



云计算与虚拟化丛书

PEARSON

Essential Virtual SAN
Administrator's Guide to VMware VSAN

VMware Virtual SAN 权威指南

[美] Cormac Hogan Duncan Epping 著 徐炯 译
VMware中国研发中心 审校

VMware资深虚拟存储专家亲笔撰写，全球第一本全面、系统讲解Virtual SAN技术的权威著作，Amazon全5星评价

从Virtual SAN的部署、安装、配置到虚拟机存储管理、架构细节和日常管理、维护等方面，深入探讨Virtual SAN的各项技术细节，并用多个实例详细讲解Virtual SAN群集的设计和实现



机械工业出版社
China Machine Press

云计算与虚拟化技术丛书

Essential Virtual SAN
Administrator's Guide to VMware VSAN

VMware Virtual SAN 权威指南

[美] Cormac Hogan Duncan Epping 著 徐炯 译
VMware中国研发中心 审校



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

VMware Virtual SAN 权威指南 / (美) 霍根 (Hogan, C.), (美) 埃平 (Epping, D.) 著; 徐炯译. —北京: 机械工业出版社, 2014.9
(云计算与虚拟化技术丛书)

ISBN 978-7-111-48023-5

I. V… II. ①霍… ②埃… ③徐… III. 虚拟处理机—指南 IV. TP338-62

中国版本图书馆 CIP 数据核字 (2014) 第 211744 号

本书版权登记号: 图字: 01-2014-5107

Authorized translation from the English language edition, entitled, *Essential Virtual SAN: Administrator's Guide to VMware Virtual SAN*, 9780133854992 by Cormac Hogan, Duncan Epping, published by Pearson Education, Inc., Copyright © 2015 VMware, Inc.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Chinese simplified language edition published by Pearson Education Asia Ltd., and China Machine Press Copyright © 2014.

本书中文简体字版由 Pearson Education (培生教育出版集团) 授权机械工业出版社在中华人民共和国境内 (不包括中国台湾地区和香港、澳门特别行政区) 独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

VMware Virtual SAN 权威指南

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 关 敏

印 刷: 北京市荣盛彩色印刷有限公司

开 本: 186mm × 240mm 1/16

书 号: ISBN 978-7-111-48023-5

责任校对: 殷虹

版 次: 2014 年 10 月第 1 版第 1 次印刷

印 张: 14.75

定 价: 59.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

购书热线: (010) 68326294 88379649 68995259

投稿热线: (010) 88379604

读者信箱: hzjsj@hzbook.com

版权所有 · 侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

推荐阅读



高性能CUDA应用设计与开发：方法与最佳实践

作者：（美）Rob Farber ISBN: 978-7-111-40446-0 定价：59.00元



CUDA并行程序设计：GPU编程指南

作者：（美）Shane Cook ISBN: 978-7-111-44861-7 定价：99.00元



CUDA专家手册：GPU编程权威指南

作者：（美）Nicholas Wilt ISBN: 978-7-111-47265-0 定价：85.00元



GPU高性能编程CUDA实战

作者：（美）Jason Sanders 等 ISBN: 978-7-111-32679-3 定价：39.00元



大规模并行处理器程序设计（英文版·第2版）

作者：（美）David B. Kirk 等 ISBN: 978-7-111-41629-6 定价：79.00元

Foreword 推荐序

Virtual SAN 是一个变革性的软件定义存储 (SDS) 产品，它是基于服务器端存储的共享分布式对象存储系统。分布式存储系统在业界和学术界已经发展多年，但是基于 Server 端存储的共享分布式存储产品发展缓慢。高速闪存和高速网络技术的发展与推广使我们能够重新审视传统 SAN 存储和服务器端存储的性能价格优劣及不同的应用场景。Virtual SAN 的成功在于挑战传统，勇于创新。

