

大话存储

终极版

存储系统底层架构原理极限剖析

张冬 著



清华大学出版社

大话存储

存储系统底层架构原理极限剖析

终极版

张冬 著

清华大学出版社
北京

内 容 简 介

网络存储是一个涉及计算机硬件以及网络协议/技术、操作系统以及专业软件等各方面综合知识的领域。目前国内阐述网络存储的书籍少之又少,大部分是国外作品,对存储系统底层细节的描述不够深入,加之术语太多,初学者很难真正理解网络存储的精髓。

本书以特立独行的行文风格向读者阐述了整个网络存储系统。从硬盘到应用程序,对这条路径上的每个节点,作者都进行了阐述。书中内容涉及:计算机 IO 基本概念,硬盘物理结构、盘片数据结构和工作原理,七种常见 RAID 原理详析以及性能细节对比,虚拟磁盘、卷和文件系统原理,磁盘阵列系统,OSI 模型,FC 协议,众多磁盘阵列架构等。另外,本书囊括了存储领域几乎所有的新兴技术,比如机械磁盘、SSD、FC/SAS 协议、HBA 卡、存储控制器、集群存储系统、FC SAN、NAS、iSCSI、FCoE、快照、镜像、虚拟化、同步/异步远程复制、Thin Provision 自动精简配置、VTL 虚拟磁带库、数据容灾、应用容灾、业务容灾、性能优化、存储系统 IO 路径、云计算与云存储等。

其中每一项技术作者都进行了建模和分析,旨在帮助读者彻底理解每一种技术的原理和本质。本书结尾,作者精心总结和多年来在论坛以及各大媒体发表的帖子内容,超过一百条的问与答,这些内容都是与实际紧密结合的经验总结,颇具参考价值。

本书第一版于 2008 年出版,受到业界一致肯定,历经 6 年技术沉淀重装出版。

本书适合初入存储行业的研发人员、技术工程师、售前工程师和销售人员进行阅读,同时适合资深存储行业人士用以互相切磋交流提高。另外,网络工程师、网管、服务器软硬件开发与销售人员、Web 开发者、数据库开发者以及相关专业师生等也非常适合阅读本书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

大话存储:终极版.存储系统底层架构原理极限剖析/张冬 著.—北京:清华大学出版社,2015
ISBN 978-7-302-38124-2

I. ①大… II. ①张… III. ①计算机网络—信息存储 IV. ①TP393

中国版本图书馆 CIP 数据核字(2014)第 224466 号

责任编辑:栾大成

封面设计:杨玉芳

责任校对:徐俊伟

责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印刷者:三河市君旺印务有限公司

装订者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:188 mm×260 mm 印 张:63.5 插 页:1 字 数:1668 千字

版 次:2015 年 1 月第 1 版 印 次:2015 年 1 月第 1 次印刷

印 数:1~4000

定 价:149.00 元

序 1

我关注张冬这个名字是在《大话存储》一书刚出版的时候。作为一个长期从事信息存储技术研究与教学的大学教师，自认为对于国内外关于网络存储方面的各种书籍和资料比较熟悉，对业界有哪些高人也算比较了解，但我在书店偶然发现一本名为《大话存储》书的时候，确实感到有点意外和惊喜。好像在熟悉的武林圈子之外，突然出现一位武林高手在那里论道。好奇心驱使我赶紧买了一本书回家研读，结果发现这本书确实与众不同。

与我们这些所谓学院派写的中规中矩的书相比，此书风格特立独行，语言形象生动，潇潇洒洒，颇具武侠之风。书中充满着智慧的思考和有趣的比喻，将各种原本枯燥深奥的技术概念和原理论述得十分透彻明白。不仅如此，该书还收集了大量的实例，使读者在系统获得网络存储知识的同时，还能了解典型实际系统的工作原理和技术细节，具有很好的实用性。我读完之后，对这本书的作者十分好奇。一个 80 后而且还是学化学出身的年轻人，如何就能写出这种行文老到而风格独特的专业技术书籍呢？上网查了一下冬瓜头（张冬的网名）的技术博客和他在各种论坛留下的文字，我得到了答案。这是一个完全由兴趣驱动而对技术极端痴迷的人，也是一位善于思考、富于想象力的人。这种纯粹的、不含任何功利成份的兴趣与痴迷，才是促进科学技术发展的真正源动力。

