

海量运维、 运营规划 之道

唐文 著

剖析 如何保障大规模互联网产品架构、网络、系统、应用稳定运行和质量、效率、成本管理。

Internet operation and maintenance
operation

Internet operation and maintenance

Internet operation and maintenance
operation

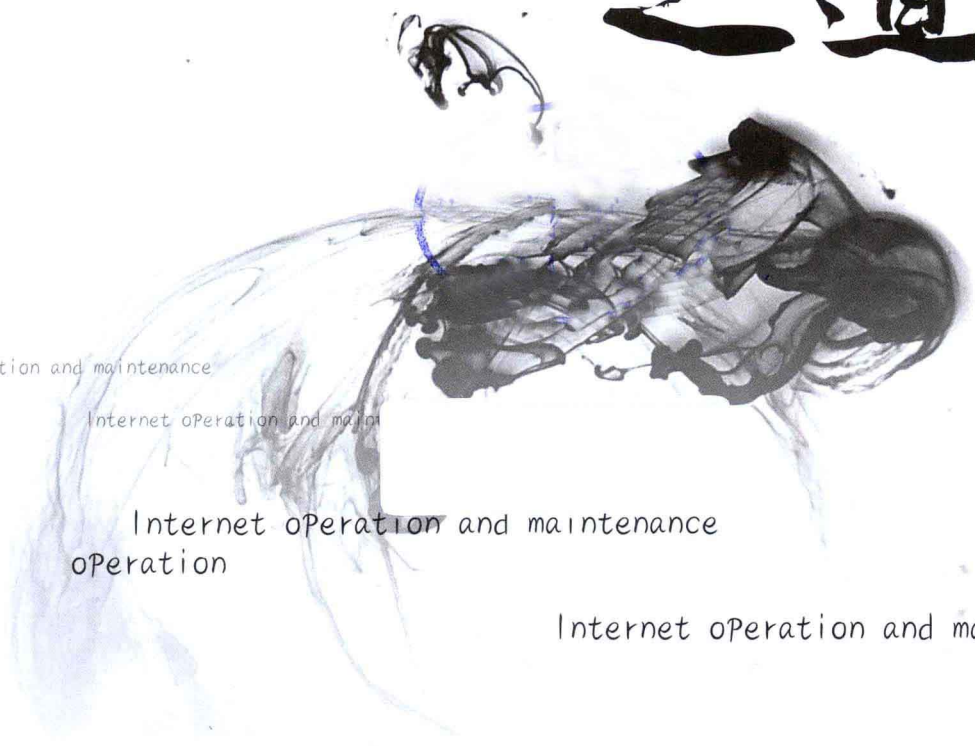
Internet operation and maintenance



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

海量运维、 运营规划 之道

唐文
著



Internet operation and maintenance
operation

Internet operation and maintenance

电子工业出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

本书作者具有腾讯、百度等中国一线互联网公司多年从业经历，书中依托工作实践，以互联网海量产品质量、效率、成本为核心，从规划、速度、监控、告警、安全、管理、流程、预案、考核、设备、带宽等方面，结合大量案例与读者分享了作者对互联网海量运维、运营规划的体会。

本书全面介绍大型互联网公司运维工作所涉及的各个方面，是每个互联网运维工程师、架构师、管理人员不可或缺的宝贵参考书。

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

海量运维、运营规划之道 / 唐文著. —北京: 电子工业出版社, 2014.1

ISBN 978-7-121-21796-8

I. ①海… II. ①唐… III. ①网站—开发 IV. ①TP393.092

中国版本图书馆CIP数据核字 (2013) 第261654号



策划编辑: 张月萍

责任编辑: 葛 娜

印 刷: 中国电影出版社印刷厂

装 订: 三河市皇庄路通装订厂

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: 13.5 字数: 339千字

印 次: 2014年1月第1次印刷

印 数: 4000册 定价: 59.00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。

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

服务热线：(010) 88258888。

推 荐 序

过去十多年，中国互联网见证了腾讯、百度等巨人的崛起，这些公司把握机会，迅速成长，在规模、效益、技术方面能与Google、Facebook、Amazon等国际巨头媲美。这些中国互联网巨头的崛起，很大程度上是由于他们有一批非常聪明、勤奋、进取的优秀员工。这些员工结合中国互联网的实际情况，不断摸索，探索出最适合中国国情的技术、产品、经营经验。这些宝贵的经验值得每一个中国互联网从业人员学习、借鉴。

