



全球资深虚拟化技术专家、VMware高级主管工程师撰写，官方出品，权威性毋庸置疑

VMware存储领域最详尽的一本著作，不仅深入探讨了存储技术的原理和VMware存储的各项技术细节，还详细讲解了vSphere存储系统的设计与实现

华章程序员书库

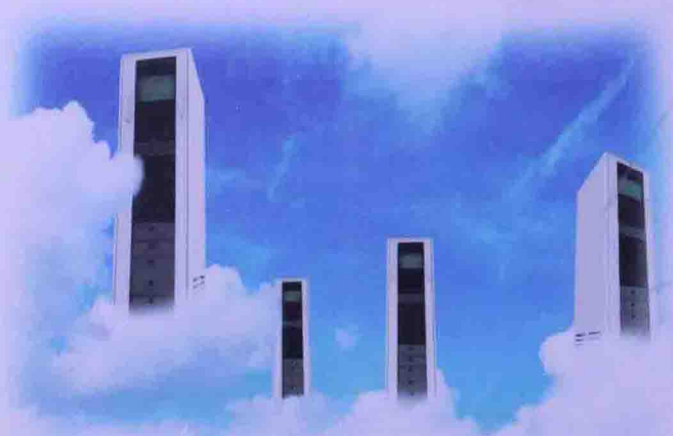
Storage Implementation in vSphere 5.0

vSphere 5.0

存储设计与实现

(美) Mostafa Khalil 著

姚军 等译



机械工业出版社
China Machine Press

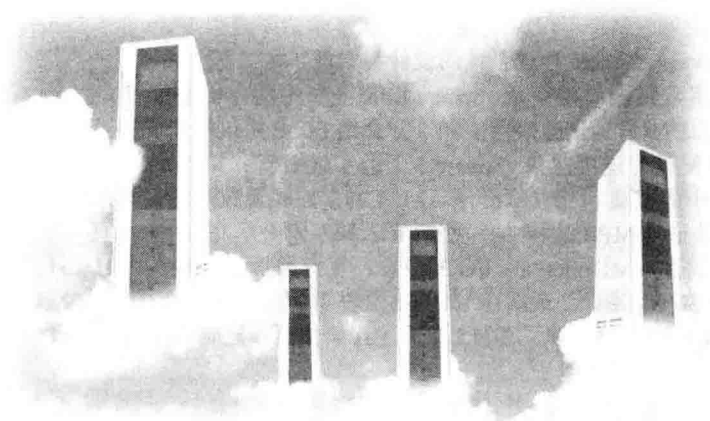
Storage Implementation in vSphere 5.0

vSphere 5.0

存储设计与实现

(美) Mostafa Khalil 著

姚军 等译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

vSphere 5.0 存储设计与实现 / (美) 哈里尔 (Khalil, M.) 著; 姚军等译. —北京: 机械工业出版社, 2013.9
(华章程序员书库)

书名原文: Storage Implementation in vSphere 5.0

ISBN 978-7-111-43866-3

I. v… II. ①哈… ②姚… III. 虚拟处理机 IV. TP338

中国版本图书馆 CIP 数据核字 (2013) 第 207226 号

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

本书版权登记号: 图字: 01-2012-7685

Authorized translation from the English language edition, entitled *Storage Implementation in vSphere 5.0*, 978-0-321-79993-7 by Mostafa Khalil, published by Pearson Education, Inc., Publishing as VMware Press. Copyright © 2013.

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 © 2013.

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。

本书是 VMware vSphere 虚拟化存储领域最为详尽和权威的一部作品, 它由全球资深虚拟化技术专家、VMware 高级主管工程师撰写。书中不仅深入探讨了存储技术的原理和 VMware vSphere 存储的各项技术细节, 还详细讲解了 VMware vSphere 存储系统的设计与实现。