真正和张冬接触，是因为他来信质疑我们实验室申报的一项专利。收到质疑的来信，我和提出这项专利的博士生经过仔细研究，发现我们提供的图上因为少了一个非门，结果将会因为反相而出错。对如此细致具体的问题，一般人是难以发现的。如果没有打破砂锅问到底的较真精神，哪里会发现如此细节的错误呢？这种质疑的精神，在科学研究中是极为宝贵的。我们学校被称为“根叔”的李培根校长，在 2010 年的新生开学典礼大会上，就以“质疑”为题作了讲演，激励青年学子发扬质疑精神。有质疑精神的人，不唯上，不唯权威，只认真理，这正是我们这个时代所稀缺的精神。

强烈的兴趣，对技术的痴迷，加上质疑精神，成就了一本存储领域的一本好书。我在研究生新生入学之后，就推荐他们先读一下《大话存储》这本书。一方面此书对研究生而言，确实是一本网络存储技术入门的好书，另一方面我还有一个用意，就是让他们知道，要从事科学研究，强烈的兴趣比什么都重要。

信息存储是信息跨越时间的传递，也是人类传承知识的主要手段。在信息存储技术上，人类有超过万年的发明创造史。早期就地取材，人类利用石刻、泥板、竹简和羊皮等来记录信息，后来发明了纸张和活字印刷来保存和传播信息，近代发明了照相、录音和录像技术来存储信息。利用这些发明和创造，人类留下了极为丰富的文字、绘画、图像、语音和视频信息。正是这些信息，记录了人类创造的知识体系，使我们能够传承文明，并在此基础上创造新的文明。

从计算机的发明为开端，人类的信息技术进入了一个以数字化为特征的历史性新阶段。各种形式的信息被转换成数字后，以统一的方式进行处理、传输和存储，然后再转换为各种形式的信息被人们所利用。这种前所未有的方式发明之后，一个以数字化为特征的信息革命浪潮就波澜壮阔地形成。各种信息都被大规模数字化，使数字化的信息呈爆炸性增长。特别是互联网的兴起和普及，大大加快了信息的流通过程，使数字信息加速产生。图灵奖获得者 Jim Gary 观察这种数据急速增长的趋势后，总结出一个规律：人类每 18 个月新增的数据量，将是历史上所有数据量之和！如此下去，对信息存储的需求将是无止境的，信息存储技术在这种强烈的需求驱动下得到了

空前的发展。

为了保存数字化的信息，当代的科学家和工程师在最近的几十年中发明了磁存储、光存储、半导体存储等多种存储技术，其中大容量的硬盘在海量信息存储中扮演了主要的角色。硬盘的密度在短短几十年中增长了一百万倍以上，在近期，硬盘密度每年增长都接近一倍，而且还有不小的增长空间。由硬盘作为基本单元，通过各种总线、网络将硬盘连接成不同层次和不同规模的存储系统，就构成了我们目前的网络存储系统。例如由硬盘组加上冗余纠错技术构成磁盘阵列，再由磁盘阵列通过局部高速网络连接形成存储区域网；又如通过包含硬盘的大规模集群和文件系统形成的海量存储系统成为大型网站和数据中心新的存储架构。人们发明了各种技术来提高存储系统的容量、性能、效率、可用性、安全性和可管理性。存储虚拟化、归档存储、集群存储、云存储、绿色存储等新名词不断涌现，SSD 固态存储、重复数据删除、连续数据保护、数据备份与容灾、数据生命周期管理等新技术层出不穷，令人应接不暇。

在这种情况下，广大的信息领域的从业人员，信息系统的用户，以及学习信息技术的大学生和研究生，迫切需要一本既全面论述网络存储技术原理，又有丰富实例；既反映最新技术进展，又通俗易懂的书来满足他们的需求。冬瓜头的《大话存储》就是这样一本恰逢其时的好书。

《大话存储》已在业界产生了很大的影响，对存储技术在中国的普及起到了良好的推动作用。该书还被引进到我国的宝岛台湾，可见其影响深远。张冬再接再厉，以他对技术的痴迷继续钻研，对第一本书作了工作量巨大的改动与增补，并增加了云存储等全新的三章内容，全面反映了他对技术的重新思考和对最新技术的深刻理解。我相信，这些新的内容将给读者带来惊喜。

在技术发展十分迅速的领域，赶时髦的书籍多如牛毛，书店里充满了应景之作，真正经过深入思考、用心写作的书是不多的。而《大话存储终极版》却是一位技术高手的呕心沥血之作，书中对每一项技术的介绍都经过深入的思考和反复的推敲，这在当前浮躁的气氛中显得弥足珍贵。在《大话存储终极版》即将出版之际，我要向作者表示深深的敬意和衷心的祝贺，并郑重向读者推荐这本学习网络存储技术的好书。