类似的分布式对象存储系统产品不少（比如 Ceph——我在 UCSC 读博士期间同师门学友的博士论文的产物），但是我认为真正能够荣登企业级 Hypervisor-converged 分布式对象存储产品的目前只有 Virtual SAN 一个。

VMware Virtual SAN 产品和研发团队就像个初创公司。从研发开始到产品正式发布整整四年，中间经历难熬的多个开发、测试和纠错周期，几经反复，一言难尽。曾经有一段时间，为了能够尽早把高质量的产品推出给客户，研发团队 100 多个日夜连续加班赶进度。Virtual SAN 中国研发团队直接参与了 Virtual SAN 1.0 核心技术的研发，并经历了产品发布前那些繁忙和揪心的日子。令人欣慰的是，结果是好的。Virtual SAN 发布时受欢迎程度就像多年前 ESX 刚刚推出时一样，在前三个月已经拥有 300 多个各行各业的客户（客户分布在美国和中国等国家）。在 VMworld 2014 大会上，Virtual SAN 也受到客户和合作伙伴的倾力追捧。客户最喜欢 Virtual SAN 的简单易用 (Simplicity)。这正是产品和研发团队在设计产品时极力追求的简单无上限 (No Limits ——这里指无上限不是无底线) 目标。

在旧金山参加 VMworld 2014 大会期间，我特意约两位作者 Cormac Hogan 和 Duncan Epping 见面，相谈甚欢。两位作者是 VMware 公司著名的技术博主，其中 Cormac 是 VMware 存储架构师，Duncan 是 VMware 云产品的首席架构师。如果没记错的话，这是 Duncan 第 7 次写书 (Cormac 是第一次写对外出版的书)。他们告诉我，这次写书与以往不同，因为这次

写书是基于他们在参与产品开发测试讨论和产品推广过程中（通过写大量的博文）使用 Virtual SAN 的实践——实践是检验好产品的唯一标准。作者在接触和推广试用 Virtual SAN 的过程中，觉得 Virtual SAN 技术太酷太好用了，自然而然就想到要写一本书来为读者介绍这中间的奥秘。

我通读了本书的英文原版书。全书通俗易懂、图片丰富，作者精通存储和分布式系统，对 VMware 和其他存储产品也是造诣颇深。作者不是底层工程师，也没有读过 Virtual SAN 产品的源代码，但是作者在书中对产品和技术的描述及把握适度，没有把本书写成一本纯产品手册或者代码层次的技术书。每一个技术章节，特别是中间的体系结构和技术细节（包括很多数据流图），都经过与 Virtual SAN 资深架构师和首席工程师讨论才拍板定调。

本书中文版的一个特点是它不是简单的“直译”。我见过好多翻译过来的书（包括当前较热门的 SDN、OpenStack 等专题），都是硬邦邦的直译，很多术语晦涩难懂，让人觉得还不如读英文原版。本书译者徐炯对存储有直接使用经验，并对翻译精益求精。在翻译之前和期间，他一直与我们 Virtual SAN 中国研发团队保持及时的沟通，确保书中内容在技术术语的使用上和语意上都能与我们达成一致的意见。

我认为这是一本当下在软件定义存储领域难得的好书。不管你是分布式对象存储新手还是专家，本书都会让你受益匪浅。

成功属于那些敢于承担风险和勇于创新的人！

林才学 博士

VMware 中国研发中心高级研发经理，Virtual SAN 中国研发团队负责人

2014 年 9 月

两周前当本书的翻译接近尾声的时候，2014 年度的 VMworld 大会在美国旧金山开幕了，会上传来了令人震惊的消息，VMware 公司推出了超融合基础架构 EVO:RAIL 和 EVO:RACK。

“融合”这个词这些年来越来越热。是呀，原本各自为政甚至是井水不犯河水的网络、服务器和存储开始相互渗透合并，其速度之快趋势之猛以至于连“融合”这个词本身都 hold 不住了，非得要搬出“超融合”才能压得住阵脚。VMware 公司的 Virtual SAN 就是 VMware 公司的这个超融合架构的核心技术。

