

Information Storage and Management

Storing, Managing, and Protecting Digital Information

ISMM

EMC Education Services



Information Storage and Management

**Storing, Managing, and Protecting
Digital Information**

Edited by
G. Somasundaram
Alok Shrivastava
EMC Education Services



WILEY

Wiley Publishing, Inc.

Information Storage and Management

Published by
Wiley Publishing, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Copyright © 2009 by EMC Corporation

Published by Wiley Publishing, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-0-470-29421-5

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Legal Department, Wiley Publishing, Inc., 10475 Crosspoint Blvd., Indianapolis, IN 46256, (317) 572-3447, fax (317) 572-4355, or online at www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Web site may provide or recommendations it may make. Further, readers should be aware that Internet Web sites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Library of Congress Cataloging-in-Publication Data is available from the publisher.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, Inc. is not associated with any product or vendor mentioned in this book.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

EMC², EMC, EMC Centera, EMC ControlCenter, AdvantEdge, AlphaStor, ApplicationXtender, Avamar, Captiva, Catalog Solution, Celerra, Centera, CentraStar, ClaimPack, ClaimsEditor, ClaimsEditor Professional, CLARAlert, CLARiON, ClientPak, CodeLink, Connectrix, Co-StandbyServer, Dantz, Direct Matrix Architecture, DiskXtender, DiskXtender 2000, Document Sciences, Documentum, EmailXaminer, EmailXtender, EmailXtract, eRoom, Event Explorer, FLARE, FormWare, HighRoad, InputAccel, Invista, ISIS, Max Retriever, Navisphere, NetWorker, nLayers, OpenScale, PixTools, Powerlink, PowerPath, Rainfinity, RepliStor, ResourcePak, Retrospect, Smarts, SnapShotServer, SnapView/IP, SRDF, Symmetrix, TimeFinder, VisualSAN, Voyence, VSAM-Assist, WebXtender, where information lives, xPpression, xPresso, Xtender, and Xtender Solutions are registered trademarks and EMC LifeLine, EMC OnCourse, EMC Proven, EMC Snap, EMC Storage Administrator, Acartus, Access Logix, ArchiveXtender, Atmos, Authentic Problems, Automated Resource Manager, AutoStart, AutoSwap, AVALONidm, C-Clip, Celerra Replicator, CenterStage, CLARevent, Codebook Correlation Technology, Common Information Model, CopyCross, CopyPoint, DatabaseXtender, Digital Mailroom, Direct Matrix, EDM, E-Lab, eInput, Engenuity, FarPoint, FirstPass, Fortress, Global File Virtualization, Graphic Visualization, InfiniFlex, InfoMover, Infoscape, InputAccel Express, MediaStor, MirrorView, Mozy, MozyEnterprise, MozyHome, MozyPro, OnAlert, PowerSnap, QuickScan, RepliCare, SafeLine, SAN Advisor, SAN Copy, SAN Manager, SDMS, SnapImage, SnapSure, SnapView, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix DMX, UltraFlex, UltraPoint, UltraScale, Viewlets, Virtual Provisioning, and VisualSRM are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners. © Copyright 2009 EMC Corporation. All rights reserved. Published in the USA. 01/09



About the Editors

G Somasundaram (Somu) is a graduate from the Indian Institute of Technology in Mumbai, India, and has over 22 years of experience in the IT industry, the last 10 with EMC Corporation. Currently he is director, EMC Global Services, leading worldwide industry readiness initiatives. Somu is the architect of EMC's open storage curriculum, aimed at addressing the storage knowledge "gap" that exists in the IT industry. Under his leadership and direction, industry readiness initiatives, such as the EMC Learning Partner and Academic Alliance programs, continue to experience significant growth and educate thousands of students worldwide on information storage and management technologies. Key areas of Somu's responsibility include guiding a global team of professionals, identifying and partnering with global IT education providers, and setting the overall direction for EMC's industry readiness initiatives. Prior to his current role, Somu held various managerial and leadership roles with EMC as well as other leading IT vendors.

Alok Shrivastava is senior director, EMC Global Services and has focused on education since 2003. Alok is the architect of several of EMC's successful education initiatives including the industry leading EMC Proven Professional program, industry readiness programs such as EMC's Academic Alliance, and most recently this unique and valuable book on information storage technology. Alok provides vision and leadership to a team of highly talented experts and professionals that develops world-class technical education for EMC's employees, partners, customers, and other industry professionals. Prior to his success in education, Alok built and led a highly successful team of EMC presales engineers in Asia-Pacific and Japan. Earlier in his career, Alok was a systems manager, storage manager, and a backup/restore/disaster recovery consultant working with some of the world's largest data centers and IT installations. He holds dual Masters degrees from the Indian Institute of Technology in Mumbai, India, and the University of Sagar in India. Alok has worked in information storage technology and has held a unique passion for this field for most of his 25-plus year career in IT.

Credits

Executive Editor

Carol Long

Senior Development Editor

Tom Dinse

Production Editor

Dassi Zeidel

Copy Editor

Luann Rouff

Editorial Manager

Mary Beth Wakefield

Production Manager

Tim Tate

Vice President and Executive**Group Publisher**

Richard Swadley

**Vice President and Executive
Publisher**

Barry Pruett

Project Coordinator, Cover

Lynsey Stanford

Composer

Jeffrey Lytle, Happenstance

Type-O-Rama

Proofreader

Nancy Bell

Indexer

Robert Swanson

Cover Image

© 2008 Ron Chapple/Ron Chapple

Stock/Photos To Go

Cover Designer

EMC Creative Development



Acknowledgments

When we embarked upon the project to develop this book, the very first challenge was to identify a team of subject matter experts covering the vast range of technologies that form the modern information storage infrastructure.