华中科技大学计算机学院 教授
信息存储系统教育部重点实验室 主任
谢长生

序 2

第一次听说冬瓜头是因为《大话存储》，而见到冬瓜头本人时，已经听他说在写《大话存储》第二版了，而今天则是欣然为《大话存储终极版》作序。我们一起有过不长时间的沟通，全是讨论最新的存储技术。有一些内容，我相信他在书中都有写到，在同他见面前，我一直在嘀咕怎么去跟他沟通。但是，看到现实中的他朴实、憨厚、腼腆，但是对技术极其敏感，说起来一套一套的，我就开始被这个山东大汉折服了，还好我是做了准备的，否则非被问倒不可。

他是个靠笔说话和表达的人，在网络论坛里，写的文字常常是洋洋洒洒，时而言辞激烈，时而意气风发，这分明是一位才高八斗的江南才子，又像是一个书场中幽默诙谐的说书人。从此之后，我经常关注他的个人博客。他会经常在博客释放一些思想出来，豪放不羁，甚至还写过长诗以及各种各样的打油诗，有些还写得非常棒，配上那个流着鼻涕的冬瓜头漫画形象，真是绝配了。

这么一个内秀的北方大汉，用豪放的气势描述一个个生涩、枯燥的技术领域，他的思想遍布他的文字，通过笔端流放在读者面前，并且还一直这么坚持。听他说，终极版出来后，还会继续写存储领域的一些细分技术，毕竟技术日新月异，尤其是在 IT 领域。正如他所言，昨天的中国同仁在存储技术还是个初学者，今天已经开始从蹒跚学步到自主创新了，而明天，有什么理由不能期盼他们引领潮流的身影呢。我想，这也是冬瓜头要写书的动力所在吧。

《大话存储终极版》的初稿篇幅已经超过了 1500 页。在浏览了全部章节之后，发现这 1500 页中真的是字字珠玑！看得出来，是冬瓜头一个字一个字写出来的。更加可贵的是，全书字里行间透着他那独特的思想，对技术、对世界的理解以及他做人的态度。能够将这些世界观的东西融入一本技术书籍，这在以前是绝无仅有的！比如书中多次提到“轮回”、“阴阳”等，最后一节是用中医的思想来“诊治”系统性能瓶颈，看后真是令我等感叹至极！世间万物都是相互联系的，都可以找到类比和轮回，这也是冬瓜头所描述的世界观的一种。

在和冬瓜头的交谈中获知，他大学学习的专业是化学，因为高中时他化学成绩最好，所以就报了化学专业，而且期间还自学过分子生物学领域的内容，我更加惊讶了！按照他的话来说，就是“兴趣是第一驱动力”。是的，好奇和探索正是人类不断发展的第一动力。说到这里我对《大话存储终极版》中关于冬瓜头所设想的“机器如何认知自身”这段内容产生了强烈共鸣，人可以认识自身认识世界，那么机器为何不能呢？强烈的好奇心可以创造奇迹！可以让机器开口，可以让机器进化！这也正是冬瓜头所表述的世界观的一种！

《大话存储终极版》对各项存储技术的细节描述已经可以说是达到了研发级别，有很多部分甚至可以指导我们的研发！但是他却并没有用代码来表述，而是用通俗的语言和详实的图示，将原本通过阅读代码才可以理解透彻的原理，就这么轻而易举地表述了出来，这是目前我所看到的任何存储书籍或者文章都没有做到的。就这一点我曾经问过冬瓜头，问他如何做到的。他每次的回答都很简单，一针见血，实实在在，他说：“因为我就是一个从不懂钻到懂的草根，我深知一个根本不懂存储的人最想了解的东西和切入角度，并且愿意毫无保留地帮助其他草根生长！”是啊，只有亲历过那悬梁刺股的学习之路的不易，才能产出精华！

在与冬瓜头的交谈中，他还常提到一句口号：“振兴民族科技。”从他说话的眼神和口气看得出来，振兴民族科技已经成为他的信仰。他也说到，他现在所做的一切都围绕着这个信仰，他愿意为中国存储事业鞠躬尽瘁、死而后已，出版《大话存储》只是他要做的第一个环节而已，今

后他还会有一系列的动作来兑现他的诺言。信仰可以改变一个人的心态与行为，我们目前太缺乏信仰，我想如果我们所有人都有这种信仰，那么“振兴民族科技”这句口号早就可以实现了。

信息存储已经成为了一个时刻影响人们生产、生活的新兴产业，它的发展也代表着世界未来的发展，让我们再来看看国产存储信息产业的发展正在经历着怎样的变革和转变。

我国处在“十二五”时期，“十二五”期间我国要实现三大转变目标：从国强到民富、从外需到内需、从高碳到低碳。这也意味着国家的发展需要依靠科技，需要大力发展新技术，尤其以信息化技术为主轴，信息化技术的发展带动重点工程的进行，势必对国产产品催生更大的需求，我相信存储业也会有更多的民族产业佼佼者诞生。

