

大型网站性能优化实战

从前端、网络、CDN到后端、大促的全链路性能优化详解

周涛明 张荣华 张新兵 著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

大型网站性能优化实战

从前端、网络、CDN到后端、大促的全链路性能优化详解

周涛明 张荣华 张新兵 著

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

性能是大型网站的一个要素,影响性能的因素非常多。本书由三位熟悉不同领域性能优化的技术专家打造,从大型网站的整体体系出发,讲述大型网站性能优化的全链路实践过程,包括核心原理、常见策略与实战案例。具体内容包括:基于用户体验的性能优化要素、前端性能优化实战、网站性能分析、服务端性能优化、TCP 优化、DNS 优化、CDN 优化、大型网站性能监控体系、大型网站容量评估、高性能系统架构模式、大促保障体系、数据分析驱动性能优化。

本书的初衷就是将实践经验分享给读者,展示性能优化相关知识的全貌。书中的很多性能优化方法和策略都是作者从实践中总结出来的,实用性非常强。本书既可供入门者了解大型网站性能优化所有的相关技术,以及解决问题的思路和方法,也可供业界同行参考,给日常工作带来启发。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

大型网站性能优化实战:从前端、网络、CDN 到后端、大促的全链路性能优化详解 / 周涛明, 张荣华, 张新兵著. —北京:电子工业出版社, 2019.1

ISBN 978-7-121-35002-3

I. ①大… II. ①周… ②张… ③张… III. ①网站—开发 IV. ①TP393.092

中国版本图书馆 CIP 数据核字(2018)第 207734 号

责任编辑:董 英

印 刷:三河市双峰印刷装订有限公司

装 订:三河市双峰印刷装订有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编:100036

开 本:787×980 1/16 印张:21 字数:455 千字

版 次:2019 年 1 月第 1 版

印 次:2019 年 1 月第 1 次印刷

定 价:79.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式:(010) 51260888-819, faq@phei.com.cn。

序 言

写书的初衷和背景

笔者是 2011 年进入阿里的，随着业务量的逐步增长，对系统的挑战越来越大，系统的容量瓶颈也越来越明显。记得刚来的时候，业务部门要做一次促销，想知道我们的系统能不能撑得住，这个问题对于没有任何经验的同学，是很难有精确答案的。后来与同事交流并结合实践，才慢慢知道了大促的技术保障过程是怎样的。

在笔者的经历中，有太多的故事想讲给那些需要的同行们，以避免他们犯同样的错，可以做到以史为鉴。即使是阿里内部的员工，也不是每个人都有机会接触大促、接触性能优化工作的。所以这本书的初衷就是将实践经验分享给同行们，让同行们也有所收获。

文字最大的魅力是提升记忆的深度，特别是美好的东西更值得记录下来，等数年或者数十年后再去回首时，也是人生中非常美好的事情。这个世界我们曾经来过，将往事沉淀，当年很多让人觉得非常有挑战的事情，如果没有文字的记录 and 时间的沉淀，再回首时，可能已经忘记了大半。再回首 2012 年以来解决的很多难题，记忆中能够留下的已经不多，这也促使笔者决定总结一下，以便帮助更多的人。

从 2012 年开始，笔者和搭档荣华、新兵开始逐步切入大促的保障工作，并同时开始负责性能领域的工作。从事和性能相关的工作十分有趣，也很容易提升知识的深度和广度。一个大型网站在构建过程中，不仅包括容量优化，还涉及体验相关的优化，而且必须能够随时回顾这些知识。本书实际上经过 3 年的沉淀和不断修整，在繁忙的工作之余写下这些文字，是需要毅力的。市面上很多与性能相关的书讲实践的偏少，本书最大的特点就是实践偏多，通过实践进行理论升华。书中的很多内容都是 3 位作者一个字一个字写下来的，这是非常不易的。

人生是一场修炼的过程，不断地突破自己的舒适圈，不断地和内心的懒惰做斗争，咬牙做下来，才能体会到这件事情的不易，轻易得到的总是不会太珍惜。我们处于飞速发展的时代，同时我们也处于浮躁的时代，这个时代造就了很多英雄，但也很容易在其中迷失自己。在迷失的时候，给自己定一个目标，只有不断地学习和努力，同时摆正自己的心态，明确自己想要的东西，才能处在浪潮之巅，在时代的浪潮里留下自己的痕迹。希望本书也能留下一丝痕迹。