几个月前接手这本关于 VMware Virtual SAN (VSAN) 技术的最新图书的翻译工作的时候，我就已经深深感受到这种技术给我们带来的震撼。作为一家纳斯达克上市的美国公司在中国工厂的 IT 技术经理，我也同时负责中国工厂的基础架构和数据中心的一些架构设计工作。去年（2013 年）我们才部署完毕 FlexPod，这是 Cisco+NetApp+VMware 合作推出的统一计算的架构，它非常巧妙地利用思科 UCS 技术将服务器的所有配置都剥离出来放在配置文件中，使得服务器纯粹成为提供计算资源和内存资源的盒子。但是 FlexPod 仍然利用了 NetApp 的共享的集中式存储，就好像 VCE 还是需要利用 EMC 的集中式存储一样。集中式存储的最大缺点就是贵，而 VSAN 利用服务器的本地存储和闪存加速技术，带给用户高性能的同时将存储成本降低到了普通机架式服务器硬盘的价位。这种低价的冲击影响是巨大的，甚至是致命的。

正如 Charles Fan 在为本书写序时说的那样，这对存储企业来说“是一种破坏性的创新”。随着企业内几乎所有的应用包括关键业务应用都随着服务器虚拟化集中到了 VMware 平台上，企业级集中式存储的应用场景也就慢慢集中到为 VMware 提供存储。此时，完全可以用 VSAN 来取代集中式存储。几年之后，目前这种中小型企业级集中式存储甚至可能会因此而被市场所彻底摒弃。

非常高兴能有机会对一本介绍前沿技术的书籍进行翻译。接到这项任务的时候，本书的英文版本还没有完全定稿（英文版于今年 8 月出版），于是我也获得了很多与本书作者 Cormac

Hogan 和 Duncan Epping 沟通的机会。这两位作者都是虚拟化领域的大拿，各自拥有自己的博客网站，发表过很多极有价值的博客文章。多年前我就是他们博客的订阅者。能有机会和自己的偶像合作非常令人激动和愉快。翻译本书的过程也是我学习的过程，在这个过程中，两位作者给予了我很多支持，在此表示深深的感谢。

还要感谢华章公司的关敏编辑和王春华编辑，她们是把这本书交到我手里的红娘，不仅如此，她们的认真仔细帮助我纠正了很多错误。此外，我也获得了 VMware 中国研发中心林才学博士及其领导下的开发团队的技术支持，在此一并谢过。

翻译本书的过程既是愉快的也是痛苦的，愉快是因为新知识的收获，痛苦是因为本书的翻译是在繁忙的工作之余利用挤出来的个人时间完成的，时间紧张而我又是拖延症患者，因此常常被小鞭子逼着抽着去做这项文字工作。所以，最后的最后，要深深地感谢我的妻子刘峥嵘同学，不是你的小鞭子时时敲打，我是不可能及时完成这项任务的。

徐炯

2014年9月8日中秋夜

2012年早些时候我加入了VMware公司，当时我很荣幸地受命领军交付下一代的vSphere——我们的旗舰产品，这个荣誉让我既羞愧又惊喜。进入角色几个月后，存储组加入团队中，我很荣幸能和这样一群敬业的工程师紧密合作。他们正在做的东西非常特别——相信这将是存储历史上一个重要的转折点。

我们开始打造一款分布式的可容错的专门为虚拟环境优化的存储系统。我们的目标是构建一个具备所有共享存储品质（弹性、性能、可扩展性等）的产品，但是这个产品不需要特殊的硬件也不需要专门的软件来维护，可以直接运行在x86服务器上。只需要插入硬盘和SSD，vSphere会搞定剩下的一切。加上基于策略的管理框架和新的运营模型，存储管理将变得前所未有的简单。