A key factor working in our favor is that at EMC, we have the technologies, the know-how, and many of the best talents in the industry. When we reached out to individual experts, they were as excited as we were about the prospect of publishing a comprehensive book on information storage technology. This was an opportunity to share their expertise with professionals and students worldwide.

This book is the result of efforts and contributions from a number of key EMC organizations led by EMC Education Services and supported by the office of CTO, Global Marketing, and EMC Engineering.

In addition to his own research and expertise, **Ganesh Rajaratnam**, from EMC Education Services, led the efforts with other subject matter experts to develop the first draft of the book. **Dr. David Black**, from the EMC CTO office, devoted many valuable hours to combing through the content and providing cogent advice on the key topics covered in this book.

We are very grateful to the following experts from EMC Education Services for developing the content for various sections and chapters of this book:

Rodrigo Alves
Charlie Brooks
Debasish Chakrabarty
Diana Davis
Amit Deshmukh
Michael Dulavitz
Ashish Garg
Dr. Vanchi Gurumoorthy
Simon Hawkshaw

Anbuselvi Jeyakumar
Sagar Kotekar Patil
Andre Rossouw
Tony Santamaria
Saravanaraj Sridharan
Ganesh Sundaresan
Jim Tracy
Anand Varkar
Dr. Viswanth VS

The following experts thoroughly reviewed the book at various stages and provided valuable feedback and guidance:

Ronen Artzi	Jack Harwood
Eric Baize	Arthur Johnson
Greg Baltazar	Michelle Lavoie
Edward Bell	Tom McGowan
Christopher Chaulk	Jeffery Moore
Roger Dupuis	Toby Morral
Deborah Filer	Peter Popieniuck
Bala Ganeshan	Kevin Sheridan
Jason Gervickas	Ed VanSickle
Nancy Gessler	
Jody Goncalves	

We also thank NIIT Limited for their help with the initial draft, **Muthaiah Thiagarajan** of EMC and DreaMarT Interactive Pvt. Ltd. for their support in creating all illustrations, and the publisher, John Wiley & Sons, for their timely support in bringing this book to the industry.

— **G. Somasundaram**

Director, Education Services, EMC Corporation

— **Alok Shrivastava**

Senior Director, Education Services, EMC Corporation

March 2009



Contents

Foreword		xvii
Introduction		xix
Section I	Storage System	1
Chapter 1	Introduction to Information Storage and Management	3
	1.1 Information Storage	5
	1.1.1 Data	5
	1.1.2 Types of Data	7
	1.1.3 Information	7
	1.1.4 Storage	8
	1.2 Evolution of Storage Technology and Architecture	9
	1.3 Data Center Infrastructure	10
	1.3.1 Core Elements	10
	1.3.2 Key Requirements for Data Center Elements	11
	1.3.3 Managing Storage Infrastructure	13
	1.4 Key Challenges in Managing Information	14
	1.5 Information Lifecycle	14
	1.5.1 Information Lifecycle Management	15
	1.5.2 ILM Implementation	16
	1.5.3 ILM Benefits	17
	Summary	18
Chapter 2	Storage System Environment	21
	2.1 Components of a Storage System Environment	21
	2.1.1 Host	22
	2.1.2 Connectivity	24
	2.1.3 Storage	26

2.2	Disk Drive Components	27
2.2.1	Platter	28
2.2.2	Spindle	28
2.2.3	Read/Write Head	28
2.2.4	Actuator Arm Assembly	29
2.2.5	Controller	29
2.2.6	Physical Disk Structure	30
2.2.7	Zoned Bit Recording	31
2.2.8	Logical Block Addressing	32
2.3	Disk Drive Performance	33
2.3.1	Disk Service Time	33
2.4	Fundamental Laws Governing Disk Performance	35
2.5	Logical Components of the Host	38
2.5.1	Operating System	39
2.5.2	Device Driver	39
2.5.3	Volume Manager	39
2.5.4	File System	41
2.5.5	Application	44
2.6	Application Requirements and Disk Performance	45
	Summary	48
Chapter 3	Data Protection: RAID	51
3.1	Implementation of RAID	52
3.1.1	Software RAID	52
3.1.2	Hardware RAID	52
3.2	RAID Array Components	53
3.3	RAID Levels	54
3.3.1	Striping	54
3.3.2	Mirroring	55
3.3.3	Parity	56
3.3.4	RAID 0	57
3.3.5	RAID 1	57
3.3.6	Nested RAID	59
3.3.7	RAID 3	59
3.3.8	RAID 4	61
3.3.9	RAID 5	62
3.3.10	RAID 6	62
3.4	RAID Comparison	63
3.5	RAID Impact on Disk Performance	66
3.5.1	Application IOPS and RAID Configurations	67
3.6	Hot Spares	68
	Summary	68
Chapter 4	Intelligent Storage System	71
4.1	Components of an Intelligent Storage System	72
4.1.1	Front End	72
4.1.2	Cache	74
4.1.3	Back End	79
4.1.4	Physical Disk	79

