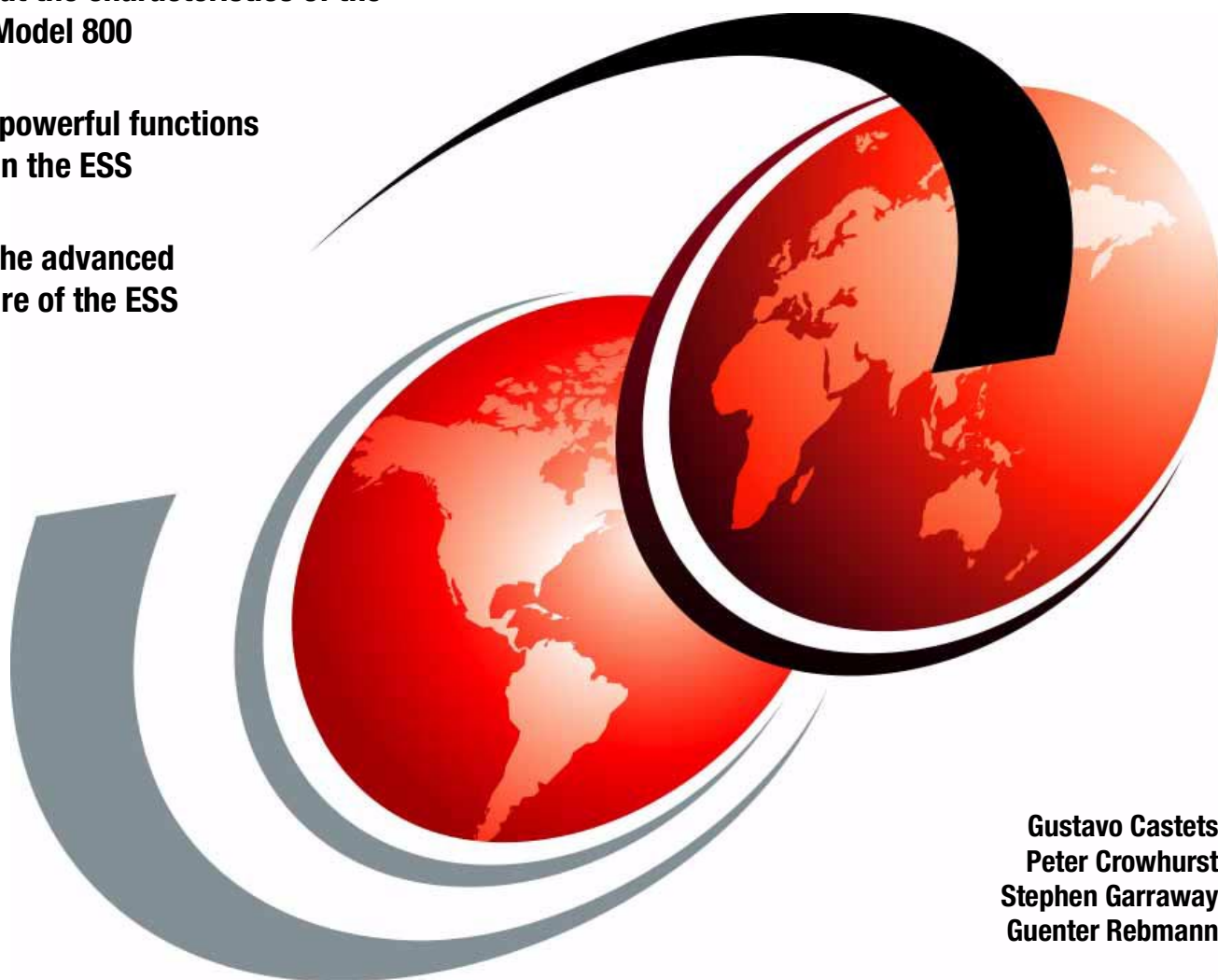


IBM TotalStorage Enterprise Storage Server Model 800

Learn about the characteristics of the
new ESS Model 800

Know the powerful functions
available in the ESS

Discover the advanced
architecture of the ESS



Gustavo Castets
Peter Crowhurst
Stephen Garraway
Guenter Rebmann

Redbooks



International Technical Support Organization

IBM TotalStorage Enterprise Storage Server Model 800

October 2002

Note: Before using this information and the product it supports, read the information in “Notices” on page xvii.