IT 环境日益复杂，数据量快速膨胀，存储业也进入了一个技术更新极为活跃的黄金发展时期，产业发展迅速，技术活跃度高，这对国内厂家来说，无疑是一个脱颖而出的良好契机。那么，我们如何在这个时代背景下产生代表着民族存储业的国产佼佼者？

在这个时代，我们应该遵守什么？我们应该坚持什么？商业道德、创新精神、客户意识，我想只有将这些融入到企业性格中才能为企业注入新的活力。作为一家有理想的企业，需要具备一定的时代精神，而在存储技术日新月异的今天，企业打造独有的技术张性，终究才会超越历史，才会产生新时代的民族企业。我相信现在越来越多的企业正朝这个方向发展。

《大话存储》以通俗易懂的语言、风趣的行文手法向读者阐述枯燥难懂的技术精髓，致力于存储信息技术发展的民族企业也同样可以在深刻理解本土文化精髓的前提下为中国写下辉煌的历史篇章。我想这个世界没有什么不可能的，只要有这份热情、专注和执著，又有什么是不可能实现的呢？

这样的一本特立独行的书，它就是时代的产物，它就是时代的精髓。

爱数软件

李基亮

序 3

存储是个大市场，有意向在数据和信息系统上做投资规划的企业逐年增加，这标志着越来越多的企业意识到自身的数据安全问题。

在我十几年前刚刚踏入存储圈子之时，数据安全问题只被金融、电信等少数行业所考虑，而如今，几乎各个行业都存在数据保护与信息安全的需求。随着用户需求的急速增长，无论是硬件设备还是软件产品都是生机一片。但是，多年来我国的这个领域一直被国外产品所垄断，究其原因，是我国存储领域技术相对滞后。

我们在经营企业的过程中，花费了大量的精力进行人才的培养。在国内，计算机行业的传统教育大多集中于软件应用与网络维护上，对于专业存储的技术培训几乎为零，而存储行业又在飞速地发展着，因此，存储市场的需求与人才滞后的落差越拉越大，我们急切渴望拥有存储专业的人才去发展存储领域。“人才为本，教育当先”，人才的培养离不开教育。多年以来，存储领域的教材乃至书籍几乎是一片空白，有的也只是太过于教条以及模式化的书籍，当看到张冬先生的《大话存储》后，我深刻地体会到我国存储领域开始有了专业的教科书，我国的存储业生机盎然。

之所以赋予《大话存储》如此高的评价，是因为它的语言通俗而不失专业，幽默而不失严谨。张冬先生用读者极易接受的语言道出了存储领域的精髓。对于初学者来说，能使存储领域不再陌生，而又充满吸引。我曾了解到，《大话存储》已经成为某院校计算机专业的教材，这不仅是存储业的幸事，同时也是现代教育的幸事。坦率地讲，我们做企业，时刻关心教育的发展，我们需要新鲜的血液来继承和发展我们的事业。《大话存储》作为能够真正做到学以致用教材之一，使我们倍感欣慰。我为我们选择的存储道路之前景充满信心，为振兴我们的民族工业充满信心，同时，为张冬这样的后继人才而倍感骄傲。

《大话存储》能够成为教材是张冬对于存储领域不懈努力的成果，《大话存储终极版》的出版，更是他不断追求与探索的结果，而《大话存储终极版》在《大话存储》的基础上更加深入地剖析了存储技术，以及存储在如今市场的广泛应用。书中不乏一些当今企业的存储实例，也包含了国内外软硬件厂家的存储技术应用，加入了更多实际范例，使读者更易理解，同时具有很强的应用性。

我相信《大话存储终极版》会给广大读者很大的帮助，同时也希望此书能够带领更多的有识青年进入存储领域，为我国民族产业的振兴而奋斗。

火星高科 总经理
龚平

序 4

认识张冬，是因他的《大话存储》，我曾在去年拜读此书，感觉一个 80 后的小伙子能用如此通俗的语言诠释存储技术，实属存储行业的一大喜事。这本书，可以让不了解存储的人认识存储，能够了解到存储并不是高深莫测的，即使一个存储行业以外的人去阅读《大话存储》，也一定能够读懂。用什么样的语言和叙述方式不重要，重要的是把要说的说明白。

张冬本人就像他的书一样，饱含着严谨的作风和真诚的态度，而又不乏幽默的风格。看过他的 BLOG，人气一直很旺，这个致力于为国产存储业做出贡献的年轻人更是让我对他刮目相看。他在博客中写到：“我所能够做的，只有让中国人，让所有中国存储行业的人，以及中国存储行业本身，有一个扎实的基础。如果能够促进国产存储软件硬件的发展，那鄙人就是鞠躬尽瘁，死而后已，死而无憾！”一个 80 后年轻人有这样的雄心壮志，我们有什么理由不去努力不去发展国产存储业呢？