4.2 Intelligent Storage Array	82
4.2.1 High-end Storage Systems	82
4.2.2 Midrange Storage System	83
4.3 Concepts in Practice: EMC CLARiiON and Symmetrix	84
4.3.1 CLARiiON Storage Array	85
4.3.2 CLARiiON CX4 Architecture	86
4.3.3 Managing the CLARiiON	88
4.3.4 Symmetrix Storage Array	89
4.3.5 Symmetrix Component Overview	91
4.3.6 Direct Matrix Architecture	93
Summary	95
Section II Storage Networking Technologies and Virtualization	97
Chapter 5 Direct-Attached Storage and Introduction to SCSI	99
5.1 Types of DAS	99
5.1.1 Internal DAS	100
5.1.2 External DAS	100
5.2 DAS Benefits and Limitations	101
5.3 Disk Drive Interfaces	101
5.3.1 IDE/ATA	101
5.3.2 SATA	102
5.3.3 Parallel SCSI	103
5.4 Introduction to Parallel SCSI	104
5.4.1 Evolution of SCSI	104
5.4.2 SCSI Interfaces	105
5.4.3 SCSI-3 Architecture	107
5.4.4 Parallel SCSI Addressing	111
5.5 SCSI Command Model	112
5.5.1 CDB Structure	112
5.5.2 Operation Code	112
5.5.3 Control Field	114
5.5.4 Status	114
Summary	115
Chapter 6 Storage Area Networks	117
6.1 Fibre Channel: Overview	118
6.2 The SAN and Its Evolution	119
6.3 Components of SAN	120
6.3.1 Node Ports	120
6.3.2 Cabling	122
6.3.3 Interconnect Devices	123
6.3.4 Storage Arrays	124
6.3.5 SAN Management Software	124
6.4 FC Connectivity	125
6.4.1 Point-to-Point	125
6.4.2 Fibre Channel Arbitrated Loop	125
6.4.3 Fibre Channel Switched Fabric	127

6.5 Fibre Channel Ports	129
6.6 Fibre Channel Architecture	131
6.6.1 Fibre Channel Protocol Stack	132
6.6.2 Fibre Channel Addressing	133
6.6.3 FC Frame	135
6.6.4. Structure and Organization of FC Data	137
6.6.5 Flow Control	137
6.6.6 Classes of Service	138
6.7 Zoning	138
6.8 Fibre Channel Login Types	141
6.9 FC Topologies	141
6.9.1 Core-Edge Fabric	142
6.9.2 Mesh Topology	144
6.10 Concepts in Practice: EMC Connectrix	145
Summary	148
Chapter 7 Network-Attached Storage	149
7.1 General-Purpose Servers vs. NAS Devices	150
7.2 Benefits of NAS	150
7.3 NAS File I/O	151
7.3.1 File Systems and Remote File Sharing	152
7.3.2 Accessing a File System	152
7.3.3 File Sharing	152
7.4 Components of NAS	153
7.5 NAS Implementations	154
7.5.1 Integrated NAS	154
7.5.2 Gateway NAS	155
7.5.3 Integrated NAS Connectivity	155
7.5.4 Gateway NAS Connectivity	156
7.6 NAS File-Sharing Protocols	157
7.6.1 NFS	158
7.6.2 CIFS	158
7.7 NAS I/O Operations	159
7.7.1 Hosting and Accessing Files on NAS	160
7.8 Factors Affecting NAS Performance and Availability	160
7.9 Concepts in Practice: EMC Celerra	164
7.9.1 Architecture	164
7.9.2 Celerra Product Family	167
Summary	168
Chapter 8 IP SAN	171
8.1 iSCSI	173
8.1.1 Components of iSCSI	173
8.1.2 iSCSI Host Connectivity	174
8.1.3 Topologies for iSCSI Connectivity	175
8.1.4 iSCSI Protocol Stack	176
8.1.5 iSCSI Discovery	177