本书共 16 章。第 1 章介绍常见存储类型及 vSphere ESXi 5 所用的存储; 第 2 章讨论 FC 协议和 FC SAN, 以及如何设计无单故障点的光纤通道连接性; 第 3 章介绍 FCoE 协议的细节和架构及其在 vSphere 5 中的实现; 第 4 章讨论 iSCSI 协议、连接性和 vSphere 5 上的实施细节; 第 5 章介绍 PSA 组件; 第 6 章介绍 ALUA 标准在 vSphere 5 中的实现; 第 7 章介绍 ALUA 和多路径及故障切换的相互作用; 第 8 章介绍 vSphere 5 在撰写本书期间认证的 MPIO 插件的一些细节; 第 9 章介绍在 vSphere 环境中使用异构存储配置的原因和方法; 第 10 章讲述 VMDirectPath I/O 的概念、工作原理和一些实用的设计实施方案; 第 11 章从 vSphere 5 的角度介绍 SVD; 第 12 章阐述 VMFS5 文件系统的历史、架构和恢复技巧; 第 13 章介绍虚拟磁盘和 RDM; 第 14 章讨论分布锁; 第 15 章讲述存储快照、复制和镜像; 第 16 章给出 VAAI 的有关细节。

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: 谢晓芳

三河市杨庄长鸣印刷装订厂印刷

2013 年 10 月第 1 版第 1 次印刷

186mm×240mm·26.25 印张

标准书号: ISBN 978-7-111-43866-3

定 价: 99.00 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

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsj@hzbook.com

译者序

在世界各个大型数据中心，虚拟化正在日益成为主流，它不仅能够用来为企业优化硬件配备，帮助企业节约成本，而且会给管理带来许多便利。

在虚拟化环境中，存储是整个结构的核心，它不仅为海量的数据提供空间，而且是所有虚拟及相关配置、信息的栖身之所，存储的设计对于整个虚拟环境的性能、管理、可用性等都是至关重要的，而存储技术的不断发展，催生了各种存储协议、存储设备和新的存储技术，存储的实现和管理也日益变得复杂。

正因为如此，每个虚拟化环境的管理人员都盼望着有一本全面、详尽的指南，讲述各种技术、协议以及虚拟化操作系统中管理与实施方面的知识，本书正是这样一本著作。

Mostafa Khalil 是 VMware 公司的资深工程师，拥有 10 多年虚拟环境实施、故障检修经验，在本书中，他以大量的插图、详尽的规程，介绍了 VMware 行业领先的最新虚拟化产品 vSphere 5.0 中的存储架构实施和管理。书中不仅涵盖了各种流行存储协议、产品的详细介绍，而且对 vSphere 5.0 中的新特性，如 VAAI、PSA 等做了重点介绍。利用书中的知识，我们可以设计出具有高可用性、高性能的存储架构，并可以轻松地利用 CLI、UI 管理整个存储环境。毫不夸张地说，对于 VMware 存储环境，本书是时下最新、最全面的规划管理宝典。

在本书的翻译过程中，我们从书中领略了虚拟化环境的许多最佳实践，对 VMware 虚拟环境有了更深的理解，我们也盼望着广大读者能够从本书中受益。

本书的翻译工作主要由姚军完成，徐锋、陈绍继、郑端、吴兰陟、施游、林起浪、陈志勇、刘建林、白龙、林耀成、陈霞、方翊、宁懿等人也为本书的翻译工作做出了贡献。在此也要感谢机械工业出版社的编辑们对翻译工作提出了许多中肯的意见，同时期待着广大读者朋友能够批评指正。

译者

前言

本书第1版涵盖我多年的VMware产品支持经验以及从VMware团队吸取的知识，也是我第一次尝试把所有最佳实践汇集到一起。我和读者分享了深入的细节，能够帮助你在出现问题的时候找出症结所在。我原来计划在一本书中介绍所有的知识，但是在我开始写作之后，篇幅不断增长，部分原因是书中包括大量的插图和屏幕截图，我希望用这种形式让你更加清晰地理解内容。结果是，我不得不将本书分成两卷，才不会因为减少篇幅而牺牲质量。我希望你能觉得这些内容和我自己的本意一样实用，并且留意本书第二卷。