记得十几年前，我刚刚进入存储领域，那时候相关的书籍非常少，完全要靠自己进行反复的试验。那时（IT 行业根本不成形，姑且称作计算机行业）计算机业的从业者都是抱着掌握 20 世纪末最具科技含量的技术的心态进行工作，从根本上说，对存储技术充满了崇拜，甚至有一丝恐惧。在探索期间，也走了不少弯路，耽误了很多时间。如果那个时候有这样一本关于存储的书籍，那简直是一大幸事！书中并没有把存储看做是多么高深的技术，而是任何一个普通人都能掌握的技术。我和张冬开玩笑说，如果你早生 10 年，你就可以带领我们走向一条存储道路的捷径。

看到张冬最新力作《大话存储终极版》时，我就感觉到这又是一本好书。不仅延续了《大话存储》中通俗易懂的语言及“武侠”式的章节回目，在技术深度上，也有很深的挖掘。书中不仅囊括了时下最先进的“云”技术以及持续数据保护（CDP）技术，还涉及到了很多非常底层的架构。在《大话存储》的基础上，有了更为深刻的剖析。值得一提的是，张冬在最后还加入了 Q&A 的内容，把几年来读者以及网友提出的问题一一列出，并作出详细的解答，能够体会张冬在这一年多的时间里，对于存储技术的探索花了很大的心思。最可贵的是，这个年轻人不以如此成就为骄傲，继续孜孜不倦地探求。

《大话存储终极版》是一本好书，作者那严谨而真诚的态度以及致力于发展本国存储业的信心注定能够成就这样一部优秀的作品。我完全有理由相信此书能够给从业者乃至热爱存储的读者带来帮助。从中，你会受益匪浅，并乐意向你的朋友推荐此书。

火星高科 技术总监

黄 疆

作者感言

各位读者好，很高兴再次为大家“大话”存储，记得上一次是在3年前。6年前，当《大话存储》一书在2008年出版面世之后，我当时就许下承诺，要写《大话存储终极版》。当时之所以敢夸下海口要继续写第二本，是因为《大话存储》介绍了存储领域最基本的概念和架构，但没有深入涉及存储领域最新的技术，比如重复数据删除、Thin Provision、动态分级存储、CDP连续数据保护、SSD固态硬盘、FCoE、SAS、云计算和云存储等。

当年的《大话存储》确实满足了广大读者的一定需求，出版之后也获得了诸多好评和官方的、民间的很多奖项。这些成果逐渐让我感觉到更大的责任和压力。正因如此，所以我深知绝对不能就此停歇，学习是永无止境的，技术是不断发展的，所以我先向大家做了承诺，这样就可以无时无刻地激励我继续学习研究下去了。

写作过程是极其困难的，尤其是当一字一句都需要精雕细琢，并且时刻以通俗表达且让所有人都能看懂的原则和基准去写的时候，其所耗费的精力和脑力是巨大的。记得在一年前撰写本书主体的时候，基本上每天都是早晨七八点钟起来，从床上直接到书桌前开始写，直到中午吃饭，吃饭过程中依然在脑海中构思着，就这样一直到晚上，最晚的一次记得是做一个实验，通宵达旦，直到第二天天亮，实在体力不支，去床上躺到中午，然后继续写。每次睡觉之前，都会带着一个疑问入睡，躺下之后就在脑海中构思、建模，一旦想到某些重要的东西，就用笔记下几个关键词，否则第二天准忘。大部分情况一般都是没想到什么思路就已经呼呼大睡了。这种状态持续了半年之久，当完成了主体稿件之后，真的有一种如释重负的感觉。可惜，好景不长，随着不断的学习和深入，逐渐发现已经写完的内容当中有大量需要补充完善、修饰的部分，在修饰完善的过程中，继续思考，结果发现又引申出更多的东西，有些甚至推翻了以前的结论。这种状态又持续了半年，最终定稿交给编辑之后，依然发现还有零碎的东西需要完善甚至推翻，结果一再将更新的内容同步给编辑，导致出版日期一推再推，出版社相关编辑、校对叫苦不迭，还好咱的老战友大成编辑一如既往地支持，我们都顶住了压力，直到最后一个月时间内没有再发现需要完善的内容，达到了最终收敛。后面这个过程感觉更加耗费精力，因为当你重新审视之前内容的时候，一旦发现不完善甚至错误，就会感觉到一种挫败感和愧疚感，使你的激情和斗志有所丧失。