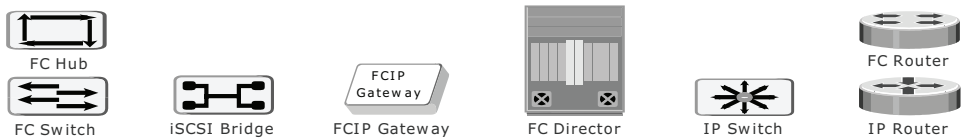
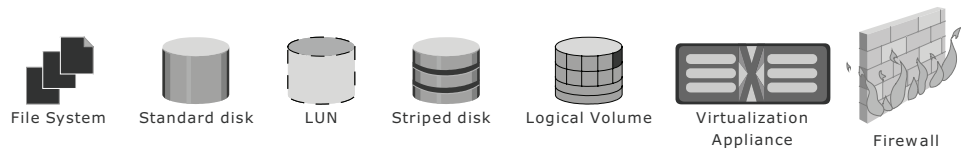
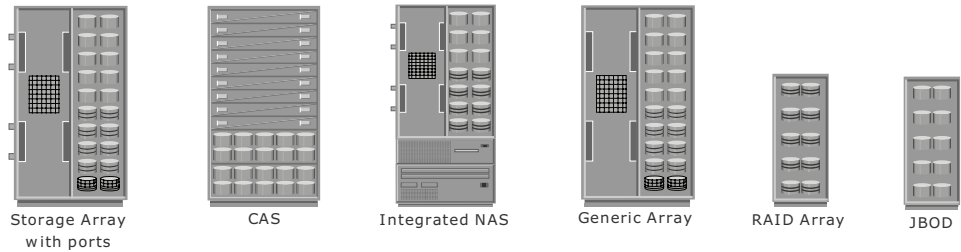
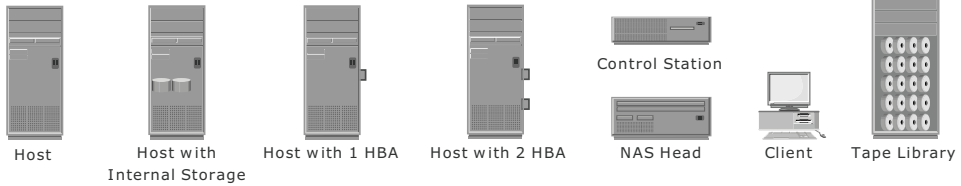
8.1.6 iSCSI Names	178
8.1.7 iSCSI Session	179
8.1.8 iSCSI PDU	180
8.1.9 Ordering and Numbering	181
8.1.10 iSCSI Error Handling and Security	182
8.2 FCIP	183
8.2.1 FCIP Topology	184
8.2.2 FCIP Performance and Security	185
Summary	186
Chapter 9 Content-Addressed Storage	189
9.1 Fixed Content and Archives	190
9.2 Types of Archives	191
9.3 Features and Benefits of CAS	192
9.4 CAS Architecture	193
9.5 Object Storage and Retrieval in CAS	196
9.6 CAS Examples	198
9.6.1 Health Care Solution: Storing Patient Studies	198
9.6.2 Finance Solution: Storing Financial Records	199
9.7 Concepts in Practice: EMC Centera	200
9.7.1 EMC Centera Models	201
9.7.2 EMC Centera Architecture	201
9.7.3 Centera Tools	203
9.7.4 EMC Centera Universal Access	204
Summary	205
Chapter 10 Storage Virtualization	207
10.1 Forms of Virtualization	207
10.1.1 Memory Virtualization	208
10.1.2 Network Virtualization	208
10.1.3 Server Virtualization	209
10.1.4 Storage Virtualization	210
10.2 SNIA Storage Virtualization Taxonomy	212
10.3 Storage Virtualization Configurations	213
10.4 Storage Virtualization Challenges	214
10.4.1 Scalability	215
10.4.2 Functionality	215
10.4.3 Manageability	215
10.4.4 Support	216
10.5 Types of Storage Virtualization	216
10.5.1 Block-Level Storage Virtualization	216
10.5.2 File-Level Virtualization	217
10.6 Concepts in Practice	219
10.6.1 EMC Invista	219
10.6.2 Rainfinity	222
Summary	225

Section III	Business Continuity	227
Chapter 11	Introduction to Business Continuity	229
	11.1 Information Availability	230
	11.1.1 Causes of Information Unavailability	230
	11.1.2 Measuring Information Availability	231
	11.1.3 Consequences of Downtime	232
	11.2 BC Terminology	233
	11.3 BC Planning Lifecycle	235
	11.4 Failure Analysis	238
	11.4.1 Single Point of Failure	238
	11.4.2 Fault Tolerance	238
	11.4.3 Multipathing Software	240
	11.5 Business Impact Analysis	240
	11.6 BC Technology Solutions	241
	11.7 Concept in Practice: EMC PowerPath	241
	11.7.1 PowerPath Features	242
	11.7.2 Dynamic Load Balancing	242
	11.7.3 Automatic Path Failover	244
	Summary	247
Chapter 12	Backup and Recovery	251
	12.1 Backup Purpose	252
	12.1.1 Disaster Recovery	252
	12.1.2 Operational Backup	252
	12.1.3 Archival	252
	12.2 Backup Considerations	253
	12.3 Backup Granularity	254
	12.4 Recovery Considerations	257
	12.5 Backup Methods	258
	12.6 Backup Process	259
	12.7 Backup and Restore Operations	260
	12.8 Backup Topologies	262
	12.8.1 Serverless Backup	265
	12.9 Backup in NAS Environments	265
	12.10 Backup Technologies	269
	12.10.1 Backup to Tape	269
	12.10.2 Physical Tape Library	270
	12.10.3 Backup to Disk	272
	12.10.4 Virtual Tape Library	273
	12.11 Concepts in Practice: EMC NetWorker	276
	12.11.1 NetWorker Backup Operation	277
	12.11.2 NetWorker Recovery	278
	12.11.3 EmailXtender	278
	12.11.4 DiskXtender	279
	12.11.5 Avamar	279
	12.11.6 EMC Disk Library (EDL)	280
	Summary	280