Second Edition (October 2002)

This edition applies to the IBM TotalStorage Enterprise Storage Server Model 800 (ESS Model 800) — IBM 2105-800.

© Copyright International Business Machines Corporation 2002. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Figures	xi
Tables	xv
Notices	xvii
Trademarks	xviii
Preface	xix
The team that wrote this redbook	xix
Become a published author	xxi
Comments welcome	xxi
Chapter 1. Introduction	1
1.1 The IBM TotalStorage Enterprise Storage Server	2
1.2 The IBM TotalStorage Enterprise Storage Server Model 800	3
1.3 Benefits	3
1.3.1 Storage consolidation	4
1.3.2 Performance	4
1.3.3 Data protection	7
1.3.4 Storage Area Network (SAN)	9
1.4 Terminology	10
1.4.1 Host attachment	10
1.4.2 Data architecture	13
1.4.3 Server platforms	15
1.4.4 Other terms	17
Chapter 2. Hardware	19
2.1 IBM TotalStorage Enterprise Storage Server Model 800	20
2.1.1 Hardware characteristics	20
2.2 ESS Expansion Enclosure	21
2.3 Photograph of the ESS Model 800	22
2.4 ESS major components	23
2.5 ESS cages	24
2.6 ESS disks	25
2.6.1 ESS disk capacity	25
2.6.2 Disk features	26
2.6.3 Disk eight-packs	26
2.6.4 Disk eight-pack capacity	26
2.6.5 Disk intermixing	27
2.6.6 Disk conversions	28
2.6.7 Step Ahead option	28
2.7 Clusters	29
2.7.1 Processors	29
2.7.2 Cache	29
2.7.3 Non-volatile storage (NVS)	30
2.8 Device adapters	31
2.8.1 SSA 160 device adapters	31
2.8.2 Disk drives per loop	32
2.8.3 Hot spare disks	32

2.9 SSA loops	33
2.9.1 SSA operation	33
2.9.2 Loop availability	33
2.9.3 Spatial reuse	34
2.10 Host adapters	35
2.11 ESCON host adapters	36
2.12 SCSI host adapters	37
2.13 Fibre Channel host adapters	39
2.14 FICON host adapters	41
2.15 Fibre distances	42
2.16 Power supplies	43
2.16.1 Power characteristics	44
2.16.2 Battery backup	44
2.17 Other interfaces	45
2.18 ESS Master Console	46
2.18.1 ESS local area network	47
Chapter 3. Architecture	49
3.1 Overview	50
3.2 Data availability architecture	51
3.2.1 Data accessibility	51
3.2.2 Data protection	52
3.3 ESS availability features	53
3.3.1 Fault-tolerant subsystem	53
3.3.2 Cluster failover/failback	53
3.3.3 CUIR	54
3.4 Maintenance strategy	54
3.4.1 Call Home and remote support	54
3.4.2 SSR/CE dispatch	56
3.4.3 Concurrent maintenance	56
3.5 Concurrent logic maintenance	56
3.6 Concurrent power maintenance	58
3.7 Sparing	58
3.7.1 Sparing in a RAID rank	59
3.7.2 Replacement DDM is new spare	60
3.7.3 Capacity intermix sparing	60
3.8 Cluster operation: failover/failback	60
3.8.1 Normal operation before failover	61
3.8.2 Failover	62
3.8.3 Failback	62
3.9 CUIR	63
3.10 RAID Data protection	63
3.10.1 RAID ranks	63
3.10.2 RAID 5 rank	64
3.10.3 RAID 10 rank	65
3.10.4 Combination of RAID 5 and RAID 10 ranks	67
3.11 SSA device adapters	68
3.12 Logical Subsystems	68
3.12.1 Device adapters mapping	68
3.12.2 Ranks mapping	69
3.13 Host mapping to Logical Subsystem	71
3.13.1 SCSI and Fibre Channel mapping	71
3.13.2 CKD server mapping	71

