

List of Figures and Tables

FIGURES

1.1 Server-Centric IT architecture	2
1.2 Inflexible allocation of free storage capacity	3
1.3 Storage-Centric IT architecture	4
1.4 Server upgrade: preparation of a new server	5
1.5 Server upgrade: testing a new server	6
1.6 Server upgrade: putting a new server into service	7
1.7 <i>Structure of this book</i>	9
2.1 Connection of servers to a disk subsystem	16
2.2 Architecture of intelligent disk subsystems	17
2.3 Flexible allocation of free storage capacity	18
2.4 Comparison: large and small internal hard disks	19
2.5 Internal I/O channels: active and active/passive	20
2.6 Internal I/O channels: active/active	21
2.7 Virtualisation by RAID	23
2.8 Replacing a defective physical hard disk	24
2.9 RAID 0 (striping)	25
2.10 RAID 1 (mirroring)	27
2.11 RAID 0+1 (mirrored stripes)	28
2.12 RAID 10 (striped mirrors)	29
2.13 RAID 0+1: failure of a physical hard disk	30
2.14 RAID 10: failure of a physical hard disk	31
2.15 RAID 4 (parity disk)	32
2.16 RAID 4 and RAID 5: write penalty	33

2.17 RAID 5 (striped parity)	35
2.18 Instant copy	42
2.19 Space-efficient instant copy	44
2.20 High availability with remote mirroring	46
2.21 Data flow of synchronous remote mirroring	47
2.22 Data flow of asynchronous remote mirroring	48
2.23 Synchronous and asynchronous remote mirroring combined	49
2.24 Consistency groups in heterogeneous environments	51
2.25 Write order consistency for asynchronous remote mirroring	52
2.26 Chaos without LUN masking	53
2.27 Order with LUN masking	54
3.1 The physical I/O path from the CPU to the storage system	60
3.2 The physical I/O path within a disk subsystem	61
3.3 Connection of storage over SCSI daisy chain	63
3.4 SCSI target IDs and SCSI LUNs	64
3.5 Priority of the SCSI target IDs	64
3.6 Twin-tailed SCSI cabling	65
3.7 Small SCSI SAN with multiport storage systems	66
3.8 The Fibre Channel protocol stack	68
3.9 The Fibre Channel topologies: point-to-point, arbitrated loop, fabric	69
3.10 Different plug types for fiber-optic cables	72
3.11 NRZ and Manchester encoding	74
3.12 Jitter	74
3.13 Fibre Channel exchange, sequence and frame	76
3.14 The Fibre Channel frame format	77
3.15 Link flow control and end-to-end flow control	78
3.16 Fibre Channel Class 2: error-free transmission	80
3.17 Fibre Channel Class 2: transmission error	81
3.18 Fibre Channel Class 3: error-free transmission	80
3.19 Fibre Channel Class 3: transmission error	81
3.20 Fabric login, N-Port login and process login	83
3.21 Addressing in the fabric	85
3.22 Fibre Channel application protocols based on the example of FCP	87
3.23 Latency of Fibre Channel components	90
3.24 Multiple bandwidths in the fabric topology	91
3.25 Bottleneck: inter switch link (ISL)	91
3.26 Fibre Channel arbitrated loop	94
3.27 Cascading of Fibre Channel arbitrated loops	94
3.28 Public loop and private loop	95
3.29 Communication between arbitrated loop and fabric	96
3.30 Fibre Channel-to-SCSI bridges	97
3.31 Typical entry-level configuration: dual fabric	98
3.32 Fibre Channel link extender	100

3.33 Today: Fibre Channel islands	101
3.34 Tomorrow: one large InterSAN?	102
3.35 Connecting storage via iSCSI	106
3.36 Booting via iSCSI	107
3.37 Gateway protocol: internet FCP (iFCP)	107
3.38 Tunnelling protocol: Fibre Channel over IP (FCIP)	108
3.39 Comparison of the protocol stacks of FCP, FCIP, iFCP and iSCSI	110
3.40 Comparison of the frame formats of FCP, FCIP, iFCP and iSCSI	110
3.41 A joint network for LAN, MAN, WAN and SAN	112
3.42 CPU loading: TCP/IP/Ethernet versus Fibre Channel	115
3.43 Migration path from Fibre Channel to iSCSI	117
3.44 Architecture of InfiniBand	118
3.45 Example scenario of an InfiniBand interconnection	120
3.46 Virtual Interface Architecture (VIA)	121
3.47 I/O consolidation	125
3.48 FCoE protocol stack	127
3.49 FCoE frame format	127
3.50 FCoE end device (FCoE Node, ENode)	128
3.51 FCoE switch	129
3.52 Virtual link	129
3.53 FCoE Initialization Protocol (FIP)	130
3.54 FCoE SAN	131
3.55 Integration of FCoE SAN and Fibre Channel	132
3.56 FCoE SAN with blade server	133
3.57 Enhanced Transmission Selection (ETS)	134
4.1 File system and volume manager	138
4.2 File systems and databases	139
4.3 Volume manager	141
4.4 RAID in the volume manager	142
4.5 Network file system	142
4.6 NAS server and NAS gateway	144
4.7 Eye of the needle: file server	145
4.8 DAFS cluster	147
4.9 DAFS data flow	148
4.10 A comparison between NFS, uDAFS and fDAFS	149
4.11 Shared disk file systems	151
4.12 GPFS architecture	153
4.13 GPFS cluster, GPFS node set and GPFS file system	154
4.14 GPFS token management for cache synchronisation	155
4.15 Comparison of the I/O paths: SCSI, Fibre Channel, FCoE, iSCSI, NAS	157
4.16 Storage networks with NAS servers	158