Chapter 13	Local Replication	283
	13.1 Source and Target	284
	13.2 Uses of Local Replicas	284
	13.3 Data Consistency	285
	13.3.1 Consistency of a Replicated File System	285
	13.3.2 Consistency of a Replicated Database	286
	13.4 Local Replication Technologies	288
	13.4.1 Host-Based Local Replication	288
	13.4.2 Storage Array–Based Replication	290
	13.5 Restore and Restart Considerations	297
	13.5.1 Tracking Changes to Source and Target	298
	13.6 Creating Multiple Replicas	300
	13.7 Management Interface	301
	13.8 Concepts in Practice: EMC TimeFinder and EMC SnapView	301
	13.8.1 TimeFinder/Clone	302
	13.8.2 TimeFinder/Mirror	302
	13.8.3 EMC SnapView	304
	13.8.4 EMC SnapSure	305
	Summary	306
 Chapter 14	 Remote Replication	 309
	14.1 Modes of Remote Replication	309
	14.2 Remote Replication Technologies	311
	14.2.1. Host-Based Remote Replication	311
	14.2.2 Storage Array-Based Remote Replication	314
	14.2.3 SAN-Based Remote Replication	321
	14.3 Network Infrastructure	323
	14.3.1 DWDM	324
	14.3.2 SONET	324
	14.4 Concepts in Practice: EMC SRDF, EMC SAN Copy, and EMC MirrorView	325
	14.4.1 SRDF Family	325
	14.4.2 Disaster Recovery with SRDF	326
	14.4.3 SRDF Operations for Concurrent Access	327
	14.4.4 EMC SAN Copy	328
	14.4.5 EMC MirrorView	329
	Summary	330
 Section IV	 Storage Security and Management	 333
 Chapter 15	 Securing the Storage Infrastructure	 335
	15.1 Storage Security Framework	335
	15.2 Risk Triad	336
	15.2.1 Assets	337
	15.2.2 Threats	338
	15.2.3 Vulnerability	339

15.3 Storage Security Domains	340
15.3.1 Securing the Application Access Domain	341
15.3.2 Securing the Management Access Domain	344
15.3.3 Securing Backup, Recovery, and Archive (BURA)	347
15.4 Security Implementations in Storage Networking	348
15.4.1 SAN	348
15.4.2 NAS	353
15.4.3 IP SAN	359
Summary	360
Chapter 16 Managing the Storage Infrastructure	363
16.1 Monitoring the Storage Infrastructure	364
16.1.1 Parameters Monitored	364
16.1.2 Components Monitored	365
16.1.3 Monitoring Examples	368
16.1.4 Alerts	374
16.2 Storage Management Activities	375
16.2.1 Availability management	375
16.2.2 Capacity management	375
16.2.3 Performance management	376
16.2.4 Security Management	376
16.2.5 Reporting	376
16.2.6 Storage Management Examples	377
16.3 Storage Infrastructure Management Challenges	382
16.4 Developing an Ideal Solution	382
16.4.1 Storage Management Initiative	383
16.4.2 Enterprise Management Platforms	385
16.5 Concepts in Practice: EMC ControlCenter	386
16.5.1 ControlCenter Features and Functionality	386
16.5.2 ControlCenter Architecture	387
Summary	392
Appendix	395
Glossary	407
Index	439

Icons used in this book





Foreword

Ralph Waldo Emerson, the great American essayist, philosopher, and poet, once said that the invariable mark of wisdom is seeing the miraculous in the common. Today, common miracles surround us, and it is virtually impossible *not* to see them. Most of us have modern gadgetry such as digital cameras, video camcorders, cell phones, fast computers that can access millions of websites, instant messaging, social networking sites, search engines, music downloads ... the list goes on. All of these examples have one thing in common: they generate huge volumes of data. Not only are we in an information age, we're in an age where information is exploding into a digital universe that requires enhanced technology and a new generation of professionals who are able to manage, leverage, and optimize *storage and information management* solutions.

Just to give you an idea of the challenges we face today, in one year the amount of digital information created, captured, and replicated is millions of times the amount of information in all the books ever written. Information is the most important asset of a business. To realize the inherent power of information, it must be intelligently and efficiently stored, protected, and managed—so that it can be made accessible, searchable, shareable, and, ultimately, actionable.

We are currently in the perfect storm. Everything is increasing: the information, the costs, and the skilled professionals needed to store and manage it—professionals who are not available in sufficient numbers to meet the growing need. The IT manager's number one concern is how to manage this storage growth. Enterprises simply cannot purchase bigger and better "boxes" to store their data. IT managers must not only worry about budgets for storage technology, but also be concerned with energy-efficient, footprint-reducing technology that is easy to install, manage, and use. Although many IT managers intend to

hire more trained staff, they are facing a shortage of skilled, storage-educated professionals who can take control of managing and optimizing the data.

I was unable to find a comprehensive book in the marketplace that provided insight into the various technologies deployed to store and manage information. As an industry leader, we have the subject-matter expertise and practical experience to help fill this gap; and now this book can give you a behind-the-scenes view of the technologies used in information storage and management. You will learn where data goes, how it is managed, and how you can contribute to your company's profitability.

If you've chosen storage and information infrastructure management as your career, you are a pioneer in a profession that is undergoing constant change, but one in which the challenges lead to great rewards.

Regardless of your current role in IT, this book should be a key part of your IT library and professional development.

Thomas P. Clancy

Vice President, Education Services, EMC Corporation

March 2009



Introduction

Information storage is a central pillar of information technology. A large quantity of digital information is being created every moment by individual and corporate consumers of IT. This information needs to be stored, protected, optimized, and managed.

Not long ago, information storage was seen as only a bunch of disks or tapes attached to the back of the computer to store data. Even today, only those in the storage industry understand the critical role that information storage technology plays in the availability, performance, integration, and optimization of the entire IT infrastructure. Over the last two decades, information storage has developed into a highly sophisticated technology, providing a variety of solutions for storing, managing, connecting, protecting, securing, sharing, and optimizing digital information.