本书作者唐文正是这些中国互联网巨头公司优秀员工中的一位。他在腾讯、百度从事了多年数据中心及后台服务运维，见证了业务量成百倍的增长，运维的系统支撑了每天上亿甚至数十亿计的用户请求。他把多年的运维经验总结出来，与大家分享，也算是对中国互联网业的一份贡献。

本书有大量腾讯、百度等公司运维系统的实际数据，能把这些“干货”分享出来，是本书与其他运维书籍、文章的最大不同。从这些实际运维数据，可以看到腾讯、百度等互联网巨头的数据中心、后台系统、用户数量的发展历程。

本书也总结了运维规划、运维质量、网站加速、网站监控、数据中心建设等方面的大量实战经验和心得体会，这些经验是每个互联网运维工程师、架构师、管理人员都必须掌握的。

衷心希望每位读者，无论是新手还是老将，都能从本书受益。

高德软件技术副总裁 陈 军

推 荐 序

对于很多互联网技术大拿来说，技术能力水平的确已经是很高，但是因为工作压力和文案水平的原因，很难有成体系的总结。这就造成了现在市面上很多前沿和先进的技术知识在市面上鲜有出书。唐文作为有心人，不但能把工作做好，并能把工作中的经验花时间总结成书，并且成体系地归纳出来，实在难能可贵。也可以算是互联网从业者们的福分了！这本书能从中国一线互联网公司，积累总结出来的这些知识，对于想初窥门径的初学者，或者已经有若干年经验的成熟老手，都是一份难能可贵的宝贵资料。在中国的市面上还很少看见这样的书，书中把当今中国一线互联网公司的运维工作，成体系地搬到了大家面前，让大家看到中国一线互联网公司是如何做技术运维工作、如何做产品速度体验优化的，以及做这些工作的方法论。

在这本书中你可以看到很多经典案例，以及在案例中给出的大量真实的运营数据和图表，非常宝贵！这些数据、方法，以及经验在当前中国互联网技术相关的书籍中，是非常少有和珍惜的。

同时，这也是一本适合互联网业务的架构师，或者想成为互联网业务架构师的技术人员阅读的书。书中介绍的很多案例，都是从技术架构上做的。让我们看到，中国一线互联网企业，在这块是如何做的。他们是如何实现全局负载均衡的，他们是如何实现容灾、柔性等特性的，以及他们是如何去实现成本、效率、质量三者之间的平衡的。我相信这些丰富的案例，以及方法论，应该给你足够的启发，让你能够想到自己的业务架构如何做是最合适的、最合理的。