写书不但是给他人共享知识的过程，它更是一个总结自身知识体系、提高自身修养以及让自己学习更多知识的途径。比如，我在写书过程中，不但通过各方面渠道纠正了之前对某项技术的一些错误认识，而且还学习了更多的知识，并且将这些知识进行深度理解分析，之后通俗地表达出来。当你发现其他人通过你的知识快速提高之后，这种感觉是最充实的。只有在奉献之后才会感到充实，而不是一味的去索取，这样只能更加空虚。

本次终极修订版，一是针对《大话存储》和《大话存储终极版》这两本书中的错误进行修改，二是将前两本书彻底整合成为一本，三是增加了一些前沿内容，四是针对前作中读者反映不太懂的地方进行了重新诠释。鄙人的下一本书已经在写作当中，本着十年磨一剑、磨不好绝不拿出来的态度，相信不会让读者失望。

感谢家人对我的支持！长达半年的无业状态，没有家人支持就没有这本书。

感谢那些曾经帮助过我的不计其数的网友和同事！没有鼓励也不会有这本书。

感谢清华大学出版社的工作人员为本书所付出的工作！没有信任更不会有这本书。

感谢华中科技大学武汉光电国家重点实验室博士生导师谢长生教授对鄙人和本书的大力支持!

感谢本书的广大读者, 你们的支持给了我持续前进的动力!

致本书最终修订版交稿之前, 我的女儿诞生了, 感谢妻子。不知道她成人之后, 能否还能看到 20 多年前的这本由他父亲所著的著作, 虽然那时候本书可能早已过时。我会努力把钻研、毅力、执着、忘我这些亘古不变的东西, 在女儿成长过程中传授给她, 我相信, 有了这些, 做什么都会成功。



作者联系方式:

QQ: 122567712

Email: 122567712@ qq.com; myprotein@ sina.com

MSN: myprotein0007@ hotmail.com

新浪微博: @ 传说中的冬瓜头

Blog: <http://space.doit.com.cn/35700>

读者交流 QQ 群: 72168388、361934810

关于书中的武侠情节

本书前几章存在一些武侠情节，这些情节自从本书第一版之后，收到了褒贬不一的评价，有些读者甚至看到这些情节之后非常愤怒，觉得作者在装 B，从而从心理上无法接受，也就没有继续阅读。对此我深表遗憾，也正因如此，个人觉得有必要在此澄清一下这些情节的由来，以及里面那些诗句的由来。

天将降大任于斯人也，必先苦其心志，劳其筋骨，饿其体肤，空乏其身，行拂乱其所为。通俗地写出存储系统的全貌，应该也算是个大任了，但是这个大任却偏偏降给了一个对存储丝毫不懂的、甚至连计算机基本原理都不懂的人身上——老天捉弄人啊！

2005 年第一次知道有 SAN 这个东西，2006 年就被某外包公司外派到北京某大客户处担任技术顾问。读者看了会纳闷，你那时候就是一个苦 B 屌丝，还能当“顾问”？这一点问得好，其实我也感觉不可思议，现在想想还得感谢那个让我非常苦 B 的外包公司，合同是和人事代理公司签订的，编制算是这个外包公司的，然后这个公司再把我派到他的客户那里，相当于倒了两次手的苦 B。我从这个公司离职的时候，邮箱里面除了一封垃圾邮件，什么都没有，往事不可追，如冷风吹。

但就是这样一家公司，竟敢让我这样一个小白担任所谓“顾问”，我非常佩服他们全球排名靠前的成本控制能力，也很感激能有这个机会。当时我研究和学习计算机网络，是某个比较知名论坛的版主，而恰逢他们那个大客户在研究和模拟一种网络协议，挺复杂搞不懂，需要看看外包服务商那里有没有懂这方面的，开价估计也不菲（现在我也不知道当时到底给我的报酬占客户开价的比例，我估计最多能有 30%），所以公司内部有人推荐了我，霸王硬上弓，就这样我从家乡只身前往帝都了。项目做了一年半，我也成功地完成了任务，把协议从上到下分析地很透彻。当然，我付出的代价就是，早 6 点起来学习，吃完晚饭继续学习。学什么呢？计算机基本原理、通信、网络、协议，当然，还有存储。一个连 $8\text{Bit}=1\text{Byte}$ 都不知道的人，学了点儿网络基本原理，就去给人当顾问，最后还成功完成了任务，听上去挺不可思议的，但就是这样发生了，笨鸟先飞，只要付出，总有回报。