本书从存储历史的简单介绍开始，然后提供各种存储连接选择和VMware支持的协议的细节：光纤通道（FC）、以太网上的光纤通道（FCoE）和互联网小型计算机系统接口（iSCSI）。从这里我们过渡到vSphere存储的基础——PSA。由此，我以多路径和故障切换（包括第三方提供的方案）以及ALUA作为基础。接下来，讨论存储虚拟设备（SVD）和VMDirectPath I/O架构、实施和配置。还介绍了虚拟机文件系统（VMFS）版本3和5错综复杂的细节，以及这种高级的群集文件系统对虚拟机文件并行访问的仲裁方法以及原始设备映射。然后讨论分布式锁的处理方法以及物理快照和虚拟机快照。最后，讨论了陈列集成vStorage API（vStorage API for Array Integration, VAAI）架构和相关存储阵列的交互方法。

考虑到这一卷是更高级内容的开始部分，我计划将内容更新为vSphere 5.1，该版本的名称为VMware云架构套件（Cloud Infrastructure Suite, CIS），并添加针对设计主题和性能优化的更多信息。

衷心希望听到你对书中主题的意见和建议。你可以在我的博客上留言：<http://vSphereStorage.com>。

致谢

我要感谢妻子Gloria的支持，还要感谢VMware全球支持服务高级副总裁Scot Bajtos和VMware全球支持服务（美洲）副总裁Eric Wansong的鼓励。