关于作者

本书的作者是在实践中逐步成长起来的架构师，在项目的实践中，全力以赴地解决各种难题。笔者 2008 年认识荣华，2011 年认识新兵，他们给了笔者很多支持。特别是 2008 年在 Cisco 工作的过程中，荣华给了笔者很多支持，笔者跟他学到了很多東西。当年 JavaEye 非常盛行的时候，荣华（以笔名 ahuaxuan）在 JavaEye 论坛里名气很大。笔者在加入 Cisco 之后，有幸与他成为同事，他教会了笔者很多东西。新兵负责过前端性能优化部分，他在前端性能方面有比较深入的研究，我们一起合作过多个优化的实践项目，并且在实践过程中取得了良好的效果。大家因为共同的爱好和兴趣而聚在一起，在策划这本书时，我们很快形成共识，大家因为志同道合而相聚，都知道这本书的意义。作为一个技术人，从一点点的知识学起，从不会、不了解，到慢慢深入，每个人在这个过程中的沉淀有所不同。本书只是从我们 3 个人的角度进行性能优化体系的阐述。

本书的体系是非常广的，可能只有很少数的人对每个部分都比较精通，碰巧的是我们 3 个人，是熟悉不同领域的工程师，这样可以各自发挥特长，让本书更有味道，也更有广度。本书是市面上第一本从前端到后端，从 CDN、DNS 到 TCP，到机房、大促全链路的关于性能优化的书籍，能够通晓这些知识，是一个工程师或者专家很难做到的。本书主要突出全面性和实战性，是目前笔者认为的最全面的关于性能优化的书籍，本书试图打造端到端的优化理论和实战体系。

本书的主要结构

本书以大型网站性能优化实战为主题，讲述了性能优化的基本理论和实践策略。

第 1 章介绍影响用户体验的几个关键要素，重点讲解白屏、首屏及页面整体加载的过程，针对它们提出了比较系统的优化策略和常见方案。

第2章主要介绍前端性能优化实战,包括延迟渲染和SEO页面的优化。重点讲解了加快页面渲染速度的基本思路及具体的实践解决方案,以及SEO页面的优化思路和解决方案。

第3章介绍如何借助站外优秀的性能测试工具,以及建立网站自己的真实用户性能监控系统,来测试和监控页面的这些关键性能指标,从而使我们能够快速对网站性能问题进行分析并做出优化。

第4章讲述服务器端性能优化理论体系和实战,包括服务器端的性能优化方法和常见的优化策略。重点包括QPS的优化、同步模型与异步模型对性能的影响、数据结构对性能的影响、算法设计不合理带来的性能问题,以及一个综合案例。

第5章阐述TCP优化,介绍了TCP的基本原理,并对TCP近几年的发展做了介绍,最后介绍了一个项目中的TCP优化实战案例。

第6章讲述DNS的一些优化方案,主要从DNS的基本原理出发,结合跨境DNS部署和实战经验,介绍了DNS优化的历程。

第7章介绍CDN的优化实践,重点包括CDN的工作原理、CDN优化的常见策略,同时介绍了大量的优化案例,最后总结了CDN的优化原则。

第8章主要讲述大型网站的性能监控需求和监控指标,以及如何实现监控,揭开大型网站性能监控体系的面纱。

第9章主要介绍大型网站如何进行容量评估,重点包括单机峰值QPS的测算、大型网站常用的容量评估方法。

第10章高性能系统架构模式,主要从宏观角度来看性能优化。好的架构是高性能的基础,只有从架构上解决问题,才能将高性能有效地持续下去。好的高扩展架构比局部优化带来的效果要大得多,如同人们经常说的,格局决定高度。

第11章重点介绍大促的整体方案和细节工作,辅以案例,加深读者的印象。大促保障是性能优化的重要力量,由于大促保障本身是成体系的,除了性能优化,还介绍了稳定性保障和资金安全保障的内容。