腾讯网络媒体事业群 运维总监 王鲁强

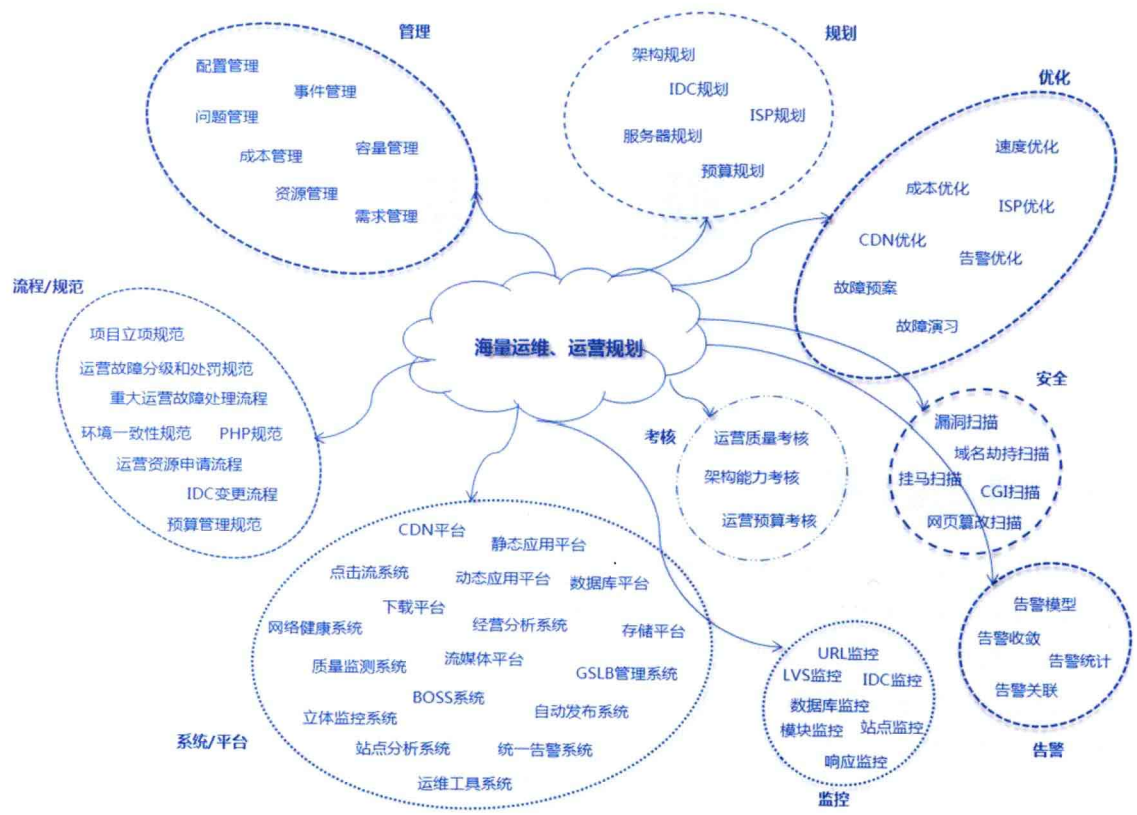
前言

为什么要写这本书

我在腾讯、百度工作多年，参与过多个海量产品的研发和维护，也见证这些产品从无到有，从小到大，并且历经数年持续迭代和稳定运行的整个生命周期。在整个生命周期中都离不开产品、研发、测试、运维等角色的全程参与和配合。而运维在上线前架构、系统、网络、硬件资源规划及上线后质量、效率、成本管理方面都发挥了不可替代的作用。优秀的架构师、团队管理者、研发工程师、测试工程师等角色都需要了解这些才能进一步做好互联网产品，本书的初衷是把运维这个角色的价值和作用介绍给所有互联网从业人员，特别是入行不久的运维朋友。

书中的内容是从2010年4月陆续记录在个人blog (<http://itkoala.com>) 上的，根据Google Analytics统计，至今浏览量有14万，访问次数为4.9万次，唯一身份访问者有2.4万人，很多热心的朋友单独联系我，给我很多支持和鼓励，这些也是我决定写这本书的动力。近3年整理的内容将作为第一版先行出版，第二版已经在梳理，时机成熟会再次出版。