With the exponential growth of information and the development of sophisticated products and solutions, there is also a growing need for information storage professionals. IT managers are challenged by the ongoing task of employing and developing highly skilled information storage professionals.

Many leading universities and colleges have started to include storage technology courses in their regular computer technology or information technology curriculum, yet many of today's IT professionals, even those with years of experience, have not benefited from this formal education, therefore many seasoned professionals—including application, systems, database, and network administrators—do not share a common foundation about how storage technology affects their areas of expertise.

This book is designed and developed to enable professionals and students to achieve a comprehensive understanding of all segments of storage technology. While the product examples used in the book are from EMC Corporation, an

understanding of the technology concepts and principles prepare the reader to easily understand products from various technology vendors.

This book has 16 chapters, organized in four sections. Advanced topics build upon the topics learned in previous chapters.

Part 1, “Information Storage and Management for Today’s World”: These four chapters cover information growth and challenges, define a storage system and its environment, review the evolution of storage technology, and introduce intelligent storage systems.

Part 2, “Storage Options and Protocols”: These six chapters cover the SCSI and Fibre channel architecture, direct-attached storage (DAS), storage area networks (SANs), network-attached storage (NAS), Internet Protocol SAN (IP-SAN), content-addressed storage (CAS), and storage virtualization.

Part 3, “Business Continuity and Replication”: These four chapters introduce business continuity, backup and recovery, local data replication, and remote data replication.

Part 4, “Security and Administration”: These two chapters cover storage security and storage infrastructure monitoring and management.

This book has a supplementary website that provides additional up-to-date learning aids and reading material. Visit <http://education.EMC.com/ismbook> for details.

EMC Academic Alliance

Universities and colleges interested in offering an *information storage and management* curriculum are invited to join the Academic Alliance program. This program provides comprehensive support to institutes, including teaching aids, faculty guides, student projects, and more. Please visit <http://education.EMC.com/academicalliance>.

EMC Proven Professional Certification



This book prepares students and professionals to take the EMC Proven Professional Information Storage and Management exam E20-001. EMC Proven Professional is the premier certification program that validates your knowledge and helps establish your credibility in the information technology industry. For more information on certification as well as to access practice exams, visit <http://education.EMC.com>.

Storage System

In This Section

Chapter 1: Introduction to Information Storage and Management

Chapter 2: Storage System Environment

Chapter 3: Data Protection: RAID

Chapter 4: Intelligent Storage Systems

Chapter 1

Introduction to Information Storage and Management

Information is increasingly important in our daily lives. We have become information dependents of the twenty-first century, living in an on-command, on-demand world that means we need information when and where it is required. We access the Internet every day to perform searches, participate in social networking, send and receive e-mails, share pictures and videos, and scores of other applications. Equipped with a growing number of content-generating devices, more information is being created by individuals than by businesses. Information created by individuals gains value when shared with others. When created, information resides locally on devices such as cell phones, cameras, and laptops. To share this information, it needs to be uploaded via networks to data centers. It is interesting to note that while the majority of information is created by individuals, it is stored and managed by a relatively small number of organizations. Figure 1-1 depicts this virtuous cycle of information.

The importance, dependency, and volume of information for the business world also continue to grow at astounding rates. Businesses depend on fast and reliable access to information critical to their success. Some of the business applications that process information include airline reservations, telephone billing systems, e-commerce, ATMs, product designs, inventory management, e-mail archives, Web portals, patient records, credit cards, life sciences, and global capital markets.

The increasing criticality of information to the businesses has amplified the challenges in protecting and managing the data. The volume of data that

KEY CONCEPTS

Data and Information

Structured and Unstructured Data

Storage Technology Architectures

Core Elements of a Data Center

Information Management

Information Lifecycle Management

business must manage has driven strategies to classify data according to its value and create rules for the treatment of this data over its lifecycle. These strategies not only provide financial and regulatory benefits at the business level, but also manageability benefits at operational levels to the organization.

Data centers now view information storage as one of their core elements, along with applications, databases, operating systems, and networks. Storage technology continues to evolve with technical advancements offering increasingly higher levels of availability, security, scalability, performance, integrity, capacity, and manageability.