第12章数据分析驱动性能优化,从数据视角介绍如何进行性能优化,重点包括性能优化相关的数据分析原理与方法,以及如何在实践中使用数据分析来进行大型网站的优化。

周涛明

2018年11月

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **提交勘误：**您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动：**在页面下方 [读者评论](#) 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/35002>



目 录

第 1 章 基于用户体验的性能优化要素	1
1.1 页面用户体验的要素介绍	1
1.2 白屏时间	3
1.2.1 白屏时间的重要性	3
1.2.2 白屏过程详解	4
1.3 首屏时间	10
1.3.1 首屏时间的定义	10
1.3.2 首屏时间的重要性	11
1.4 页面整体加载完成	15
第 2 章 前端性能优化实战	16
2.1 延迟渲染	16
2.1.1 挑战和困难	17
2.1.2 解决方案	17
2.2 SEO Ajax	20
2.2.1 挑战和困难	21
2.2.2 解决方案	21
第 3 章 网站性能分析	24
3.1 快速了解网站性能	24
3.1.1 使用 YSlow 进行性能分析	24
3.1.2 使用 PageSpeed 进行性能分析	25

3.1.3	使用 WebPagetest 进行性能分析	27
3.2	真实用户前端性能监控	29
3.2.1	真实用户前端性能数据采集	29
3.2.2	数据采集可行性分析	30
第 4 章	服务端性能优化	36
4.1	最大 QPS 推算及验证	36
4.1.1	RT	37
4.1.2	单线程 QPS	38
4.1.3	最佳线程数	38
4.1.4	最大 QPS	39
4.1.5	实验数据验证公式	44
4.1.6	压力测试最佳线程数和 QPS 的临界点	47
4.2	同步模型与异步模型	49
4.2.1	同步模型	49
4.2.2	异步模型	50
4.2.3	为什么异步模型需要的线程数少	58
4.2.4	两个模型的对比及异步模型适用场景	59
4.2.5	小结	60
4.3	数据结构对性能的影响	61
4.3.1	HashMap 的问题	61
4.3.2	HashMap 的结构	62
4.3.3	碰撞	64
4.3.4	Hash 算法	65
4.3.5	题外话：ConcurrentHashMap 中的 Hash	72
4.3.6	HashMap 综述	73
4.3.7	均摊	74
4.4	算法设计不合理带来的性能问题	78
4.4.1	某应用 A 的现象	78
4.4.2	某应用 B 的现象	78
4.4.3	分析	79
4.4.4	方案	81

4.4.5	验证	82
4.4.6	小结	86
4.5	综合案例：电商活动页面性能优化	86
4.5.1	第一轮：通过 APC 使 QPS 提高近 3 倍	86
4.5.2	第二轮：解决消耗 CPU 资源大户 Gzip	92
4.5.3	小结	105
第 5 章	TCP 优化	107
5.1	TCP 传输原理	108
5.1.1	TCP 传输的简要说明	108
5.1.2	滑动窗口——接收端流量控制	108
5.1.3	拥塞窗口——发送端流量控制	109
5.1.4	传统 TCP 拥塞控制问题	110
5.2	Linux 内核升级中的 TCP 优化技术	110
5.2.1	调整接收窗口	111
5.2.2	初始拥塞窗口调整（Linux 2.6.38 开始支持）	111
5.2.3	Early Retransmit（Linux 3.5 开始支持）	112
5.2.4	初始 RTO 调整（Linux 2.6.18 开始支持）	114
5.2.5	TFO	114
5.2.6	TSO	115
5.3	TIME_WAIT 问题案例分析	116
5.3.1	问题现象	116
5.3.2	问题分析	117
5.3.3	问题初步解决	118
5.3.4	问题再分析	118
5.3.5	问题后记	119
5.4	总结	119
第 6 章	DNS 优化	120
6.1	DNS 基本原理	121
6.1.1	DNS 的一些关键术语	121
6.1.2	DNS 查询过程	122
6.1.3	NS 选择策略和机制	124