3.14	Architecture characteristics	72
3.15	ESS Implementation - Fixed block	73
3.15.1	SCSI mapping	73
3.15.2	FCP mapping	74
3.16	ESS Implementation - CKD	75
3.17	CKD server view of ESS	76
3.18	CKD Logical Subsystem	77
3.19	SCSI server view of ESS	78
3.20	FB Logical Subsystem - SCSI attachment	79
3.21	Fibre Channel server view of ESS	80
3.22	FB Logical Subsystem - Fibre Channel attachment	81
3.23	iSeries	82
3.23.1	Single level storage	82
3.23.2	iSeries storage management	83
3.24	Data flow - host adapters	83
3.25	Data flow - read	84
3.25.1	Host adapter	85
3.25.2	Processor	85
3.25.3	Device adapter	85
3.25.4	Disk drives	85
3.26	Data flow - write	85
3.26.1	Host adapters	86
3.26.2	Processor	86
3.26.3	Device adapter	86
3.27	Cache and read operations	87
3.28	NVS and write operations	88
3.28.1	Write operations	89
3.28.2	NVS	89
3.29	Sequential operations - read	90
3.30	Sequential operations - write	91
3.31	zSeries I/O accelerators	92
3.31.1	Parallel Access Volumes (PAV)	92
3.31.2	Multiple Allegiance	92
3.31.3	I/O priority queuing	92
Chapter 4.	Configuration	93
4.1	Overview	94
4.2	Physical configuration	95
4.3	Storage capacity	95
4.3.1	Flexible capacity configurations	95
4.3.2	Step Ahead configurations	96
4.4	Logical configuration	97
4.4.1	Logical standard configurations	97
4.5	Base enclosure	98
4.6	Expansion Enclosure	99
4.7	Base and Expansion Enclosure loop configuration	100
4.8	Upgrade with eight-packs	100
4.9	Physical configuration	101
4.9.1	Cluster Processors	101
4.9.2	Cache	101
4.9.3	Host adapters	101
4.9.4	Device adapters	102
4.9.5	Performance accelerators	103

4.9.6	ESS copy functions	103
4.10	Loop configuration	103
4.11	Loop configuration – RAID 5 ranks	104
4.12	Loop configuration – RAID 10 ranks	105
4.13	Loop configuration – Mixed RAID ranks	106
4.13.1	RAID 5 versus RAID 10 considerations	106
4.13.2	Reconfiguration of RAID ranks	107
4.14	ESSNet setup	108
4.14.1	ESSNet and ESS Master Console	108
4.14.2	User local area network	109
4.14.3	Physical setup	109
4.14.4	Web browser	109
4.15	The ESS Specialist	110
4.15.1	ESS Specialist configuration windows	110
4.15.2	Logical standard configurations	111
4.16	ESS logical configuration	111
4.17	Logical configuration process	112
4.18	SCSI and Fibre Channel hosts and host adapters	113
4.19	ESCON and FICON host adapters	115
4.20	Defining Logical Subsystems	116
4.20.1	CKD Logical Subsystems	116
4.20.2	FB Logical Subsystems	118
4.21	Disk groups – ranks	118
4.22	RAID 5 and RAID 10 rank intermixing	120
4.23	Balancing RAID 10 ranks	122
4.24	Configuring CKD ranks	123
4.25	Configuring FB ranks	124
4.26	Assigning logical volumes to a rank	125
4.26.1	Rank capacities	126
4.26.2	Adding CKD logical volumes	126
4.26.3	Assigning iSeries logical volumes	128
4.26.4	Assigning fixed block logical volumes	128
4.27	Defining CKD logical devices	128
4.28	Defining FB logical devices	130
4.28.1	SCSI attached hosts	130
4.28.2	Fibre Channel-attached hosts	131
4.29	LSS/ranks configuration example	133
4.30	SCSI host connectivity	134
4.30.1	Single host connection	134
4.30.2	SCSI connection for availability	134
4.30.3	Multi-connection without redundancy	135
4.30.4	Daisy-chaining host SCSI adapters	135
4.31	Fibre Channel host connectivity	136
4.31.1	Fibre Channel topologies	137
4.31.2	Fibre Channel connection for availability	137
4.32	ESCON host connectivity	138
4.32.1	ESCON control unit images	138
4.32.2	ESCON logical paths establishment	139
4.32.3	Calculating ESCON logical paths	139
4.33	FICON host connectivity	140
4.33.1	FICON control unit images	141
4.33.2	Calculating FICON logical paths	142
4.34	ESCON and FICON connectivity intermix	144