我们遇到了很多困难——就像所有长期的软件项目一样，长夜漫漫、士气低落、主次纠葛、计划多变。尽管如此，我们坚持下来了。2013年6月特别痛苦，当时团队已经准备好发布vSphere 5.5了，但是我不得不告诉团队还不能交付VSAN，相反，在宣布产品“可以商用”之前，它们必须经过更广泛的公开beta测试和更多严苛的实验。

风险实在太高，特别是因为这是VMware在软件定义的存储领域的一次突袭，是我们软件定义的数据中心愿景的关键组成部分。

当然，他们很失望。因为我们不能在VMworld的舞台上展示这个产品，但是大家坚持了下来。我觉得我们的选择是正确的。经过了6个月和12 000名beta测试员的不懈努力，Virtual SAN终于完成了：它可靠、成熟，做好了登台表演的准备。VSAN可以从用于小型分支办公室的最少3个节点的配置，横向扩展到数TB存储、上百万IOPS的可以支持整个企业所有存储需求的庞然大物。

我们的团队交付的是业界真正特别优秀的产品：一个完全分布式的与hypervisor无缝融合的存储架构。

VSAN 不是外挂，它是内核。

关于 VSAN 已经有了很多文章，都提到它将给存储世界带来一种全新的理念。然而本书很不一般。Duncan 和 Cormac 在整个项目过程中始终与我们的开发团队密切合作。他们不仅极度熟悉 VSAN 架构，也亲手进行过大规模的部署和管理。他们是专家。

Ben Fathi
VMware CTO

今年的早些时候，我有幸参加了克莱顿·克里斯坦森（Clayton Christensen）的一个研讨会。他的极具创意的作品——《创新者的困境》（*Innovator's Dilemma*）是我最喜欢的商业读本之一，因此能有机会亲耳聆听克莱顿本人的深刻见解实在是太棒了。整个研讨会中我都有一个幻觉，好像房间里面没有其他人，只有克莱顿和我，而且我们一直在探讨 Virtual SAN。

讨论的主题是——VSAN 到底是一种破坏性创新还是维持性创新？

在克莱顿的书里面是这样定义的：维持性创新是一种使事物更好、更快、更强大的技术进步，可以满足客户不断增长的需求。维持性创新不需要改变当前的业务模式、业务流程或目标客户。这是大公司如何变得更大的方法。只要给它们足够的资源和客户关系，它们总能通过维持性创新战胜较小的公司。

然而，总是会发生某些技术进步超过了客户需求发展的情况。此时，创新来自于底部。这些创新通过提供一种不同的方式来解决这个问题。在开始的时候它们可能无法提供同样水准的特性或性能，但是它们更便宜、更简单，常常会带来更多的和不同领域的客户，有时候它们会彻底改变业务模式。这就是破坏性创新。对于现任的市场领导者来说，采纳破坏性创新极为困难。这种类型的创新会重新定义整个业界，新的领导者由此而诞生。

那么，VSAN 是一种破坏性创新还是维持性创新呢？这个问题看上去好像很傻，当然这是一种破坏性创新。它是一种极度简单的、纯粹软件化的、融合在 hypervisor 中的分布式存储解决方案，它完全和 vSphere 整合在了一起，可以运行在普通商用服务器上。它对存储的经济模型和消费模型都做出了重新定义。尽管它还缺少（就目前而言）一些典型的存储的特性和优点，但是它提供了比传统企业级存储阵列更为简单易用的品质，以更低的价格卖给一群与以往不同的存储管理员用户。因此，这是一种典型的破坏性创新，就好像傻瓜相机一样。比起“真正的”相机，傻瓜相机开始只有很少的功能，但是它们却非常简单，目标客户是完全不同