6.1.4	DNS 扩展协议 EDNS	125
6.1.5	常用 DNS 相关命令	126
6.2	实战案例：超远距离 DNS 性能问题分析和优化	130
6.2.1	现象描述	130
6.2.2	DNS Lookup 耗时长的问题分析	131
6.2.3	DNS 解析性能解决方案	133
6.3	总结	136
第 7 章	CDN 优化	138
7.1	CDN 优化概述	138
7.2	CDN 的相关术语	140
7.3	从应用看 CDN 的基本原理	141
7.3.1	CDN 基本架构	141
7.3.2	CDN 全局调度	141
7.3.3	CDN 基本调度方式	142
7.3.4	CDN 加速的基本实施流程	145
7.4	CDN 优化常见策略	146
7.4.1	静态化缓存优化	146
7.4.2	动态内容静态边缘化	147
7.4.3	动态加速优化	150
7.4.4	用户序列优化原理	153
7.4.5	域名合并优化	153
7.4.6	多级缓存架构优化	154
7.4.7	301、302 跳转边缘化访问和多终端边缘化判断	154
7.5	CDN 优化实战	155
7.5.1	CDN 的不合理架构造成 304 请求耗时长优化实战	155
7.5.2	静态资源命中率优化实战	159
7.5.3	CDN 动态加速优化实战	164
7.5.4	CDN 静态化的问题和优化实战	171
7.5.5	CDN 调度优化实战	178
7.6	总结	179

第 8 章 大型网站性能监控体系	182
8.1 监控设计	183
8.1.1 应用监控存在的问题	183
8.1.2 从问题排查思路看监控的设计	183
8.1.3 监控的设计步骤	184
8.1.4 监控常见法则总结	187
8.2 大型网站性能监控体系设计目标和原则	188
8.2.1 准确性	188
8.2.2 完整性	189
8.2.3 实时性	189
8.2.4 细分化	189
8.2.5 聚合化	189
8.2.6 图表化	190
8.2.7 可追溯	190
8.3 性能指标和监控项及实现	190
8.4 性能监控的关键指标	194
8.4.1 应用监控	194
8.4.2 系统监控	196
8.5 常用监控命令详解	201
第 9 章 大型网站容量评估	205
9.1 容量评估概述	205
9.2 容量评估的特点	206
9.3 单机峰值 QPS 的测算	206
9.3.1 单机测算方法	207
9.3.2 两种常用的引流压力测试方法	207
9.3.3 引流压力测试停止时间的判断	208
9.3.4 如何避免单机压力测试出现问题	209
9.4 大型网站常用的容量评估方法	210
9.4.1 二八原则评估法——新业务评估的基本方法	210
9.4.2 有历史数据参考的容量评估——GMV 线性比例评估法和 GMV 转化评估法	210
9.4.3 流量占比评估法	215

9.5 总结	216
第 10 章 高性能系统架构模式	218
10.1 无状态架构	219
10.1.1 解决方案一——Session 复制	219
10.1.2 解决方案二——Session Sticky	220
10.1.3 解决方案三——Session 集中式存储	220
10.1.4 解决方案四——基于浏览器 Cookie 的无状态架构	222
10.2 基于负载均衡器的水平扩展架构	222
10.3 基于 DNS 的负载均衡	224
10.4 读写分离架构	224
10.5 基于数据水平切分的水平扩展架构	225
10.6 缓存架构	228
10.6.1 缓存的基本属性	229
10.6.2 缓存的分类	229
10.6.3 缓存使用常见的问题和误区	230
10.6.4 缓存使用场景	231
10.6.5 缓存使用规范和原则	232
10.7 近端架构	233
10.8 异步化架构	234
10.9 排队缓冲架构	235
10.10 多机房架构	236
10.10.1 同城架构	236
10.10.2 异地架构	238
10.11 基于服务的可扩展架构	240
10.12 日结架构	242
10.13 热点避免架构	243
第 11 章 大促保障体系	246
11.1 大促保障概述	246
11.1.1 大促保障简介	246
11.1.2 大促保障整体流程	247
11.2 大促保障体系详解	249