我要衷心感谢那些在百忙之中抽出时间审校本书稿件的人对本书的反馈：

Craig Risinger, VMware咨询架构师

Mike Panas, VMware 技术组高级成员

AboubacarDiar, HP Storage

Vaughn Stewart, NetApp

Jonathan van Meter

特别感谢 VMware 高级技术营销架构师 Cormac Hogan 允许我使用他的插图。

还要感谢 Pearson 的技术审校者（我只知道他们名字的首字母）以及编辑 John Murray 和 Ellie Bru，感谢他们帮助我完成本书。

最后的感谢留给一路上教育和培养我的所有人，名字太多，无法一一列出。感谢你们所有人！

目 录

译者序

前言

3.13 关键提示 / 58

3.14 小结 / 59

第 1 章 存储类型 / 1

1.1 存储的历史 / 1

1.2 小结 / 7

第 2 章 光纤通道存储连接性 / 8

2.1 SCSI 标准和协议 / 8

2.2 小结 / 34

第 3 章 FCoE 存储连接性 / 35

3.1 FCoE / 35

3.2 FCoE 初始化协议 / 36

3.3 FCoE 发起方 / 38

3.4 克服以太网的局限性 / 39

3.5 FCoE 所需的协议 / 40

3.6 802.1p 标记 / 42

3.7 硬件 FCoE 适配器 / 43

3.8 在 ESXi 5 中如何实现软件 FCoE / 44

3.9 配置 FCoE 网络连接 / 45

3.10 启用软件 FCoE 适配器 / 48

3.11 删除或者禁用软件 FCoE 适配器 / 50

3.12 FCoE 故障检修 / 52

第 4 章 iSCSI 存储连接性 / 60

4.1 iSCSI 协议 / 60

4.2 小结 / 113

第 5 章 vSphere PSA / 114

5.1 原生多路径 / 114

5.2 SATP / 115

5.3 PSP / 116

5.4 第三方插件 / 117

5.5 多路径插件 (MPP) / 118

5.6 PSA 组件剖析 / 118

5.7 通过 PSA 和 NMP 的 I/O 流 / 120

5.8 列出多路径细节 / 123

5.9 声明规则 / 132

5.10 MP 声明规则 / 133

5.11 插件注册 / 135

5.12 SATP 声明规则 / 135

5.13 用 UI 修改 PSA 插件配置 / 139

5.14 用 CLI 修改 PSA 插件 / 142

5.15 小结 / 158

第 6 章 ALUA / 159

6.1 ALUA 定义 / 159

6.2 小结 / 174

第 7 章 多路径和故障切换 / 175

7.1 什么是路径 / 175

7.2 活动路径在哪里 / 179

7.3 LUN 发现和路径枚举 / 182

7.4 LUN 发现和路径枚举日志条目
示例 / 183

7.5 影响多路径的因素 / 186

7.6 故障切换触发器 / 189

7.7 路径状态 / 193

7.8 路径选择插件 / 195

7.9 何时、如何改变默认 PSP / 196

7.10 PDL 和 APD / 198

7.11 路径排名 / 207

7.12 小结 / 211

第 8 章 第三方多路径 I/O 插件 / 212

8.1 MPIO 在 vSphere 5 上的实现 / 212

8.2 EMC PowerPath/VE 5.7 / 212

8.3 HDLM / 225

8.4 Dell EqualLogic PSP Routed / 234

8.5 小结 / 237

第 9 章 使用异构存储配置 / 238

9.1 “异构”存储环境 / 238

9.2 异构存储方案 / 238

9.3 ESXi 5 异构存储视图 / 239

9.4 小结 / 246

第 10 章 使用 VMDirectPath I/O / 247

10.1 VMDirectPath / 247

10.2 VMDirectPath 支持哪些 I/O
设备 / 247

10.3 VMDirectPath I/O 配置 / 249

10.4 利用 VMDirectPath I/O 的
VM 设计方案实例 / 256

10.5 vMDirectPath Gen. 2 / 257

10.6 VMDirectPath I/O 故障检修 / 260

10.7 小结 / 262

第 11 章 SVD / 263

11.1 SVD 概念 / 263

11.2 小结 / 270

第 12 章 VMFS 架构 / 271

12.1 VMFS 的历史 / 271

12.2 小结 / 308

第 13 章 虚拟磁盘和 RDM / 309

13.1 概述 / 309

13.2 虚拟磁盘 / 309

13.3 用 UI 创建虚拟磁盘 / 313

13.4 用 vmkfstools 创建虚拟磁盘 / 316

13.5 原始设备映射 / 322

13.6 列出 RDM 属性 / 327

13.7 虚拟存储适配器 / 332

13.8 虚拟机快照 / 335

13.9 快照操作 / 343

13.10 恢复快照 / 349

13.11 链接克隆 / 350

13.12 小结 / 351

第 14 章 分布式锁 / 352

14.1 基本锁 / 352

- 14.2 小结 / 367

第 15 章 快照处理 / 368

- 15.1 什么是快照 / 368
- 15.2 什么是复制 / 368
- 15.3 什么是镜像 / 368
- 15.4 VMFS 签名 / 369
- 15.5 快照对 VMFS 签名的影响 / 370
- 15.6 如何处理快照 LUN 上的 VMFS 数据存储 / 371
- 15.7 重新签名 / 371
- 15.8 强制安装 / 376
- 15.9 小结 / 381

第 16 章 VAAI / 382

- 16.1 什么是 VAAI / 382
- 16.2 VAAI 原语 / 382

- 16.3 全拷贝原语 / 383
- 16.4 块置零原语 / 384
- 16.5 硬件加速的加锁原语 / 384
- 16.6 精简配置 API / 385
- 16.7 NAS VAAI 原语 / 386
- 16.8 启用和禁用原语 / 387
- 16.9 VAAI 插件和 VAAI 过滤器 / 393
- 16.10 列出注册的过滤器和 VAAI 插件 / 396
- 16.11 列出 VAAI 过滤器和插件配置 / 397
- 16.12 列出 VAAI vmkernel 模块 / 400
- 16.13 确定设备支持的 VAAI 原语 / 400
- 16.14 用 ESXTOP 显示块设备 VAAI I/O 统计 / 404
- 16.15 VAAI T10 标准命令 / 406
- 16.16 VAAI 原语故障检修 / 407
- 16.17 小结 / 409

第1章 存储类型

1.1 存储的历史

历史上，存储的概念可以上溯到古埃及人建造地窖避免谷物发霉，把谷物储藏起来以备饥荒时使用。著名的例子是约瑟的故事：他监督谷物的配给和储存，并记录总数。经过很多代人和国家的发展，现在我们存储数据。古埃及的历史讲得太多了，让我们回到正题：说一说电脑存储的历史！

从计算机的角度看，最小的数据单位是“位”（bit），数值是0或者1。你可能已经知道，位来源于最早的小型计算机架构，这种架构的形式是处于两个位置之一的拨动开关：关（0）或者开（1），这意味着，它使用的是所谓的二进制数据。8位组成一个字符，称为一个数据字节（byte）。