《大话存储》就是在这期间写了一大半，可想而知，写作过程是痛苦的也是值得回味的，大话存储一开始其实就是个笔记，从不懂到慢慢懂，慢慢积累，慢慢提高。无时无刻不在思考，吃饭、睡觉、上厕所，满脑子都是“为什么”，然后就去求证，求证过程是痛苦的，每个材料上描述方式、结论都不一样，这直接影响了自己的判断，这个过程很痛苦。

正是因为自己当时的苦 B 经历，姥姥不疼舅舅不爱，只身一人在北京待了一年半，这一年半里说的话都能数过来有多少句，再加上对知识的原发性兴趣和渴望，让我产生了某种升华，也就是所谓“侠道”，置之死地而后生，闭关苦心修炼，冥思苦想，孤独一世，独孤求败，这正是武侠的情节。我在写作过程中，不知不觉就开始自编自演，仿佛自己就是书中所描写的那些角色、那些遭遇，同时自己也希望有书中那些绝世高手的指导和提携，但实际上没有，我就是一颗无人知道的小草，我就自己提携自己，相当于“自举”，自己激励自己，从而有了那些武侠情节。退一步讲，就算不出现武侠，也可能会出现另一种方式来自我激励，比如咆哮体、自嘲体、暗讽体等等，但是当时毕竟年轻，心态还是积极的，所以武侠也可能是必然的。没有这些情节激励自己，就根本无法坚持下去，何况一个外行来写书，“大任”就得“行拂乱其所为”才能成功，大任，就得去疯狂地做、用异于常规的办法来做，才能成功！当然，不是我故意乱的，而是它必定要疯狂，必定要乱！

有人问我平时是不是个武侠迷，我的回答是：不是，我小学看过射雕英雄传但是基本就是为了看而看，因为那时候电视上播，就想找书看看，但是根本就没看懂。再往后就没看过武侠小说，武侠剧倒是看过一些，比如白眉大侠之类。这也说明了，侠道并不是刻意为之，而是由环境、经历和思考自然产生的一种境界，看武侠的不一定有侠道，能体会侠道的，不一定爱看武侠。我相信对技术有着纯粹追求和渴望的人，心中都有侠道。

看完了上述描述，我想读者如果再看到那些情节，不妨想想当年自己有多苦 B，如果产生了共鸣，我想一定也会理解作者写书时候的苦 B 场景，这样就不至于一开始就被作者的所谓“装 B”封闭了阅读下去的道路了。

最后，借书中一首诗，表达我此时此刻的感想：

七星阵里论七星，
北斗光前参北斗。
不知天上七星侠，
如今过活要饭否？

感谢各位读者，如今我已经没那么惨了，也写不出这种情节了。所谓生于忧患，死于安乐。

冬瓜头 于 北京

目 录

第 1 章 混沌初开——存储系统的前世今生	1
1.1 存储历史	2
1.2 信息、数据和数据存储	6
1.2.1 信息	6
1.2.2 什么是数据	7
1.2.3 数据存储	8
1.3 用计算机来处理信息、保存数据	9
第 2 章 IO 大法——走进计算机 IO 世界	11
2.1 IO 的通路——总线	12
2.2 计算机内部通信	13
2.2.1 IO 总线是否可以看作网络	14
2.2.2 CPU、内存和磁盘之间通过网络来通信	15
2.3 网中之网	16
第 3 章 磁盘大挪移——磁盘原理与技术详解	19
3.1 硬盘结构	20
3.1.1 盘片上的数据组织	21
3.1.2 硬盘控制电路简介	27
3.1.3 磁盘的 IO 单位	28
3.2 磁盘的通俗演绎	29
3.3 磁盘相关高层技术	31
3.3.1 磁盘中的队列技术	31
3.3.2 无序传输技术	32
3.3.3 几种可控磁头扫描方式概论	32
3.3.4 关于磁盘缓存	34
3.3.5 影响磁盘性能的因素	35
3.4 硬盘接口技术	36
3.4.1 IDE 硬盘接口	36
3.4.2 SATA 硬盘接口	39
3.5 SCSI 硬盘接口	42
3.6 磁盘控制器、驱动器控制电路和磁盘控制器驱动程序	50
3.6.1 磁盘控制器	50