5.1 Basic idea: virtualisation in the storage network	162
5.2 Virtualisation locations in the I/O path	163
5.3 Virtualisation in the disk subsystem	164
5.4 Virtualisation in the storage network	164
5.5 Virtualisation in the host bus adapter	165
5.6 Virtualisation in the volume manager	165
5.7 Mirroring in the volume manager for high fault-tolerance	166
5.8 Striping in the volume manager for high write performance	167
5.9 Limitations of storage sharing	169
5.10 Limitations due to incompatible device drivers	170
5.11 Storage virtualisation as an abstraction layer	173
5.12 Virtualisation on block level	179
5.13 Virtualisation on file level	180
5.14 Virtualisation on different levels of the storage network	181
5.15 Symmetric storage virtualisation	185
5.16 Structure of the metadata controller in symmetric virtualisation	186
5.17 Asymmetric storage virtualisation	189
5.18 Structure of the metadata controller in asymmetric virtualisation	190
6.1 Layering of the transmission techniques	196
6.2 Storage networks in the I/O path	197
6.3 Storage networks within disk subsystems	198
6.4 Inflexible: storage assignment in server-centric systems	200
6.5 Better: storage pooling in storage-centric IT systems	200
6.6 Partitioning of a tape library	201
6.7 Tape library sharing	202
6.8 Real-time data sharing	205
6.9 Partitioning of the data set	205
6.10 Parallel databases and file systems	206
6.11 Protection against the failure of an I/O bus: redundant SCSI cable	207
6.12 Protection against the failure of an I/O bus: dual storage network	208
6.13 Problem: operating system recognises hard disk several times	208
6.14 Solution: multipathing software	209
6.15 Protection against the failure of a server: server cluster	210
6.16 Case study: cluster for file servers – failure of a server	211
6.17 Server cluster and redundant I/O buses	211
6.18 Protection against the failure of a disk subsystem: remote mirroring	212
6.19 Protection against the failure of a disk subsystem: volume manager mirroring	213
6.20 Volume manager mirroring with server cluster and redundant I/O buses	214
6.21 The logical I/O path for volume manager mirroring with redundant I/O buses	215
6.22 Case study: protection of an important database	216
6.23 Case study: protection of an important database (failure of the primary data centre)	217
6.24 Shared-null configuration	220

6.25 Shared-null configuration (failure of a server)	221
6.26 Shared-nothing cluster	222
6.27 Shared-nothing cluster (failure of a server)	223
6.28 Enhanced shared-nothing cluster	224
6.29 Enhanced shared-nothing cluster (load balancing)	224
6.30 Shared-everything cluster	225
6.31 Shared-everything cluster (failure of a server)	225
6.32 Three tier architecture	226
6.33 Three tier architecture (implementation)	227
6.34 Web architecture (five tier architecture)	228
6.35 Web architecture (implementation)	229
6.36 Web architecture based upon the ‘travel portal’ case study	231
6.37 Extension of the travel portal	232
6.38 Extension of the travel portal: cache server	234
7.1 Example environment for network backup	240
7.2 Storage hierarchy in the backup server	243
7.3 Performance bottlenecks for traditional network backup	247
7.4 Separate LAN for network backup	249
7.5 Multiple backup servers	250
7.6 Backup server and application server on the same computer	251
7.7 Server-free backup	253
7.8 LAN-free backup	255
7.9 LAN-free backup with shared disk file system	256
7.10 Application server-free backup	258
7.11 Tape library sharing	260
7.12 Backup of a NAS server by means of a network backup system	264
7.13 Architecture of the Network Data Management Protocol (NDMP)	266
7.14 NDMP: Backup to a local tape drive	268
7.15 NDMP: Backup to a remote tape drive	269
7.16 NDMP: LAN-free backup with NDMP	270
7.17 Translator services in NDMP Version 5	271
7.18 Architecture of databases	273
7.19 Backup of archive log files	275
8.1 Reference architecture for digital archive systems	284
8.2 General infrastructure for network backup	287
8.3 General infrastructure for digital archiving	287
8.4 General infrastructure for backing up archive data	288
8.5 Life cycle of classical data and information	289
8.6 Life cycle of archive data	290
8.7 Example: multilevel guarantee of data integrity	304
8.8 Data mirroring in the archive storage	308
8.9 Hierarchical storage management (HSM) for archive data	309