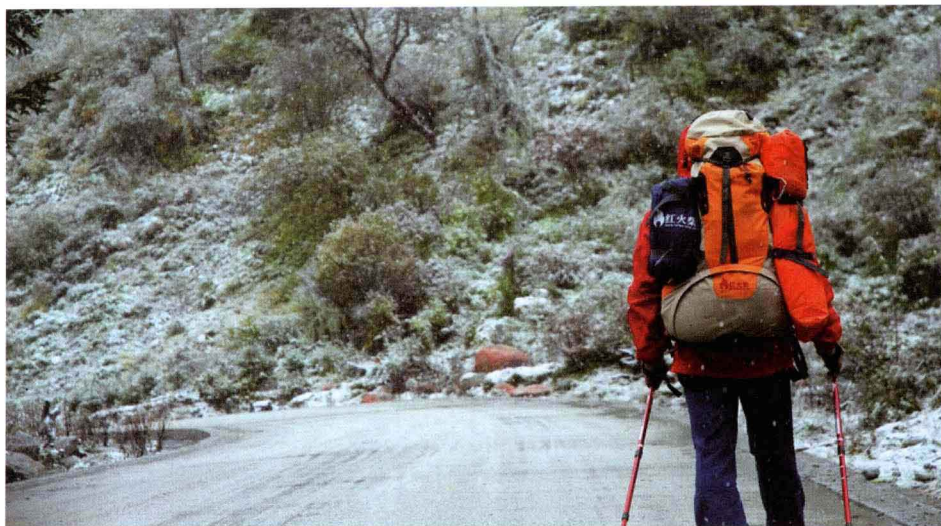
本书的名字叫《海量运维、运营规划之道》，海量运维是指支撑大规模产品的架构、网络、系统、应用的维护，运营规划是指产品运行的质量、效率、成本的管理，道在这里的意思是方向、方法，所以本书是表达如何做好大规模互联网产品维护管理的方法论。关于海量运维、运营规划，我想业界都没有准确的定义，假如说互联网的架构师用能否设计多高的摩天大楼来衡量架构能力，那么运维、运营更多的是在关注互联网服务的稳定、质量、效率、成本、故障、瓶颈及用户的忍耐、抱怨等问题。在本书中，将以质量、效率、成本为核心，从运营规划、管理、流程/规范、系统/平台、监控、告警、安全、优化、考核等几个维度结合案例来与大家分享自己的体会，内容大致如下。



痛并快乐着！

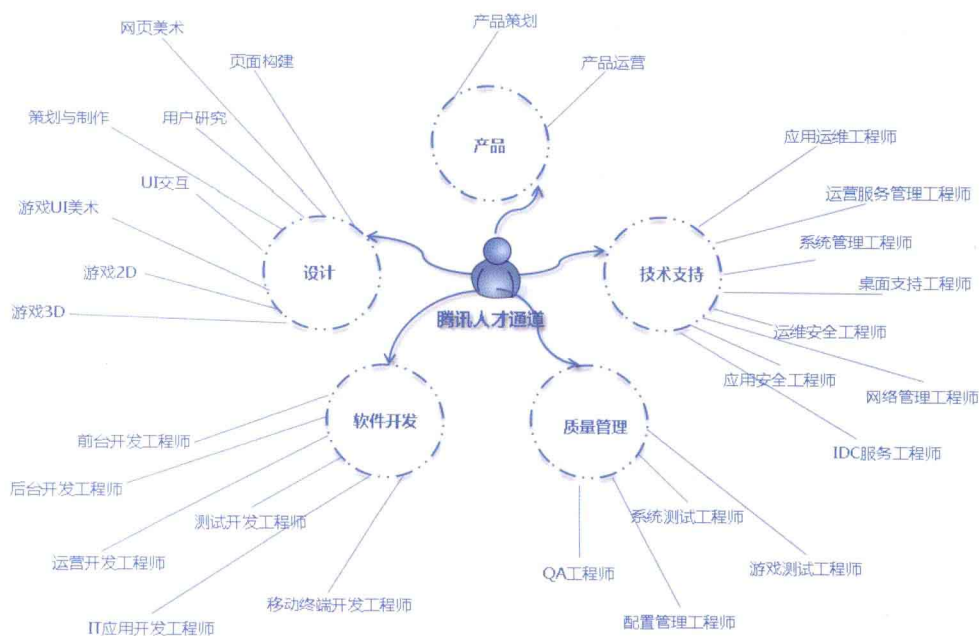
从未有过这样的充实，没有停留，离开深圳，来到上海，跟之前一样的感觉，长沙—北京，北京—深圳，深圳—上海，9年工作生涯，随风飘过。

可能，人生，在经历着无数的痛苦和快乐后，慢慢的老去，有希望达成的，也有变梦想的，结局都归属黄土。正因如此，决定留下些什么，于是有了itkoala.com这个站点，希望在这里回忆过去，在这里记录和分享自己的所感、所悟。