的一群用户。你猜后来怎样了？很快傻瓜相机的用户数量就超过了传统相机。

那为什么会有这样的问题？作为一款存储产品，是的，VSAN 毫无疑问是一款革命性的产品，它会颠覆整个存储工业并将其推向一个新的时代。然而，如果换一个角度来看，作为 vSphere 服务器虚拟化平台的自然延伸，进化成软件定义的数据中心，这是一种维持性创新。VSAN 仍然卖给同样的 vSphere 客户，并帮助他们实现更多的功能。它将服务器环境扩展成一个融合性的基础架构，将 vSphere 抽象层通过基于策略的自动化从计算领域扩展到了存储领域。

因此，我们有一半胜算掌握在我们自己手里。这是一个维持性创新和破坏性创新的组合。一方面它构筑在我们的 VMware 原生的 hypervisor 平台之上，扩展了提供给客户的价值，另一方面它同时是一种将要重塑存储工业的破坏性创新，这个产品对存储造成的影响就像当年 vSphere 对服务器造成的一样。

VSAN 产品是整个 VSAN 产品团队过去 4 年辛勤劳动的结果。这个团队不仅仅包括核心架构师、程序员、测试员和产品经理，Duncan 和 Cormac 也是团队的 2 个关键人物。他们带来了真实世界的经验和客户的想法，他俩也是将我们的想法带到全世界的两个最有力的声音。我非常高兴他们能及时推出这本书，希望你们能和我一样觉得它非常有用。VSAN 是一个非常特别的产品，它会给整个业界带来持续的影响。欢迎你加入我们这个令人兴奋的旅程中。

Charles Fan

VMware 研发、存储和可用性高级副总裁

说到虚拟化及其依赖的底层基础架构，经常会提起一个组件——存储。原因相当简单：在很多环境中，存储是痛点。尽管存储市场已经因为闪存技术的引入发生了变化，很多传统的存储问题得到了缓解，但是很多机构还没能采纳这些新的架构，因而仍然会遇到挑战。

存储问题的范围包括运营上的复杂性到性能问题甚至是可用性的限制。这些问题中的大部分都起因于同样的根本问题：老旧的系统架构。这是因为大多数存储平台架构是在虚拟化技术出现之前开发出来的，而虚拟化已经改变了使用这些共享存储平台的方法。

某种程度上，可以说是虚拟化迫使存储业界去寻找新的方法来构建存储系统。不再是通过单台服务器连接到单台存储设备（也称为逻辑单元或简称为 LUN），虚拟化通常由一台（或多台）物理服务器承载很多虚拟机连接到一个或多个存储设备上。这不仅仅增加了这些存储系统的负载，也改变了工作负载的模式并增加了对总容量的需求。

可以想象，对于大多数存储管理员来说，这要求思考模式的大改变。LUN 的大小应该是多少？对性能有什么要求？最终需要多少个磁盘？这些 LUN 将提供何种数据服务？虚拟机将存放在什么地方？不仅思考模式要改变，而且要求和其他 IT 团队协力合作。过去服务器管理员、网络管理员和存储管理员都可以活在他们自己的独立的小天地内，现在他们需要相互沟通并齐心协力才能保证他们构建的平台的高可用性。在过去，一个错误（例如错误配置或过低的置备）只会影响一台服务器，现在则会影响很多虚拟机。

当虚拟化出现时，我们对于如何运营和构建 IT 基础架构的思维曾经发生过集体性的根本变化。如今集体性的转变再次发生，这一次是由软件定义的网络和软件定义的存储引起的。但我们不应该再重复历史，重复那些在虚拟化首次出现时我们曾经犯过的错误。我们应该坦率而开放地和数据中心管理员们讨论这个问题，并一起迎接数据中心架构和运营的革命。

写作本书的动机

我俩很早就都参与进了 VSAN 的产品开发，并立刻意识到这是每个人都会提起的产品，人们会想要了解更多的东西。在各种美妙的来来往往的交谈和探讨中，我们认识到这些信息不曾被记录下来过。考虑到我俩都是狂热的博客作者，我们决定各自独立地开始撰写文章。很快我们就攒起了大量的材料，多到不可能将它们用博客的方式都发表出来，于是我们决定合力出版一本书。初步接触后，VMware 出版社愿意出版这本书。

