

《大型分布式网站架构设计与实践》

图书基本信息

书名：《大型分布式网站架构设计与实践》

13位ISBN编号：9787121238853

出版时间：2014-9-1

作者：陈康贤

页数：460

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu111.com

《大型分布式网站架构设计与实践》

内容概要

《大型分布式网站架构设计与实践》主要介绍了大型分布式网站架构所涉及的一些技术细节，包括SOA架构的实现、互联网安全架构、构建分布式网站所依赖的基础设施、系统稳定性保障和海量数据分析等内容；深入地讲述了大型分布式网站架构设计的核心原理，并通过一些架构设计的典型案例，帮助读者了解大型分布式网站设计的一些常见场景及遇到的问题。作者结合自己在阿里巴巴及淘宝网的实际工作经历展开论述。《大型分布式网站架构设计与实践》既可供初学者学习，帮助读者了解大型分布式网站的架构，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。

《大型分布式网站架构设计与实践》

作者简介

陈康贤，淘宝花名龙隆，淘宝技术部研发工程师，2011年加入淘宝网，参与了阿里云手机商城、口碑网迁移、店铺建站、offer、支付宝卡宝、生活商城、淘宝同学等项目，在分布式系统架构设计、高并发系统设计、系统稳定性保障等领域积累了较为丰富的实践经验。曾在程序员上发表过《漫谈基于http协议的SOA架构》《浅析HTTP平台的安全性架构》两篇文章，对基于HTTP协议的SOA架构有深入研究，在排查解决线上问题和故障方面有丰富的实践经验，擅于利用数据分析解决实际问题，对新技术有浓厚的兴趣。