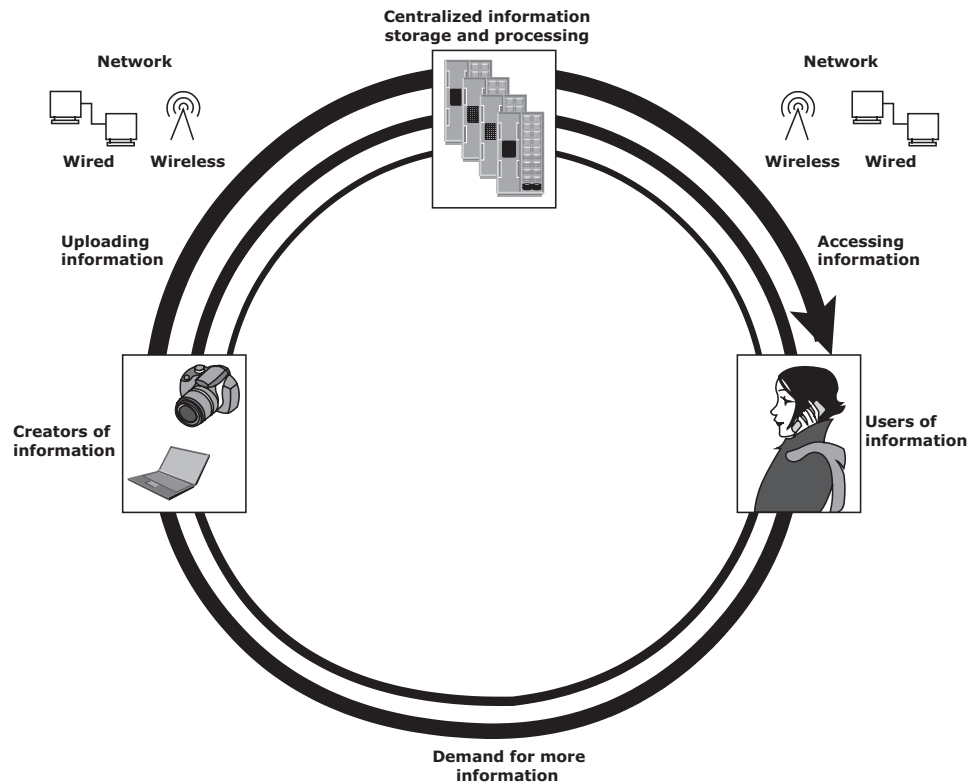


Figure 1-1: Virtuous cycle of information

This chapter describes the evolution of information storage architecture from simple direct-attached models to complex networked topologies. It introduces the information lifecycle management (ILM) strategy, which aligns the information technology (IT) infrastructure with business priorities.

1.1 Information Storage

Businesses use data to derive information that is critical to their day-to-day operations. Storage is a repository that enables users to store and retrieve this digital data.

1.1.1 Data

Data is a collection of raw facts from which conclusions may be drawn. Handwritten letters, a printed book, a family photograph, a movie on video tape, printed and duly signed copies of mortgage papers, a bank's ledgers, and an account holder's passbooks are all examples of data.

Before the advent of computers, the procedures and methods adopted for data creation and sharing were limited to fewer forms, such as paper and film. Today, the same data can be converted into more convenient forms such as an e-mail message, an e-book, a bitmapped image, or a digital movie. This data can be generated using a computer and stored in strings of 0s and 1s, as shown in Figure 1-2. Data in this form is called *digital data* and is accessible by the user only after it is processed by a computer.

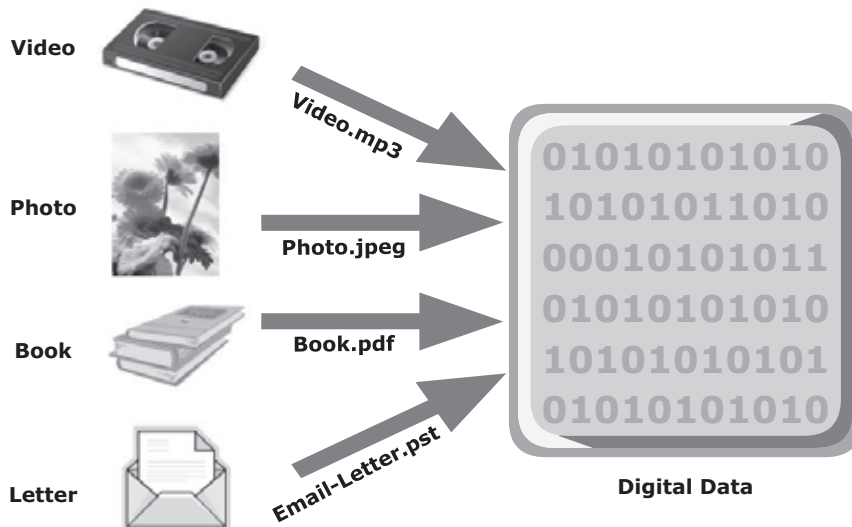


Figure 1-2: Digital data

With the advancement of computer and communication technologies, the rate of data generation and sharing has increased exponentially. The following is a list of some of the factors that have contributed to the growth of digital data:

- **Increase in data processing capabilities:** Modern-day computers provide a significant increase in processing and storage capabilities. This enables the conversion of various types of content and media from conventional forms to digital formats.
- **Lower cost of digital storage:** Technological advances and decrease in the cost of storage devices have provided low-cost solutions and encouraged the development of less expensive data storage devices. This cost benefit has increased the rate at which data is being generated and stored.
- **Affordable and faster communication technology:** The rate of sharing digital data is now much faster than traditional approaches. A handwritten letter may take a week to reach its destination, whereas it only takes a few seconds for an e-mail message to reach its recipient.

Inexpensive and easier ways to create, collect, and store all types of data, coupled with increasing individual and business needs, have led to accelerated data growth, popularly termed the *data explosion*. Data has different purposes and criticality, so both individuals and businesses have contributed in varied proportions to this data explosion.

The importance and the criticality of data vary with time. Most of the data created holds significance in the short-term but becomes less valuable over time. This governs the type of data storage solutions used. Individuals store data on a variety of storage devices, such as hard disks, CDs, DVDs, or Universal Serial Bus (USB) flash drives.

EXAMPLE OF RESEARCH AND BUSINESS DATA



- **Seismology:** Involves collecting data related to various sources and parameters of earthquakes, and other relevant data that needs to be processed to derive meaningful information.
- **Product data:** Includes data related to various aspects of a product, such as inventory, description, pricing, availability, and sales.
- **Customer data:** A combination of data related to a company's customers, such as order details, shipping addresses, and purchase history.
- **Medical data:** Data related to the health care industry, such as patient history, radiological images, details of medication and other treatment, and insurance information.