读者对象

本书的目标读者是和 VMware vSphere 环境相关的 IT 专业人员。你最好已经用过一阵子 VMware vSphere，或许已经参加过 vSphere 的课程，例如“安装、配置和管理”课程。本书不是一本初学者读物，但是书中提供的信息应该已经足够覆盖各种不同水平的管理员和架构师。

如何使用本书

本书分 10 章，分别如下：

- ❑ 第 1 章概要介绍了软件定义的存储和 VSAN。
- ❑ 第 2 章从物理和虚拟的角度描述了对 VSAN 安全地进行实施所需的要求。
- ❑ 第 3 章介绍了安装和配置 VSAN 的各个步骤。
- ❑ 第 4 章介绍了基于策略的存储管理。
- ❑ 第 5 章深入介绍了 VSAN 的架构细节。
- ❑ 第 6 章描述了虚拟机存储策略是如何用来简化虚拟机部署的。
- ❑ 第 7 章描述了常用的管理和维护任务的步骤。
- ❑ 第 8 章覆盖了 Virtual SAN 和其他 VMware 功能及产品之间的互操作性。
- ❑ 第 9 章提供了多个例子来介绍如何设计一个 VSAN 群集，包含了一些容量规划的练习。
- ❑ 第 10 章覆盖了各种可用于进行 VSAN 排错和监控的（命令行）工具。

致谢

我们两个作者都在 VMware 公司工作。在本书中表达的意见都是我们根据自己对产品的经验表达的个人意见。本书中的陈述不一定反映出 VMware 公司的意见和观点。

我们要感谢 Christos Karamanolis 和 Paudie O’Riordan，作为我们的技术编辑，他们始终让我们保持坦诚的态度。当然，我们要感谢 Virtual SAN 工程团队，特别要指出 2 个名字：

Christian Dickmann 和 Christos Karamanolis (再次感谢!), 他们在 VSAN 方面的渊博知识和对 VSAN 的深入理解贯穿到了本书的各个角落。我们还想感谢 William Lam、Wade Holmes、Rawlinson Rivera、Simon Todd、Alan Renouf 和 Jad El-Zein 对本书提供的帮助和贡献。

最后, 我们想感谢 VMware 公司的管理团队 (Phil Weiss、Adam Zimman 和 Mornay van der Walt) 对本书和其他项目的支持。

VSAN 加油!

Cormac Horgan 和 Duncan Epping

关于作者 *About the Authors*

Cormac Hogan 是 VMware 集成工程团队的存储架构师。他是位于爱尔兰科克市的 VMware EMEA[⊖] 总部 2005 年的第一批雇员之一，曾经在 VMware 技术市场部门和支持部门任职。Cormac 撰写过很多存储相关的白皮书，并做过大量关于存储最佳实践和新特性方面的演讲。Cormac 是 CormacHogan.com 网站的站长，这是一个专注在存储和虚拟化方面的博客网站。

可以通过 twitter 关注他：@CormacJHogan。

Duncan Epping 是 VMware 研发中心的首席架构师。他负责在现有产品和功能中挖掘新的潜力，并为 VMware 寻找新的业务机会。Duncan 专门研究软件定义的存储、hypervisor 融合平台以及高可用解决方案。Duncan 是最早的 VMware 认证设计专家之一（VCDX007），他是 Yellow-Bricks.com 网站的站长和多本著作的作者，包括 VMware vSphere 群集技术深入探讨（*VMware vSphere Clustering Technical Deepdive*）系列。

可以通过 twitter 关注他：@DuncanYB。

⊖ EMEA 指的是 Europe, Middle East and Africa。——译者注