图 1.1 展示了早期的模型计算机，利用拨动开关来对二进制位 0 或 1 编程。

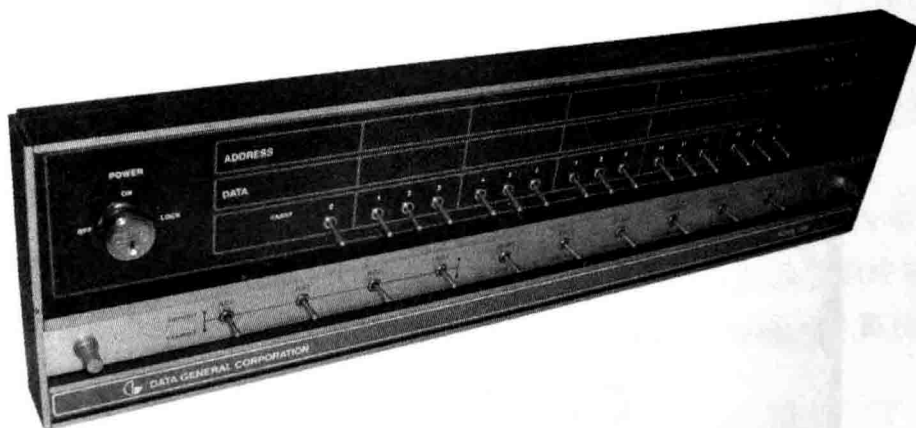


图 1.1 Data General Nova 1200 小型计算机显示了一个 16 位设计（16 个拨动开关）

照片来源：Arnold Reinhold，在 GNU 免费许可证下使用

（细节：Data General 公司是 EMC 的 CLARiiON® 存储阵列家族的创造者。）

所以，计算机处理的任何数据都是 0 和 1 的序列。1.1.4 节会讨论数据和地址空间计量单位。数据存储于易失性和永久性媒体上。

易失性数据存储也称作随机存取存储器（Random Access Memory, RAM），从 RAM 读取数据比从其他形式的存储中读取快得多，因为它们没有移动部件。这类存储被计算机用来加载程序和数据供运行时使用。在 RAM 中作了修改的数据在计算机系统关闭之前的某个时间间隔内被写入永久性存储器。

永久性数据存储媒体根据存储或者读取的数据类型和连接的控制器而有所不同。最早的数据存储形式是大型主机系统使用的磁带。数据被写入磁道中，这多少有些像 20 世纪 70 年代因为美国大部分汽车上的立体声磁带播放机而流行起来的旧式 8 轨音频磁带格式。这些磁道沿着磁带长度方向平行延伸，由与磁道数量相同的磁头读出。20 世纪 50 年代的大型主机使用宽度为 0.5 英寸的 10.5 英寸磁带卷。这种磁带已经演化为现代个人电脑用于数据备份的 1/4 英寸盒式磁带。

之后产生的可移动永久性数据存储形式是从 8 英寸一直到 2 英寸的软盘。5.25 英寸曾经是最流行的尺寸，而后是 3.5 英寸。

（细节：IBM PC 最初型号使用的 DOS 1.0（磁盘操作系统）装载在小型磁带以及 5.25 英寸软盘上。）

IBM PC（型号 5150）带有两个 5.25 英寸的软盘驱动器（见图 1.2）。后续的型号（5160 或者 XT）带有一个 10MB MFM 硬盘。这种硬盘的规格是 5.25 英寸全高（也就是说，占据完整的一个插槽），代替第二个磁盘驱动器。我仍然收藏着这两种系统，而且认为在加州山景城（Mountain View）的计算机历史博物馆中还展示着一台或者多台这样的系统。



图 1.2 IBM PC（5150 型）带有两个 5.25 英寸软盘驱动器
资料来源：维基百科（关闭的文档）

1.1.1 硬盘的诞生

随着应用程序的增大，对永久性存储的内部结构的需求越来越大，这些存储称为硬盘或者硬盘驱动器。“硬”这个词的使用是因为这些磁盘与软盘和磁带相比较不灵活。它们最先出现在 IBM PC/XT（5160 型）上。它有 10 MB 的大容量，直径为 5.25 英寸，平均查找时间为 150 毫秒。在 20 世纪 80 年代下半叶，查找时间降低到 1/10，大约为 15 毫秒。