Businesses generate vast amounts of data and then extract meaningful information from this data to derive economic benefits. Therefore, businesses need to maintain data and ensure its availability over a longer period.

Furthermore, the data can vary in criticality and may require special handling. For example, legal and regulatory requirements mandate that banks maintain account information for their customers accurately and securely. Some businesses handle data for millions of customers, and ensures the security and integrity of data over a long period of time. This requires high-capacity storage devices with enhanced security features that can retain data for a long period.

1.1.2 Types of Data

Data can be classified as structured or unstructured (see Figure 1-3) based on how it is stored and managed. Structured data is organized in rows and columns in a rigidly defined format so that applications can retrieve and process it efficiently. Structured data is typically stored using a database management system (DBMS).

Data is unstructured if its elements cannot be stored in rows and columns, and is therefore difficult to query and retrieve by business applications. For example, customer contacts may be stored in various forms such as sticky notes, e-mail messages, business cards, or even digital format files such as .doc, .txt, and .pdf. Due its unstructured nature, it is difficult to retrieve using a customer relationship management application. Unstructured data may not have the required components to identify itself uniquely for any type of processing or interpretation. Businesses are primarily concerned with managing unstructured data because over 80 percent of enterprise data is unstructured and requires significant storage space and effort to manage.

1.1.3 Information

Data, whether structured or unstructured, does not fulfill any purpose for individuals or businesses unless it is presented in a meaningful form. Businesses need to analyze data for it to be of value. *Information* is the intelligence and knowledge derived from data.

Businesses analyze raw data in order to identify meaningful trends. On the basis of these trends, a company can plan or modify its strategy. For example, a retailer identifies customers' preferred products and brand names by analyzing their purchase patterns and maintaining an inventory of those products.

Effective data analysis not only extends its benefits to existing businesses, but also creates the potential for new business opportunities by using the information in creative ways. Job portal is an example. In order to reach a wider set of prospective employers, job seekers post their résumés on various websites offering job search facilities. These websites collect the résumés and post them on centrally accessible locations for prospective employers. In addition, companies post available positions on job search sites. Job-matching software matches keywords from

résumés to keywords in job postings. In this manner, the job search engine uses data and turns it into information for employers and job seekers.

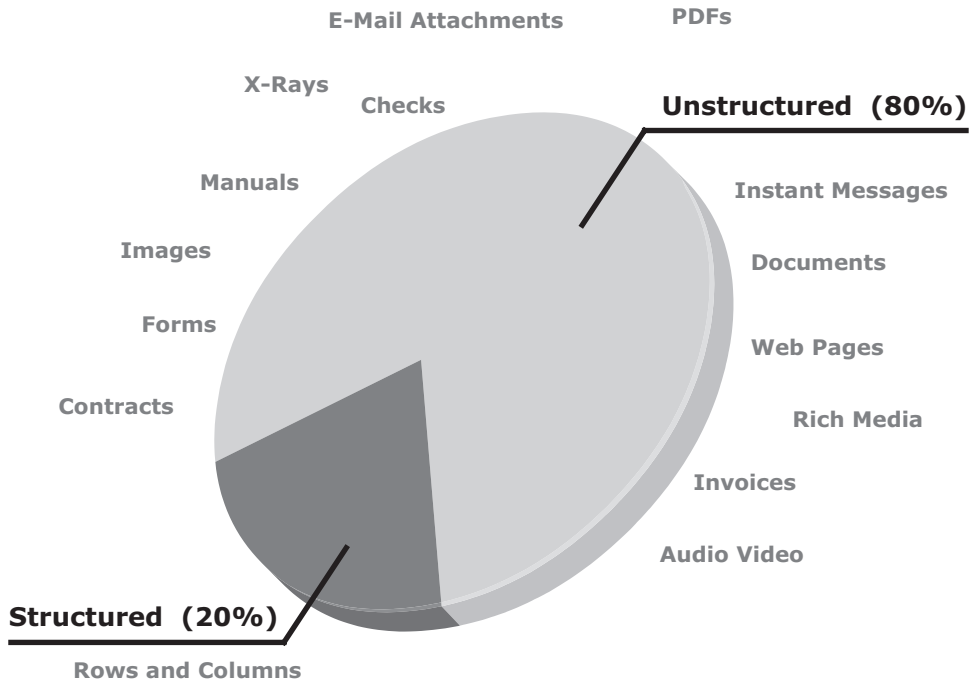


Figure 1-3: Types of data

Because information is critical to the success of a business, there is an ever-present concern about its availability and protection. Legal, regulatory, and contractual obligations regarding the availability and protection of data only add to these concerns. Outages in key industries, such as financial services, telecommunications, manufacturing, retail, and energy cost millions of U.S. dollars per hour.

1.1.4 Storage

Data created by individuals or businesses must be stored so that it is easily accessible for further processing. In a computing environment, devices designed for storing data are termed *storage devices* or simply *storage*. The type of storage used varies based on the type of data and the rate at which it is created and used. Devices such as memory in a cell phone or digital camera, DVDs, CD-ROMs, and hard disks in personal computers are examples of storage devices.

Businesses have several options available for storing data including internal hard disks, external disk arrays and tapes.