8.10	Interfaces in the reference architecture for archive systems	312
8.11	Protocol converters for incompatible interfaces	313
8.12	Interface between multiple DMS	319
8.13	Standardised interfaces for archive systems	320
8.14	Archiving of emails using archive storage	323
8.15	Archiving of emails using archive storage and DMS	324
8.16	Archiving of files using archive storage and file server emulation	327
8.17	Archiving of files using archive storage and protocol converter	328
8.18	Archiving of files using archive storage and DMS	330
8.19	Archiving of ERP data using archive storage	333
8.20	Archiving of ERP data using archive storage and DMS	334
8.21	Archiving of PACS or HIS data using archive storage	336
8.22	Central archiving of PACS and HIS data using DMS	337
8.23	Central archiving using one DMS	339
8.24	Central archiving using federated DMS	340
9.1	Phases and dependencies of system restart	350
9.2	Business continuity for the web architecture	352
9.3	The cost–time window of business continuity	353
9.4	Cycle of risk management	354
9.5	Strategies for business continuity	357
9.6	Parameters of availability: MTTR, MTTF and MTBF	360
9.7	Availability of sequentially coupled components	361
9.8	Availability of parallel-coupled components	362
9.9	Parameter of business continuity: RTO	362
9.10	Parameter of business continuity: RPO	363
9.11	The seven-tier model for business continuity	367
9.12	Solution segments of the seven-tier model	371
9.13	D2D2T for rapid restore	374
9.14	RPO for synchronous and asynchronous remote mirroring	375
9.15	Remote mirroring combined with instant copy for disaster recovery	376
9.16	Volume manager mirroring combined with remote mirroring for high availability and disaster recovery	378
9.17	Simplification without volume manager mirroring	379
9.18	Remote mirroring with instant copy for rolling disasters	380
9.19	Protection against rolling disasters in backup data centres	381
9.20	Combination of high availability and disaster recovery	382
10.1	Devices in the storage network	392
10.2	In-band management	395
10.3	Out-band management	399
10.4	SNMP architecture	400
10.5	MIB hierarchy	402
10.6	Web Based Enterprise Management (WBEM)	404

10.7	The three columns of WBEM	404
10.8	CIM: relationships and inheritance of attributes	406
10.9	CIM schemas	407
10.10	WBEM architecture	408
10.11	SMI-S architecture	410
10.12	DMI architecture	412
11.1	Independent versus attached media changer	423
11.2	A tape library with proprietary interfaces	424
11.3	Central removable media management	425
11.4	Drive sharing with mount request queues	430
11.5	Scratch pools spanning multiple libraries	432
11.6	Drive pools spanning multiple libraries	433
11.7	Integration of offline storage (vaulting)	434
11.8	Cartridge life cycle	435
11.9	Architecture of the IEEE standard 1244	440
12.1	Main components of the SNIA shared storage model	452
12.2	The layers of SNIA shared storage model	454
12.3	Implementation of the file/record layer	455
12.4	Functionality of the block aggregation layer	457
12.5	Combination of block and file layer	458
12.6	Access paths	459
12.7	Caching	461
12.8	Clustering	463
12.9	Resource and data sharing	465
12.10	Service subsystem	466
12.11	Example – direct attached block storage	467
12.12	Example – storage network attached block storage	468
12.13	Example – block storage aggregation in a storage device: SAN appliance	469
12.14	Example – network attached block storage with metadata server: asymmetric block service	470
12.15	Example – multi-site block storage	471
12.16	Example – file server	472
12.17	Example – NAS head	473
12.18	Example – file server metadata manager	474
12.19	Example – object-based storage device (OSD)	476
12.20	Tape modelling (logical)	477
12.21	Tape modelling (physical)	478
12.22	Tape extension of the SNIA shared storage model	479
12.23	Example – file backup (tar)	481
12.24	Example – file system volume backup (dump)	482
12.25	Example – volume backup (dd)	483
12.26	Example – file backup to virtual tape	483

12.27 Example – direct attached tape	484
12.28 Example – LAN attached tape	485
12.29 Example – shared tape drive	486
12.30 Example – partitioned tape library	487
12.31 Example – virtual tape controller	487
12.32 Example – virtual tape controller with disk cache	488
12.33 Example – data mover for tape	489
12.34 Example – file server with tape drive	490
12.35 Example – file server with external tape	491
12.36 Example – file server with data mover	492

TABLES

2.1 Failure probability of a RAID 5 array	36
2.2 Comparison of RAID levels	38
3.1 SCSI cable lengths and transmission rates	63
3.2 Fibre Channel roadmap	71
3.3 Well-defined Fibre Channel addresses	86
3.4 Comparison of iSCSI, iFCP, FCIP and FCP	110
4.1 Comparison of snapshot and instant copy	140
4.2 Comparison of Fibre Channel, FCoE, iSCSI and NAS	158
8.1 Comparison of backup and archiving	285
8.2 Comparison of the three basic WORM technologies	301
9.1 Risk factors for IT operations	347
9.2 Availability and outage times for a 24×7 operation	360
9.3 RTO and RPO for high availability and disaster recovery	366
9.4 The seven-tier model for business continuity	367
11.1 The IEEE 1244 data model	441
12.1 Data – container relationships	464
A.1 Calculation of parity block for RAID 4 and RAID 5	536