图 1.2 中的 IBM PC/XT 配置了一个 MFM（Modified Frequency Modulation，改进频率调制）ST-506 接口。20 世纪 80 年代末到 20 世纪 90 年代初，长度限制编码（Run Length Limited, RLL）ST-412 接口也得到使用。后者要求安装特殊的软件，提供 BIOS 扩展，以扩充逻辑块寻址（Logical Block Addressing, LBA）。

还有一些其他形式的永久性存储和硬盘驱动器，如增强的小型磁盘接口（Enhanced Small Disk Interface, ESDI），该接口在 20 世纪 80 年代末的 AT&T Unix 以及 XENIX 操作系统早期版本中相当常见。

1.1.2 SCSI 的出现

PC 中对更大、更快和更多磁盘的需求催生了小型计算机系统接口（Small Computer System Interface, SCSI）。已经证明这种接口是目前为止最成功和最可靠的接口。它成为连接各类 I/O 设备的行业标准，包括：扫描仪、打印机、磁带驱动器和存储设备，以及磁盘阵列机箱（Disk Array Enclosure, DAE）中的大量磁盘。近来最常用的 SCSI 设备是磁盘和磁带驱动器或者磁带库。SCSI 协议和标准在 2.1 节、第 3 章和第 4 章中介绍。

SCSI 磁盘及其缓冲区已经变得更大更快。这一接口也从并行发展为串行连接 SCSI（Serial Attached SCSI, SAS）。下面是简化的 SAS 概念：在控制器和电源终端之间不使用磁盘菊花链，而是将磁盘连接到控制器上的专用通道，或者插入连接到计算机外部 SAS 控制器通道的外部 SAS 磁盘机箱中。

1.1.3 PATA 和 SATA——SCSI 的远亲

推出 IBM PC/AT 计算机之后，下一代的磁盘常称作 IDE 磁盘。这些磁盘使用的电子集成驱动器（Integrated Drive Electronics, IDE）接口实际上是 AT 嵌入式接口 /AT 附加分组接口（AT Attachment/AT Attachment Packet），后来它改名为并行 ATA——PATA，以便和新一代的串行 ATA——SATA 区分。

PATA 局限于每个控制器接口（主接口和从接口）两个驱动器，而 SATA 受到控制器提供的通道数量限制。

下面是 SAS 和 SATA 的一些不同。

- SCSI 驱动器比 SATA 驱动器快得多，也昂贵得多，这是因为其设计和该技术实现的

性能。

- SCSI 使用标记命令队列实现，允许保留许多命令。这种设计明显提高了驱动器 / 控制器的性能，它们可以重新排列命令的顺序，尽可能达到最优的执行方式。
- SCSI 驱动器还使用一个处理器执行命令和处理接口，同时使用一个单独的处理器通过伺服系统处理磁头定位。
- SCSI 磁盘可以用于包含 ECC（Error Checking and Correcting，错误检查和校正）缓存的存储阵列。这对于数据完整性来说很重要，尤其是在 vSphere 环境中。
- SATA 1.0 使用任务命令队列（Task Command Queuing，TCQ）。这种队列技术旨在帮助弥补 SCSI 和 PATA/SATA 驱动器之间的差距；然而，它的开销相当高而且不够有效。TCQ 也称作中断队列。
- SATA 2.0 引入了原生命令队列（Native Command Queuing，NCQ），这和 SCSI 使用的请求等待队列很相似。这种技术从根本上减少了 TCQ 使用的中断数量，这归功于它继承了第一方直接内存访问（Direct Memory Access，DMA）引擎以及智能命令排序，在最优的时候发起中断。
- 在较新的 SATA 磁盘中缓存也已经增大，但是大部分不能提供 ECC 功能。