11.2.1	容量保障体系	249
11.2.2	风险保障体系	253
11.2.3	组织保障	255
11.2.4	运维保障	255
11.2.5	中间件保障	256
11.3	大促容量峰值保障策略	257
11.4	大促风险保障策略	259
11.4.1	风险保障概述	259
11.4.2	风险保障常见风险	259
11.4.3	风险识别和风险分类	260
11.4.4	风险保障策略	263
11.4.5	分组隔离策略	265
11.4.6	业务降级策略	265
11.4.7	监控发现策略	265
11.5	大促资金安全保障策略	265
11.5.1	常见的资金安全防护策略	265
11.5.2	大促资金安全防护	268
11.6	大促经验沉淀	268
11.7	大促保障实战分析	269
11.7.1	机房网络瓶颈问题分析	269
11.7.2	集群个体异常造成的容量问题分析	275
11.7.3	诡异的网络瓶颈	278
11.7.4	多机房压力测试流量不均问题分析	283
11.7.5	Tengine 限流案例	291
11.8	总结	292
第 12 章	数据分析驱动性能优化	293
12.1	WebP 性能优化案例背景	293
12.1.1	WebP 格式开始兴起	294
12.1.2	WebP 改造使 L-D 转化率下降	295
12.2	性能优化中的数据分析原理与方法	296
12.2.1	数据分析简介	296

12.2.2	数据分析之杜邦分析	297
12.2.3	数据分析之多维分析	299
12.3	通过数据分析来诊断 WebP 的性能问题	303
12.3.1	指标定义	303
12.3.2	基于指标树自动诊断 WebP 的性能问题	305
12.4	案例：通过数据分析进行 OLAP 分析和 RT 优化	308
12.4.1	在线分析系统响应指标基线的定义	308
12.4.2	性能问题诊断	309
12.4.3	数据的获取及觉察	311
12.4.4	方案的推导	313
12.4.5	小结	315
12.5	通过函数抽象进行性能优化	316
12.5.1	优化过程简介	316
12.5.2	函数抽象	317
12.5.3	统计分析	319
12.5.4	小结	321

1

第 1 章

基于用户体验的性能优化要素

1.1 页面用户体验的要素介绍

说到用户体验，它给人的第一印象总是：

- 抽象，带有强烈的主观意识。
- 难以量化。

想象这样一个场景，几个人在会议室里讨论一个产品功能的雏形，没过多久，会议室就充斥着各种你来我往：

- 我觉得这样会更好。
- 我认为用户喜欢这样或者那样。
-

到目前为止，用户体验已经渗透到很多领域，囊括了从硬件到软件的诸多极致体验设计。而最为大家熟知的，就是“苹果”手机了，从视觉、交互等细节上的设计，都能感受到“苹果”对用户体验的重视，并且好的用户体验能够作为一个非常重要的元素为产品加分，也切切实实带来了好的口碑。

在竞争激烈的电商领域，用户体验优化也能作为一个重要的要素，帮助我们脱颖而出。而网站性能的好坏，会直接影响用户的第一感觉，应该没有人能忍受一个经常卡顿的系统或者产品，即使它是“苹果”手机。

以现实举例，我们假设各个电商平台是一家家 SuperMarket 或者大商场，并且它们都拥有以下几个优秀 Market 的要素：

- 商品丰富、质量有保证等。
- 信誉好、品牌影响力大等。
- 价廉物美等其他因素。

假设 A-Market 是这样的：

- 拥挤不堪。
- 导购服务不到位。
- 结账让充满挫败感。

这些都令人生畏，对吧，即使它可能拥有更丰富的商品。

B-Market 是这样的：

- 井然有序。
- 导购服务快速到位，能帮助你更快地找到想要的商品。
- 商品详细介绍专业有效。
- 结账过程快速有效。

显然，消费者更倾向于到 B-Market 购物。

接下来回归我们的主战场，作为一家服务全球的跨境电商网站，它给予用户的服务体验与线下 Market 本质是相似的，并且会更加庞大。不像普通的区域化服务的本地 Market，我们要服务更多不同国家的用户，随之而来的是，不同国家复杂的网络环境带来的各种网站性能问题，我们需要在考虑投入产出比的情况下，尽可能让不同国家的用户在网站上都能有“快”的用户体验。