书籍目录

第1章 面向服务的体系架构（SOA）

1

本章主要介绍和解决以下问题，这些也是全书的基础：

HTTP协议的工作方式与HTTP网络协议栈的结构。

如何实现基于HTTP协议和TCP协议的RPC调用，它们之间有何差别，分别适应何种场景。

如何实现服务的动态注册和路由，以及软负载均衡的实现。

1.1 基于TCP协议的RPC

3

1.1.1 RPC名词解释

3

1.1.2 对象的序列化

4

1.1.3 基于TCP协议实现RPC

6

1.2 基于HTTP协议的RPC

9

1.2.1 HTTP协议栈

9

1.2.2 HTTP请求与响应

15

1.2.3 通过HttpClient发送HTTP请求

16

1.2.4 使用HTTP协议的优势

17

1.2.5 JSON和XML

18

1.2.6 RESTful和RPC

20

1.2.7 基于HTTP协议的RPC的实现

22

1.3 服务的路由和负载均衡

30

1.3.1 服务化的演变

30

1.3.2 负载均衡算法

33

1.3.3 动态配置规则

39

1.3.4 ZooKeeper介绍与环境搭建

40

1.3.5 ZooKeeper API使用简介

43

1.3.6 zkClient的使用

47

1.3.7 路由和负载均衡的实现

50

1.4 HTTP服务网关

54

第2章 分布式系统基础设施

58

本章主要介绍和解决如下问题：

分布式缓存memcache的使用及分布式策略，包括Hash算法的选择。

常见的分布式系统存储解决方案，包括MySQL的分布式扩展、HBase的API及使用场景、Redis的使用等。

如何使用分布式消息系统ActiveMQ来降低系统之间的耦合度，以及进行应用间的通信。

垂直化的搜索引擎在分布式系统中的使用，包括搜索引擎的基本原理、Lucene详细的使用介绍，以及基于Lucene的开源搜索引擎工具Solr的使用。

2.1 分布式缓存

60

2.1.1 memcache简介及安装

60

2.1.2 memcache API与分布式

64

2.1.3 分布式session

69

2.2 持久化存储

71

2.2.1 MySQL扩展

72

2.2.2 HBase

80

2.2.3 Redis

91

2.3 消息系统

95

2.3.1 ActiveMQ & JMS

96

2.4 垂直化搜索引擎

104

2.4.1 Lucene简介

105

2.4.2 Lucene的使用

108

2.4.3 Solr

119

2.5 其他基础设施

125

第3章 互联网安全架构

126

本章主要介绍和解决如下问题：

常见的Web攻击手段和防御方法，如XSS、CRSF、SQL注入等。

常见的一些安全算法，如数字摘要、对称加密、非对称加密、数字签名、数字证书等。

如何采用摘要认证方式防止信息篡改、通过数字签名验证通信双方的合法性，以及通过HTTPS协议保障通信过程中数据不被第三方监听和截获。

在开放平台体系下，OAuth协议如何保障ISV对数据的访问是经过授权的合法行为。

3.1 常见的Web攻击手段

128
3.1.1 XSS攻击
128
3.1.2 CRSF攻击
130
3.1.3 SQL注入攻击
133
3.1.4 文件上传漏洞
139
3.1.5 DDoS攻击
146
3.1.6 其他攻击手段
149
3.2 常用的安全算法
149
3.2.1 数字摘要
149
3.2.2 对称加密算法
155
3.2.3 非对称加密算法
158
3.2.4 数字签名
162
3.2.5 数字证书
166
3.3 摘要认证
185
3.3.1 为什么需要认证
185
3.3.2 摘要认证的原理
187
3.3.3 摘要认证的实现
188
3.4 签名认证
192
3.4.1 签名认证的原理
192
3.4.2 签名认证的实现
193
3.5 HTTPS协议
200
3.5.1 HTTPS协议原理
200
3.5.2 SSL/TLS
201
3.5.3 部署HTTPS Web
208
3.6 OAuth协议
215

3.6.1 OAuth的介绍

215

3.6.2 OAuth授权过程

216

第4章 系统稳定性

218

本章主要介绍和解决如下问题：

常用的在线日志分析命令的使用和日志分析脚本的编写，如cat、grep、wc、less等命令的使用，以及awk、shell脚本的编写。

如何进行集群的监控，包括监控指标的定义、心跳检测、容量评估等。

如何保障高并发系统的稳定运行，如采用流量控制、依赖管理、服务分级、开关等策略，以及介绍如何设计高并发系统。

如何优化应用的性能，包括前端优化、Java程序优化、数据库查询优化等。

如何进行Java应用故障的在线排查，包括一系列排查工具的使用，以及一些实际案例的介绍等。

4.1 在线日志分析

220

4.1.1 日志分析常用命令

220

4.1.2 日志分析脚本

230

4.2 集群监控

239

4.2.1 监控指标

239

4.2.2 心跳检测

247

4.2.3 容量评估及应用水位

252

4.3 流量控制

255

4.3.1 流量控制实施

255

4.3.2 服务稳定性

260

4.3.3 高并发系统设计

265

4.4 性能优化

277

4.4.1 如何寻找性能瓶颈

277

4.4.2 性能测试工具

285

4.4.3 性能优化措施

292

4.5 Java应用故障的排查

314

4.5.1 常用的工具

314

4.5.2 典型案例分析

331	
第5章 数据分析	
337	
本章主要介绍和解决如下问题：	
分布式系统中日志收集系统的架构。	
如何通过Storm进行实时的流式数据分析。	
如何通过Hadoop进行离线数据分析，通过Hive建立数据仓库。	
如何将关系型数据库中存储的数据导入HDFS，以及从HDFS中将数据导入关系型数据库。	
如何将分析好的数据通过图形展示给用户。	
5.1 日志收集	
339	
5.1.1 inotify机制	
339	
5.1.2 ActiveMQ-CPP	
343	
5.1.3 架构和存储	
359	
5.1.4 Chukwa	
362	
5.2 离线数据分析	
369	
5.2.1 Hadoop项目简介	
370	
5.2.2 Hadoop环境搭建	
374	
5.2.3 MapReduce编写	
384	
5.2.4 Hive使用	
389	
5.3 流式数据分析	
403	
5.3.1 Storm的介绍	
404	
5.3.2 安装部署Storm	
407	
5.3.3 Storm的使用	
418	
5.4 数据同步	
422	
5.4.1 离线数据同步	
423	
5.4.2 实时数据同步	
429	
5.5 数据报表	
431	
5.5.1 数据报表能提供什么	
431	
5.5.2 报表工具Highcharts	
432	