表 1.1 列出了各种存储总线及其性能特性。

表 1.1 存储接口特性

接 口	原始带宽（Mbit/s）	最大传输速度（MB/s）	每通道设备
SATA 3.0	6 000	600	1
SATA 2.0	3 000	300	
SATA 1.0	1 500	150	每线 1 个
PATA 133	1 064	133.5	2
SAS 600	6 000	600	1（通过扩展器可以超过 65 000）
SAS 300	3 000	300	
SAS 150	1 500	150	
SCSI Ultra-640	5 120	640	15（加上 HBA）
SCSI Ultra-320	2 560	320	
光纤通道（通过光纤）	10 520	2 000	126 256 ³ =16 777 216（交换光纤）
光纤通道（通过铜缆）	4 000	400	
InfiniBand	10 000	1 000	1（点对点）
Quad Rate			许多（交换光纤）

资料来源：http://en.wikipedia.org/wiki/Sata#Comparison_with_other_buses

根据各种磁盘接口的不同性能特征、容量和成本，存储可以分为多个层次（见表 1.2），满足不同的服务水平协议（Service Level Agreement, SLA）。

表 1.2 存储层次示例

层 次	存储类型	优 势	劣 势
0	固态硬盘（SSD）	最高的原始带宽和传输速度。在顺序和随机读写上均有最高的速度。无移动部件	非常昂贵，容量较小，寿命以写入操作数量计算
1	SCSI/SAS- 光纤	较高的原始带宽和传输速度	昂贵，容量较小
2	SCSI/SAS- 铜缆	平均水平的传输速度，比第 1 层便宜	有些昂贵，容量比第 3 层小
3	SATA 2 或者 3	最便宜，容量较大	传输率较低

存储层次（storage tier）的概念在 vSphere 5.0 推出的存储 DRS 功能中起到重要的作用。

1.1.4 存储容量计量单位

存储容量（二进制）以位的数量级作为计量，最小的单位是 2^0 （1 位），一个字节是 2^3 。下一个单位是 2^{10} （KB），然后是 2^{20} （MB），以 2 的 10 次方递增。实际存储容量按照国际电工协会（IEC）的规范，从 KB 开始以 1 000 倍递增。前者常用于表示 RAM 容量，而后者用于磁盘。

表 1.3 列出了存储容量的单位。

表 1.3 存储容量单位

单 位	缩 写	二 进 制 数	值	磁 盘 容 量
位（Bit）	Bit	2^0	1 位	1 位
字节（Byte）	Byte	2^3	8 位	8 位
千字节（Kilobyte）	KB	2^{10}	1 024 字节	1 000 字节
兆字节（Megabyte）	MB	2^{20}	1 024 KB	1 000 KB
千兆字节（Gigabyte）	GB	2^{30}	1 024 MB	1 000 MB
万亿字节（Terabyte）	TB	2^{40}	1 024 GB	1 000 GB
千万亿字节（Petabyte）	PB	2^{50}	1 024 TB	1 000 TB
百亿亿字节（Exabyte）	EB	2^{60}	1 024 PB	1 000 PB

注意 带宽的计量单位基于每秒位数，增量为 1 024 倍，缩写使用小写的字母“b”，代替表示字节的大写字母“B”。例如，每秒 10 兆位写作 10 Mb/s 或者 10 Mbps。常常疏忽地互换使用 B 和 b。必须注意，不要采用错误的命名惯例，否则可能使你协商的速度提高或者降低 8 倍！

1.1.5 与 vSphere 5 相关的永久性存储类型

ESXi 5 可以安装在本地磁盘、存储区域网络 (SAN) 中的 LUN 或者 iSCSI 存储中的 LUN 上 (见第 2 ~ 4 章)。

(细节: ESX/ESXi 4.1 之前的版本不支持从 iSCSI 启动。)

1. 支持的本地存储媒体

vSphere 5 支持如下类型的本地存储媒体:

- SCSI 磁盘 (并行)
- 串行 SCSI 磁盘 (SAS)
- 串行 ATA 磁盘 (SATA)
- SD 闪存盘和 USB key (这适用于 ESXi 3.5 嵌入 / 可安装配置)
- 固态硬盘 (SSD)

2. 共享存储设备

vSphere 5.0 的某些功能需要共享存储才能生效——例如, 高可用性 (HA)、分布式资源调度器 (Distributed Resource Scheduler, DRS)、vMotion、存储 vMotion 及存储 DRS 等。

