大型网站为了应对日益复杂的业务场景，通过使用分而治之的手段将整个网站业务分成不同的产品线。如大型购物交易网站都会将首页、商铺、订单、买家、卖家等拆分成不同的产品线，分归不同的业务团队负责。

具体到技术上，也会根据产品线划分，将一个网站拆分成许多不同的应用，每个应用独立部署。



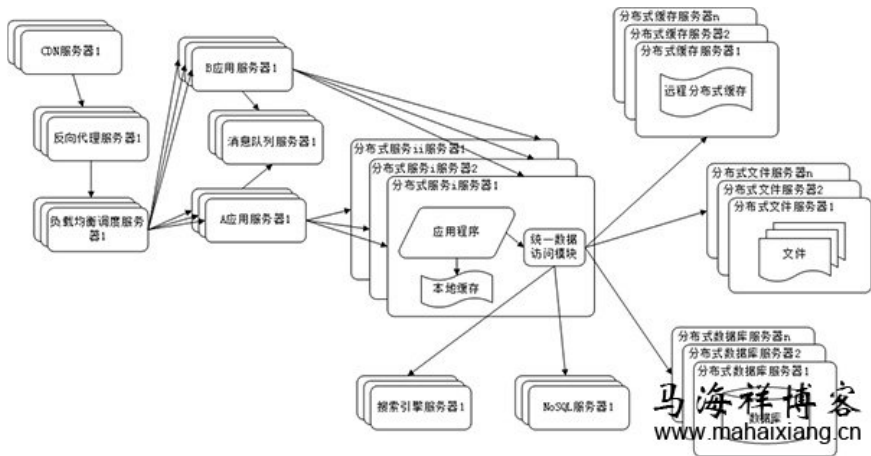
应用之间可以通过一个超链接建立关系（在首页上的导航链接每个都指向不同的应用地址），也可以通过消息队列进行数据分发，当然最多的还是通过访问同一个数据存储系统来构成一个关联的完整系统。

10、分布式服务

随着业务拆分越来越小，存储系统越来越庞大，应用系统的整体复杂度呈指数级增加，部署维护越来越困难。

由于所有应用要和所有数据库系统连接，在数万台服务器规模的网站中，这些连接的数目是服务器规模的平方，导致数据库连接资源不足，拒绝服务。

既然每一个应用系统都需要执行许多相同的业务操作，比如用户管理、商品管理等，那么可以将这些共用的业务提取出来，独立部署。由这些可复用的业务连接数据库，提供共用业务服务，而应用系统只需要管理用户界面，通过分布式服务调用共用业务服务完成具体业务操作。如下图所示：