3.6.2	驱动器控制电路	50
3.6.3	磁盘控制器驱动程序	50
3.7	内部传输速率和外部传输速率	51
3.7.1	内部传输速率	51
3.7.2	外部传输速率	52
3.8	并行传输和串行传输	52
3.8.1	并行传输	52
3.8.2	串行传输	54
3.9	磁盘的 IOPS 和传输带宽（吞吐量）	54
3.9.1	IOPS	54
3.9.2	传输带宽	56
3.10	固态存储介质和固态硬盘	56
3.10.1	SSD 固态硬盘的硬件组成	57
3.10.2	从 Flash 芯片读取数据的过程	60
3.10.3	向 Flash 芯片中写入数据的过程	61
3.10.4	Flash 芯片的通病	63
3.10.5	NAND 与 NOR	64
3.10.6	SSD 给自己开的五剂良药	65
3.10.7	SSD 如何处理 Cell 损坏	68
3.10.8	SSD 的前景	68
3.11	Memblaze 闪存产品介绍	69
3.11.1	技术能力	70
3.11.2	产品能力	73
3.12	小结：网中有网，网中之网	75
第 4 章	七星北斗——大话/详解七种 RAID	77
4.1	大话七种 RAID 武器	78
4.1.1	RAID 0 阵式	78
4.1.2	RAID 1 阵式	80
4.1.3	RAID 2 阵式	82
4.1.4	RAID 3 阵式	85
4.1.5	RAID 4 阵式	88
4.1.6	RAID 5 阵式	90
4.1.7	RAID 6 阵式	94
4.2	七种 RAID 技术详解	96
4.2.1	RAID 0 技术详析	99

4.2.2 RAID 1 技术详析.....	100
4.2.3 RAID 2 技术详析.....	102
4.2.4 RAID 3 技术详析.....	104
4.2.5 RAID 4 技术详析.....	106
4.2.6 RAID 5 技术详析.....	109
4.2.7 RAID 6 技术详析.....	113
第 5 章 降龙传说——RAID、虚拟磁盘、卷和文件系统实战	115
5.1 操作系统中 RAID 的实现和配置.....	116
5.1.1 Windows Server 2003 高级磁盘管理.....	116
5.1.2 Linux 下软 RAID 配置示例.....	121
5.2 RAID 卡.....	123
5.3 磁盘阵列.....	134
5.3.1 RAID 50.....	134
5.3.2 RAID 10 和 RAID 01.....	135
5.4 虚拟磁盘.....	136
5.4.1 RAID 组的再划分.....	136
5.4.2 同一通道存在多种类型的 RAID 组.....	137
5.4.3 操作系统如何看待逻辑磁盘.....	137
5.4.4 RAID 控制器如何管理逻辑磁盘.....	137
5.5 卷管理层.....	138
5.5.1 有了逻辑盘就万事大吉了么?.....	138
5.5.2 深入卷管理层.....	140
5.5.3 Linux 下配置 LVM 实例.....	141
5.5.4 卷管理软件的实现.....	143
5.5.5 低级 VM 和高级 VM.....	144
5.5.6 VxVM 卷管理软件配置简介.....	146
5.6 大话文件系统.....	149
5.6.1 成何体统——没有规矩的仓库.....	149
5.6.2 慧眼识人——交给下一代去设计.....	150
5.6.3 无孔不入——不浪费一点空间.....	150
5.6.4 一箭双雕——一张图解决两个难题.....	151
5.6.5 宽容似海——设计也要像心胸一样宽.....	153
5.6.6 老将出马——权威发布.....	153
5.6.7 一统江湖——所有操作系统都在用.....	154
5.7 文件系统 IO 方式.....	154

第 6 章 阵列之行——大话磁盘阵列	157
6.1 初露端倪——外置磁盘柜应用探索	158
6.2 精益求精——结合 RAID 卡实现外置磁盘阵列	159
6.3 独立宣言——独立的外部磁盘阵列	161
6.4 双龙戏珠——双控制器的高安全性磁盘阵列	163
6.5 龙头凤尾——连接多个扩展柜	165
6.6 锦上添花——完整功能的模块化磁盘阵列	166
6.7 一脉相承——主机和磁盘阵列本是一家	167
6.8 天罗地网——SAN	169
第 7 章 熟读宝典——系统与系统之间的语言 OSI	171
7.1 人类模型与计算机模型的对比剖析	172
7.1.1 人类模型	172
7.1.2 计算机模型	173
7.1.3 个体间交流是群体进化的动力	174
7.2 系统与系统之间的语言——OSI 初步	174
7.3 OSI 模型的七个层次	176
7.3.1 应用层	176
7.3.2 表示层	176
7.3.3 会话层	176
7.3.4 传输层	177
7.3.5 网络层	177
7.3.6 数据链路层	178
7.3.7 物理层	181
7.4 OSI 与网络	182
第 8 章 勇破难关——Fibre Channel 协议详解	185
8.1 FC 网络——极佳的候选角色	186
8.1.1 物理层	186
8.1.2 链路层	186
8.1.3 网络层	188
8.1.4 传输层	194
8.1.5 上三层	195
8.1.6 小结	195
8.2 FC 协议中的七种端口类型	196
8.2.1 N 端口和 F 端口	196