参考文献
445

精彩短评

- 1、目录有点用。
- 2、这不是一本介绍技术细节的书，却是目前主流大型网站可无限拓展的架构设计。看了这本书，虽然不能深入到技术细节里去了解具体的设计开发步骤，却可以一窥当前大型网站的架构思路。受益匪浅。
- 3、算是大型网站开发指南吧，入门了解读物
- 4、书的内容太愧对这么高大上的名字。没有看到太多有价值的东西。一本关于分布式架构的书竟然有大篇幅的代码，以及工具介绍，也真是让人醉了，作者看问题的层次还是不够高。
- 5、教科书类说明手册
- 6、很不错，只是安装截图和java代码太多了
- 7、名字取的太高大上了，就是一些基础知识的大杂烩
- 8、说实话，这本书有点水。
有一些拼写错误(P140)，然后注释放在错误的地方(P142)。
没有深度，谈得很浅，而且技术过时，不是淘宝在用的。书的章节结构也有点乱。
现在出个书那么容易？
- 9、才看了刚开始几页，觉得讲的挺琐碎的，ZooKeeper, HTTP, RPC, 这些简单的概念一句话带过就行，没必要写这么多代码演示，你是讲架构，不是讲如何编程
- 10、干货很少，应该叫编著
- 11、每个主题都只是浅尝辄止，读起来实在是不过瘾。
- 12、刚收到书就迫不及待的看完了，确实很不错，从书中可以看得出作者深厚的底蕴，特别是稳定性和故障排查方面，有一个系统性的介绍
- 13、作者太嫩了，搞了一堆配置说明和实例代码以为就是一本书了。
- 14、唉
- 15、分布式网站架构中用到的各个方面的技术及相关工具介绍
- 16、看了看目录，这么多名词，这本书是名词解释书籍么？我也是醉了。
- 17、谈的很宽泛，可作为学习其他相关技术栈的提纲
- 18、内容和书名有点儿不符合。深入程度的话，一般。作为入门或普及用的，还可以。
- 19、很专业，案例解析清晰，很适合初学者。强烈推荐！
- 20、作为新人入手书籍看看还行，虽都有涉及，但是都点到即止，没有深入，而且作为一本技术书籍，文中大量的代码片断，程序截图和工具介绍，实有不妥。
- 21、广、泛
- 22、怎么说呢，实在是太垃圾了。每一章节都讲得很水，就是把各种技术搬过来定义一下，基本没有什么个人的经验总结。差评，给个一星好了。
- 23、把安装截图和代码去掉至少能薄2/3，不是架构书
- 24、前两章还可以，讲的比较基础，对架构也说的比较清楚，后面就更多地侧重于目前流行的工具了，全书感觉好多工具的安裝和搭建，这不是实践，架构的书讲这些是不是有点凑数的嫌疑
- 25、受益良多 真应该早点看
- 26、入门级大杂烩读物，大段的贴代码, 贴安装方法, 非要取名 "大"、"分布式"、"架构", 醉了
- 27、偏于技
- 28、常识读物
- 29、普及读本
- 30、坦白的说有很多知识是我不具备的，从中可以了解到网站设计的架构，基础组件。但是书中的代码和，命令的使用，基础软件安装部分有拼凑页数之嫌。同期的部分书籍(阿里巴巴系列丛书)推荐不用购买，很多已购买的同事反馈是拼凑页数的。
- 31、入门级书，介绍了很多工具，顶层的设计却不多
- 32、安装指导太多了，水分很足。
- 33、与李智慧编的那本比起来差很多，感觉华而不实
- 34、推荐想转互联网开发的coder读读

《大型分布式网站架构设计与实践》

- 35、这根本就是本安装工具书呵呵，写这样骗钱的书还请水军，作者自己是不是也感到心虚了呢？
- 36、作为入门书籍总结概括得很不错，适合新人看
- 37、书中普及了一些基础知识~
- 38、书名有点夸大啊！

看目录只是在说，如何使用一些工具和中间件来创建一个网站而已
并没有体现出"大型"和"分布式"这两个词，背后所蕴藏的真正技术要点

而且也不够系统化，浅尝辄止

现在的书也越来越水了，卖点就是阿里巴巴/淘宝的工程师吗

39、形如博客

40、这本书感觉写得很不好，虽然介绍的都是分布式的概念，但是我阅读之后的感觉更多像是在堆砌概念，而不是真的用心想好怎么把知识分享出来，组织没做好，不够连贯，有待提高。