关于互联网运维、运营

关于互联网运维、运营，相信大家对运维比较熟悉，但对于运营，可能会有歧义。运维、运营这两个名词在腾讯会特别被强调和区分。说说我对这个岗位的理解，首先看看腾讯（百度类似）有哪些岗位及分类。



综上所述，这里对运维、运营的理解又综合、抽象了技术支持类工程师的属性，比如应用运维、运营服务管理、系统管理等，甚至是研发工程师，可以理解为一种综合素养和协调能力。具体的工作大致如下：

- 1 运营规划，提供运营综合规划支持，分析业务发展需要的资源趋势，规划IDC资源，驱动实现并与运营预算对齐。
- 2 运营预算，负责维护产品设备预算、带宽预算、专线预算的申请和滚动更新，并对运营预算结果和运营成本负责。
- 3 运营支持，负责运营服务器的日常管理与相关系统的信息变更管理；负责访问策略、域名管理、IP管理等基础运营支持。
- 4 系统运维，负责产品运营服务器的系统运维、系统监控与安全保障、系统性能分析和优化等工作；随时待命，处理产品可能遇到的问题，以及突发事件管理。
- 5 应用运维，负责业务的部署、新版本的发布、服务端的变更；负责监控业务的运行状况，及时处理业务运行中出现的故障，保障业务服务正常可用。
- 6 运维需求，负责与各开发项目组的日常沟通交流，接受并处理项目组提出的运维需求。
- 7 运营数据分析、挖掘，负责产品的业务数据分析与挖掘、性能分析与系统优化、问题跟踪与管理、负责定期给出业务运维状况报告。
- 8 运营流程、规范、制度，负责产品运维流程的探索、产品运维工作范畴与深度方面的文档建设，进行与运维相关的新技术研究；负责系统运维的知识管理体系、流程与文档建设。
- 9 运营接口、平台、培训，负责承担部门内的运营/运维技术培训；负责公共类运维支撑平台的建设和在部门内推行；负责与外部门的交互，包括外部信息对内的知会、内部需求对外的反馈。

读者对象

本书适合以下读者阅读：

- 互联网行业运维工程师、测试工程师、前端研发工程师、后端研发工程师
- 互联网行业技术经理、项目经理、产品经理、CTO
- 致力于从全局把握运维和优化网站及互联网产品的所有互联网从业人员

勘误和支持

由于编写的时间仓促，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正，书中的内容会不断更新在我的blog上，也欢迎您将错误和建议发送邮件至我的邮箱11599096@qq.com，期待能够听到读者的真挚反馈。

致谢

感谢原腾讯同事王辉和电子工业出版社张月萍老师花费大量的时间、精力来阅读我的书稿，指出许多文字和格式上的错误。感谢原腾讯同事丁吉亮、熊卫协助校稿和提供素材。

目 录

第1部分 质量

第1章 规划 /2

1.1 架构规划 /2

- 1.1.1 新趋势、新机遇、新挑战 /2
- 1.1.2 产品、架构、成本的生命周期 /3
- 1.1.3 质量、速度、成本 /5
- 1.1.4 专业化、平台化、持续化 /7
- 1.1.5 可伸缩、可调度、可扩展 /8
- 1.1.6 Set模型，集装箱改变世界 /10
- 1.1.7 全网调度，速度与容灾并进 /12
- 1.1.8 海量导致架构取向，适用是王道 /15
- 1.1.9 大系统简单做、小做 /17

1.2 IDC规划 /19

- 1.2.1 全球最复杂的基础网络 /19
- 1.2.2 IDC属性、瓶颈分析及建议 /21
- 1.2.3 IDC策略，分布制胜、速度第一 /24

1.3 CDN规划 /27

- 1.3.1 CDN发展现状与趋势分析 /27
- 1.3.2 CDN厂商分析及外包细节 /28
- 1.3.3 CDN节点分布及解析策略 /32
- 1.3.4 CDN外包质量和风险控制 /34