这些共享存储, 不管是网络连接存储 (Network Attached Storage, NAS) 还是块设备, 都必须在 VMware 的硬件兼容性列表 (HCL) 中。在该列表中列出意味着这些设备已经过测试和认证, 符合最低性能标准, 具备多路径和故障切换能力, 也可能受到某些 VMware API (如 vSphere Storage API for Array Integration (VAAI) 和 vSphere Storage API for Storage Awareness (VASA)) 的支持。

符合 VMware HCL 要求的典型块存储设备由如下部分组成。

- 一个或者多个存储处理器 (Storage Processor, SP) ——也称作存储控制器。
- 每个 SP 有两个或者更多不同连接方式和速度的端口 (例如, 光纤通道、iSCSI)。进一步的细节参见第 2 ~ 4 章。
- 有些 EMC 存储阵列提供多个 SP (称作控制器 (Director)), 每个控制器中有多个端口 (例如, EMC DMX 阵列提供多个 FA 控制器, 每个有 4 个端口)。
- SP 的后端连接到 1 个或者多个 DAE, 这些 DAE 存放表 1.2 中列出的不同类型磁盘。
- 有些存储阵列通过光纤通道回路交换机将 SP 连接到 DAE。

3. 存储设备选择技巧

当你设计一个 vSphere 5 环境时, 对存储组件的选择是成功设计和实施的关键。下列原则和技巧能够帮助你做出正确的选择:

- 列出要进行虚拟化的应用程序。
- 确定这些应用程序的磁盘 I/O 条件。

- 确定带宽需求。
- 计算应用数据的磁盘容量需求。
- 确定这些应用程序的 SLA。

注意，设计 I/O 峰值往往比容量更重要。性能不足的存储架构是虚拟化环境性能问题最常见的根源。

1.2 小结

本章介绍了存储、常见存储类型以及 vSphere ESXi 5 所用的存储。接下来的几章提供进一步的细节。

第2章 光纤通道存储连接性

在外交领域，协议（protocol）的定义是，“一组指导活动执行方法的规则。”相比之下，协议在技术领域的定义与此相去不远；协议在技术上也用于指导某些活动的执行方法！

本章提供对光纤通道（Fibre Channel，FC）存储协议和连接性的概述，而后续的两章介绍以太网上的光纤通道（Fibre Channel over Ethernet，FCoE）和互联网小型计算机系统接口（iSCSI）协议。

2.1 SCSI 标准和协议

SCSI（Small Computer System Interface，小型计算机系统接口）是一组在电脑和 SCSI 外设之间物理连接和传输数据的标准。这些标准定义了命令和协议。

2.1.1 SCSI-2 和 SCSI-3 标准

SCSI-2 和 SCSI-3 标准由 T10 技术委员会管理（见 <http://www.t10.org/drafts.htm>）。

SCSI-2 指的是第二代 SCSI 标准，SCSI-3 则指第三代 SCSI 标准。但是，后续的标准去掉了数字“-3”。在修订 SCSI-3 架构模型（SCSI Architecture Model，SAM）时，它变成 SCSI 架构模型-2（SAM-2）。换句话说，不存在 SCSI-4 标准，而代之以 SAM 的修订号，后续的各代称为 SAM-2、SAM-4 等。

图 2.1 展示了 SCSI 标准架构和相关的协议。

ESXi 5 主要使用 SCSI-2 标准，在某些操作和配置中也使用 SCSI-3。我在本书中会标出 vSphere 5 功能使用的 SCSI 标准。

2.1.2 光纤通道协议

光纤通道协议（Fibre Channel Protocol，FCP）由 T11 技术委员会（最新草案列表参见 <http://www.t11.org/stat.nsf/fcproj?OpenView&Count=70>）管理。

FCP 用于不同线速的光纤通道网络（目前的速度范围是 1 ~ 8 Gb/s，但是正在酝酿更高的速率，如 16 ~ 20 Gb/s）。光纤通道连接的基本单元是帧（frame）。图 2.2 展示了 FC 帧的结构。这和以太网帧或者 IP 封包有些相似。

重要的是，理解 FC 帧结构能够帮助你解读本书后面讨论的各种 vSphere 日志中列出的存储相关信息。