41、有的地方很啰嗦，不过还是介绍了分布式网站的各个方面，还是不错的。

42、基本都是软件安装的介绍，粗略的过了一下技术点，没有深入，最后用了7页的代码介绍了怎么给HighCharts报表页面换主题，感觉是忽悠。。。

43、断断续续读了几个月，总体来讲对架构有一定认识

44、很全面

45、初略看了一遍，收获还是蛮大的。。开阔了自己的眼界，了解了很多技术~后续会继续看相关的几本书~

ps1：看到很多负面评论，大概说是该书基础太多，没有深度什么的。。。大牛太多了，我觉得，这本书对于我这个新手，还是帮助挺大的

ps2：里面各种安装截图确实有点啰嗦，我直接跳过了。。（难怪看得这么快！^_^）

精彩书评

- 1、看过很多介绍分布式网站的书，唯独这个书即介绍了架构设计的思路，又介绍了很多实践的案例，当然，架构设计必须得到一定层次一定功力才能够看的懂，实践案例中包括soa架构，高并发系统设计案例，java原子变量的实现细节，都非常精彩非常有技术含量
- 2、昨天在某个地方看到有人推荐这本书，由于最近的项目和目录里面提到的内容关联性还是蛮大的，所以就买了一本，今天中午到货，迫不及待的看了一下，总体感觉还不错，知识点非常全面，也有很多实现细节，比如soa架构设计的实现，服务路由等等，安全方面提到的也比较多，一般的书上不常看到。
- 3、才看了刚开始几页，觉得讲的挺琐碎的，ZooKeeper, HTTP, RPC, 这些简单的概念一句话带过就行，没必要写这么多简单的代码演示，这书是在讲架构，不是讲如何编程。还会继续看下去，评论也会继续更新。
- 4、写得好不好先不说，弄一堆小号上来刷评论实在恶心。对于技术类书籍我一向抱着哪怕扔几十块钱无所谓可以买回来翻翻。一看这么多小号刷评论就不打算买的，可想而知又是一本没有内容的粘贴文。
- 5、这是一本框架性的书，它从分布式网站架构设计需要考虑的内容出发，介绍了相关的技术。虽然每一部分讲解不是特别深入，但是原理清晰，案例丰富，非常不错。现在的网站都越做越大，分布式是必须的选择，通过这本书我看到了淘宝是怎么做的，对于今后的工作很有帮助。作者也是一位非常年轻的工程师，很不错。
- 6、书中提到了很多大型分布式网站设计所必须的知识点，如soa、分布式系统的基础设施，系统稳定性设计，数据分析，系统稳定设计的章节写的很有实践性，自检脚本的编写、日志分析命令的妙用，这些知识虽然之前都听过，但是运用的不熟练，数据很多实际的案例，另外互联网安全方面之前讲类似知识点的书也比较少，书中介绍的比较详细，不错，赞一个
- 7、作者身处淘宝一线，对大型分布式网站有非常全面的认识，书中列举的案例也非常实用。一般来说架构设计是非常难的，也是非常重要的。但是本书的作者将其讲解的非常通俗易懂。当然很多内容还是可以再深入的，但是对于实践来说，最难的是全面掌握这些最基本的内容，然后再根据具体需求取舍，深入学习，等等。随着技术的发展，时间的推移，网站必然越来越大，也越来越分布式，架构就变得异常重要，本书给了一个全景概貌，很难得，非常具有参考价值。
- 8、书拿到了，很不错，印刷质量很好，书的大纲很清晰，知识点的讲解也非常细，并且有实践案例，还没来得及细读，感觉应该不错，详细介绍了离线和实时的数据分析，以及报表，这块如果结合稳定性保障中提到的那些指标，对于大型分布式场景下的集群监控应该是蛮有作用的
- 9、和作者认识有一段时间了，他写的书当然要买，作者在系统架构方面的功底是非常深厚的，也是有目共睹的，大型分布式网站的核心技术包括SOA架构的实现，缓存的使用，db的分库分表扩容等等，难能可贵的是作者还将自己在系统稳定性保障方面，包括日志分析、服务分级等经验系统的充分的进行了总结，平心而论，这是一本好书，当然，在阿里有一部分系统是开源的，还有很多系统并不是开源的，作者在相关技术原理的介绍时，使用了外界目前能够接触到的开源技术，以便介绍的时候能够既有原理，又有实现细节，这是一般介绍大型分布式网站设计的书籍所不具备的。当然，并不是所有人都能够达到这种高度，看这本书得先修炼内功。

《大型分布式网站架构设计与实践》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu111.com