1.4 预算规划 /37

- 1.4.1 预算的作用及预算推导 /37
- 1.4.2 服务器定制，切合应用与性价比 /39
- 1.4.3 服务器需求，用数据说话 /42

第2章 速度 /45

2.1 网站速度优化介绍 /45

- 2.1.1 什么是网站速度优化 /45
- 2.1.2 网站速度优化的价值 /46
- 2.1.3 网站速度关键指标 /47
- 2.1.4 影响网站速度的因素 /50

2.2 网站速度监测实践 /51

- 2.2.1 网站速度监测概述 /51
- 2.2.2 网站速度主动监测 /53
- 2.2.3 网站速度被动监测 /54

2.3 网站速度分析实践 /57

- 2.3.1 网站速度分析概述 /57
- 2.3.2 网站速度分析工具 /62
- 2.3.3 网站速度分析案例 /67

2.4 网站速度优化实践 /72

- 2.4.1 网站速度优化概述 /72
- 2.4.2 网络速度优化及案例 /73
- 2.4.3 系统速度优化及案例 /80
- 2.4.4 前端速度优化及案例 /83

2.5 防止网站速度退化 /88

- 2.5.1 通过规范防止速度退化 /88
- 2.5.2 通过工具防止速度退化 /90

第3章 监控 /94

3.1 监控概述 /94

- 3.1.1 监控定义和价值 /94
- 3.1.2 监控体系和实现途径 /95

3.2 用户监控实践 /97

- 3.2.1 访问监控，获取用户访问速度 /97
- 3.2.2 URL监控，服务实时访问状态 /103
- 3.2.3 劫持监控，了解全国劫持状态 /105
- 3.2.4 内容监控，掌握网页元素变化 /108

3.3 机器监控实践 /109

3.4 网络监控实践 /113

- 3.4.1 网络监控，时刻掌控广域网变化 /113
- 3.4.2 响应监控，IDC内服务响应监控 /118

3.4.3	域名监控，做精准的DNS解析	/119
3.5	应用监控实践	/124
3.5.1	模块监控，动态应用模块监控	/124
3.5.2	数据库监控，数据库运行可视化	/126
第4章	告警	/129
4.1	告警概述	/129
4.1.1	告警定义和途径	/129
4.1.2	告警问题和规避方法	/130
4.2	告警模型	/131
4.2.1	告警分类，将告警分类区分对待	/131
4.2.2	告警模型，如何建立告警模型	/133
4.3	告警优化	/134
4.3.1	告警收敛，规避被告警淹没	/134
4.3.2	告警关联，让告警更具价值	/135
4.3.3	告警分析，用数据说话	/135
第5章	安全	/138
5.1	Web漏洞扫描	/138
5.2	域名劫持扫描	/140
5.3	DDoS攻击扫描	/142
5.4	敏感信息扫描	/144
5.5	网页篡改扫描	/146

第2部分 效率

第6章	管理	/150
6.1	配置管理，标识所有的资源并记录在册	/150
6.2	事件管理，记录、分类、诊断、解决问题	/152
6.3	问题管理，消除引起问题的根本原因	/154
6.4	变更管理，串联变更原因、过程、结果	/157
6.5	容量管理，对运营指标未来的量化、预测	/159
第7章	流程	/163
7.1	业务运维管理规范	/163

- 7.2 系统用户管理规范 /165
- 7.3 设备使用管理规范 /168
- 7.4 应用安全管理规范 /169
- 7.5 开发与编译环境安全规范 /171
- 7.6 故障分级和处罚规范 /172