大型网站的架构演化到这里，基本上大多数的技术问题都得以解决，诸如跨数据中心的实时数据同步和具体网站业务相关的问题也都可以通过组合改进现有技术架构解决。

三、大型网站架构的模式

最后，我们在聊聊最常见的架构模式，主要就是将系统在横向维度上切分成几个部分，每个部分单一职责。

1、分层

分层是应用系统中最常见的一种架构模式，在计算机世界中无处不在，可以更好的把庞大的软件系统切分成不同的部分，便于分工合作和维护。

分层中的挑战就是合理规划层次的边界和接口，严格遵守分层的约束，禁止跨层次的调用以及逆向调用。

2、分割

分割在纵向层面对软件进行切分。比如网购系统，在应用层，分割为交易、物流、优惠、库存等系统。

3、分布式

分割和分层便于切分后的模块能够便于分布式部署。分布式应用和服务、分布式文件系统、分布式数据和存储、分布式计算、分布式配置、分布式锁。

4、集群

多台服务器部署相同的应用构成一个集群，保证负载均衡的同时来提升系统的可用性，互相灾备。

5、缓存

CDN、反向代理、本地缓存、分布式缓存、数据库缓存、浏览器缓存。

6、异步

异步架构中最典型的的就是生产者和消费者模式，要解耦，就用异步，异步离不开队列，有效的利用异步，能够消除并发访问的高峰。

7、冗余

在集群中机器数量达到一定数量的时候，部分机器宕机是常态，因此需要数据冗余备份，数据库定期备份称之为冷备份，主从分离实时同步称之为热备份。

8、自动化

发布过程自动化、代码管理自动化、自动化测试、自动化安全扫描、自动化低级bug扫描、自动化监控、自动化报警、自动化失效转移、自动化降级。

9、安全

互联网开放的特性注定了在安全方面提升了很大的挑战。

好的设计绝对不是模仿，不是生搬硬套一个特定的模式，而是对问题深刻理解之上的创造和创新。

马海祥博客点评：

架构师最大的价值不在于掌握了多少先进的技术，而在于将一个大系统切分为N个低耦合的子模块的能力，这些子模块包含横向的业务模块，也包含纵向的基础技术模块，这种能力一部分源自于专业的技术和经验，还有一部分源自架构师对于业务场景的理解，对人性的把握，甚至对世界的认知。

本文发布于马海祥博客文章，如想转载，请注明原文网址摘自于
<http://www.mahaixiang.cn/internet/1799.html>，注明出处；否则，禁止转载；谢谢配合！

打赏

相关标签搜索： 大型网站 网站架构 网站系统

上一篇：关于大型网站架构的负载均衡技术详解

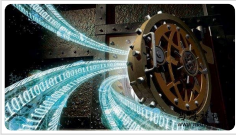
下一篇：基于眼球追踪技术对用户调研的探讨研究

相关文章推荐：

- 1 深入解析互联网协议的原理
- 2 关于大型网站架构的负载均衡技术详解
- 3 如何开启苹果系统的两步验证机制，避免
- 4 HTTP与HTTPS的区别
- 5 HTTP、SSL/TLS和HTTPS协议的区别与联系
- 6 基于眼球追踪技术对用户调研的探讨研究
- 7 HTTP服务的七层架构技术解析及运用
- 8 HTTPS建设使用的方案教程解析
- 9 今日头条的个性化推荐算法
- 10 自然语言处理的单词嵌入及表征方法



您可能还会对以下这些文章感兴趣！



详解内存数据库中的索引技术

传统的数据库管理系统把所有数据都放在磁盘上进行管理，所以称作磁盘数据库（DRDB:Disk-Resident Database），磁盘数据库需要频繁地访问磁盘来进行数据的操作，磁盘的读写速度远远小于CPU处理数据的速度，所以磁盘数据库的瓶颈出现在磁盘读写上，基于此，内存数据库的概.....【查看全文】

阅读：3257 关键词：内存数据库 索引技术 数据库 日期：2015-01-09



关于大型网站架构的负载均衡技术详解

负载均衡是将负载（工作任务，访问请求）进行平衡、分摊到多个操作单元（服务器，组件）上进行执行，是解决高性能，单点故障（高可用），扩展性（水平伸缩）的终极解决方案。面对大量用户访问、高并发请求，海量数据，可以使用高性能的服务器、大型数据库，存储设备，高性能W.....【查看全文】

文】
阅读：809 关键词：大型网站 网站架构 负载均衡 日期：2016-08-05



计算机的开机启动原理

计算机从打开电源到开始操作，整个启动可以说是一个非常复杂的过程。总体来说，计算机的整个启动过程分成四个阶段：第一阶段：BIOS；第二阶段：主引导记录；第三阶段：硬盘启动；第四阶段：操作系统；直至执行/bin/login程序，跳出登录界面，等待用户输入用户名和密码。.....【查看全文】

阅读：3039 关键词：计算机 计算机启动 计算机原理 开机启动原理 日期：2014-01-16



HTTPS建设使用的方案教程解析

百度已对部分地区开放HTTPS加密搜索服务，随后，百度实行全站化HTTPS安全加密服务，百度HTTPS安全加密已覆盖主流浏览器，旨在用户打造了一个更隐私化的互联网空间、加速了国内互联网的HTTPS化。同时也希望更多网站加入到HTTPS的队伍中来，为网络安.....【查看全文】