第8章 预案 /175

- 8.1 预案，从容有序应对危机 /175
- 8.2 预演，提高故障执行力 /177

第9章 考核 /179

- 9.1 运维能力考核 /179
- 9.2 架构能力考核 /180
- 9.3 运营预算考核 /182

第3部分 成本

第10章 基础 /184

- 10.1 规模化导致运营成本凸显 /184
- 10.2 运营成本优化方向和策略 /185
- 10.3 工具协助成本可持续优化 /187

第11章 设备 /190

- 11.1 服务器分类、分级、分层定制 /190
- 11.2 业务分类平台化减少运营成本 /192
- 11.3 混布最大限度提高资源利用率 /192

第12章 带宽 /194

- 12.1 通过改变产品形态减少数百万成本 /194
- 12.2 通过降低带宽单价减少成本 /197
- 12.3 通过自动优化减少带宽成本 /199

第1部分

质 量

第1章 规 划

1.1 架构规划

1.1.1 新趋势、新机遇、新挑战

时代在变，世界在变，期待也在改变，我们经历了传统互联网、Web 2.0、SNS、移动互联网的变迁，下面我就互联网工作的所想、所感、所悟表达一些自己的体会，有几点分享。

- 1 全球网站TOP10变迁，搜索引擎成为网民获取内容的重要入口，UGC（用户原创内容）已经是主流产品形态，如表1-1所示。

表1-1 全球网站TOP10变迁

2004年8月全球TOP10	2009年8月全球TOP10	2013年6月全球TOP10
1. Yahoo!	1. Google	1. Google.com
2. Microsoft Network	2. Yahoo!	2. Facebook.com
3. 新浪	3. Facebook	3. YouTube.com
4. Google	4. YouTube	4. YaHoo.com
5. 搜狐	5. Windows Live	5. Baidu.com
6. 网易	6. Microsoft Network	6. Wikipedia.org
7. Baidu.com	7. Blogger	7. Live.com
8. 3721网络实名	8. Wikipedia	8. Twitter.com
9. Yahoo!カテゴリ	9. Baidu.com	9. QQ.com
10. Microsoft Corporation	10. Yahoo!カテゴリ	10. Amazon.com

- 2 UGC成为网站内容的重要来源，人际分享与SNS成为资讯阅读的另一重要渠道，半数以上网民贡献过内容。
- 3 互动和参与空前增强，多媒体、SNS、图片分享、微博等增长最快，渗透率尚低的领域，仍将快速发展。
- 4 中国互联网规模持续增长，CNNIC 2012年年底发布的数据显示，中国网民总人数达 5.64 亿，使用手机上网的用户达到了4.2亿。中国互联网用户数已于2008年超越美国，成为全球第一，与日韩、欧美相比，仍有较大增长空间，海量服务、云支撑已成为互联网主流关键词，中国的海量会成为世界级的、史无前例的海量。
- 5 大互联网公司的沟通+娱乐+社区+搜索+电子商务多业务组合战略进入中盘鏖战阶段。E级存储、T级带宽、亿级在线、20万级服务器，量变引发质变，成本+质量+柔性运营=高手生存之道。
- 6 海量将给互联网带来艰巨的挑战，千万级用户在线、每日数十亿次PV、数千台服务器、数百万用户访问带宽、P级UGC内容存储、每秒百万级动态数据请求。上千台网络设备，硬件故障随时发生；上千种后台服务，软件故障不可避免；复杂的内外网环境，网络问题家常便饭。
- 7 海量服务的规模随着用户量、信息存储量、访问量的高速增长，对基础架构平台，尤其是互联网业务架构、服务性能与容量，带宽、网络和服务器性能与容量，产生了巨大的压力和挑战，能够以低建设成本、低运营成本促进业务的可持续发展，形成互联网公司海量服务的核心竞争力，低成本的平衡扩展性成为生死要素。

1.1.2 产品、架构、成本的生命周期

关于架构规划，首先应该了解产品、架构、成本的生命周期，任何互联网企业都由多个领域、多个产品支撑，有新旧、大小之分。互联网产品经过产品策划、设计、研发、运营，然后上线进入市场，它的市场生命周期才算开始。产品、架构、成本都会随时间的推移而改变，呈现一个由少到多、由简单到复杂的过程，就如同人的生命一样，由诞生、成长到成熟，最终走向衰亡。产品的生命周期如图1-1所示，架构、成本的生命周期如图1-2所示。