阅读：42 关键词：seo https 日期：2018-02-01



今日头条的个性化推荐算法

互联网给用户带来了大量的信息，满足了用户在信息时代对信息的需求，但也使得用户在面对大量信息时无法从中获得对自己真正有用的那部分信息，对信息的使用效率反而降低了，而通常解决这个问题最常规的办法是推荐系统。推荐系统能有效帮助用户快速发现感兴趣和高质量的信.....【查看全文】

阅读：12908 关键词：今日头条 日期：2016-01-20



HTTP服务的七层架构技术解析及运用

一般来说，计算机领域的体系结构普遍采用了分层的方式，从最底层的硬件往高层依次有：操作系统->驱动程序->运行库->系统程序->应用程序等等。从网络分层模型OSI来讲，由上至下为：应用层->表示层->会话层->传输层->网络层->数据链路层->物理层。当然实际应用的TCP/IP协.....【查看全文】

阅读：4386 关键词：七层架构解析 七层架构运用 七层架构技术 http服务 日期：2014-09-



HTTP、SSL/TLS和HTTPS协议的区别与联系

HTTPS是为了安全性而设置的，要验证很多的信息，相对应http请求的速度肯定有点慢，如果使用HTTPS的话很麻烦的，无意给服务器和客户端增加了很大的压力，所以平时最好不要使用HTTPS，如果牵扯到个人隐私或者是其他的什么重要信息就一定要这么做了，很多的时候你感觉有点问题，.....【查看全文】

文】

阅读：14035 关键词：http ssl https https协议 日期：2016-05-13



详解大型网站系统的特点和架构演化发展历程

大型网站的挑战主要来自庞大的用户，高并发的访问和海量数据，任何简单的业务一旦需要处理数以P计的数据和面对数以亿计的用户，问题就会变得棘手，大型网站架构主要就是解决这类问题。大型网站不是从无到有一步就搭建好一个大型网站，而是能够伴随小型网站业务的渐进发..... [【查看全文】](#)

阅读：853 关键词：大型网站 网站架构 网站系统 日期：2017-03-02



如何开启苹果系统的两步验证机制，避免iCloud帐号遭到攻击

首先，你需要登录至苹果的网页版Apple ID管理系统，你需要点击“管理你的Apple ID”，随后输入帐号密码信息。在登录之后，你需要从左侧导航栏中选择“密码和安全”选项，在这里，你将需要验证安全问题，随后下拉至“两步验证”区域，点击蓝色的“开始”链接并阅读其中的..... [【查看全文】](#)

阅读：1407 关键词：苹果系统 验证机制 icloud攻击 icloud帐号 icloud 日期：2014-09-



基于贝叶斯推断应用原理的过滤垃圾邮件研究

随着电子邮件的应用与普及，垃圾邮件的泛滥也越来越多地受到人们的关注。而目前正确识别垃圾邮件的技术难度非常大。传统的垃圾邮件过滤方法，主要有关键词法和校验码法等。前者的过滤依据是特定的词语；后者则是计算邮件文本的校验码，再与已知的垃圾邮件进行对比。它们..... [【查看全文】](#)

阅读：855 关键词：贝叶斯推断 贝叶斯应用 贝叶斯原理 过滤垃圾邮件 垃圾邮件 日期：

↓ 点击查看更多 ↓



马海祥博客
www.mahaixiang.cn

SEO优化 网站制作 网络营销 运营思维

网站导航

SEO新闻 SEO思维 移动SEO 站外SEO 站内SEO
营销策划 竞价技巧 微信微博 内容营销 营销案例
电子商务 O2O模式 App运营 网赚教程 创新思维

关注博主：

关注微信公众号



Copyright 2012~2020 马海祥博客（www.mahaixiang.cn） 版权所有 | 网站备案：苏ICP备12048125号 | 站长统计 | 警告：本站禁止镜像、反向代理和采集，文章转载请注明来源网址，否则将追究相关责任。

www.mahaixiang.cn/internet/1799.html

11/11