4.35 Standard logical configurations	145
Chapter 5. Performance	147
5.1 Performance accelerators	148
5.2 Third-generation hardware	148
5.3 RISC SMP processors	149
5.4 Caching algorithms	150
5.4.1 Optimized cache usage	150
5.4.2 Sequential pre-fetch I/O requests	150
5.5 Efficient I/O operations	150
5.5.1 SCSI command tag queuing	151
5.5.2 z/OS enhanced CCWs	151
5.6 Back-end high performance design	151
5.6.1 Serial Storage Architecture (SSA)	151
5.6.2 Striping	152
5.6.3 RAID 10 vs RAID 5	153
5.6.4 High-performance disk drives	154
5.7 Configuring for performance	155
5.7.1 Host adapters	155
5.7.2 RAID ranks	156
5.7.3 Cache	156
5.7.4 Disk drives	156
5.8 Subsystem Device Driver	157
5.9 Measurement tools	159
5.10 IBM TotalStorage Expert	160
5.10.1 ESS Expert overview	160
5.10.2 How does the ESS Expert work	162
5.10.3 ESS Expert performance reports	162
5.11 z/OS environment tools	163
Chapter 6. zSeries performance	165
6.1 Overview	166
6.2 Parallel Access Volume	166
6.2.1 Traditional z/OS behavior	167
6.2.2 Parallel I/O capability	168
6.2.3 Benefits of Parallel Access Volume	169
6.2.4 PAV base and alias addresses	170
6.2.5 PAV tuning	171
6.2.6 Configuring PAVs	172
6.2.7 Querying PAVs	173
6.2.8 PAV assignment	174
6.2.9 WLM support for dynamic PAVs	174
6.2.10 Reassignment of a PAV alias	176
6.2.11 Mixing PAV types	177
6.2.12 PAV support	178
6.3 Multiple Allegiance	179
6.3.1 Parallel I/O capability	179
6.3.2 Eligible I/Os for parallel access	180
6.3.3 Software support	180
6.3.4 Benefits of Multiple Allegiance	181
6.4 I/O Priority Queuing	181
6.4.1 Queuing of channel programs	181
6.4.2 Priority queuing	182

6.5 Custom volumes	183
6.6 FICON host adapters	184
6.6.1 FICON benefits	185
6.6.2 FICON I/O operation	186
6.6.3 FICON benefits at your installation	190
6.7 Host adapters configuration	191
Chapter 7. Copy functions	193
7.1 ESS Copy Services functions	194
7.2 Managing ESS Copy Services	195
7.2.1 ESS Copy Services Web user interface	196
7.2.2 ESS Copy Services command-line interface (CLI)	198
7.2.3 TSO commands	199
7.3 ESS Copy Services setup	199
7.4 FlashCopy	200
7.4.1 Overview	201
7.4.2 Consistency	202
7.4.3 FlashCopy management on the zSeries	202
7.4.4 FlashCopy management on the open systems	203
7.5 Peer-to-Peer Remote Copy (PPRC)	203
7.5.1 PPRC overview	204
7.5.2 PPRC volume states	205
7.5.3 PPRC with static volumes	207
7.5.4 PPRC management on the zSeries	207
7.5.5 PPRC management on the open systems	209
7.6 PPRC implementation on the ESS	210
7.7 PPRC Extended Distance (PPRC-XD)	212
7.7.1 PPRC-XD operation	213
7.7.2 Data consistency	215
7.7.3 Automation	216
7.7.4 PPRC-XD for initial establish	216
7.7.5 Implementing and managing PPRC Extended Distance	216
7.8 PPRC connectivity	216
7.8.1 PPRC supported distances	217
7.8.2 PPRC channel extender support	217
7.8.3 PPRC Dense Wave Division Multiplexor (DWDM) support	218
7.9 Concurrent Copy	218
7.10 Extended Remote Copy (XRC)	220
7.10.1 Overview	220
7.10.2 Invocation and management of XRC	221
7.10.3 XRC implementation on the ESS	221
7.10.4 Coupled Extended Remote Copy (CXRC)	223
7.10.5 XRC FICON support	224
7.11 ESS Copy Services for iSeries	224
7.11.1 The Load Source Unit (LSU)	224
7.11.2 Mirrored internal DASD support	225
7.11.3 LSU mirroring	225
7.11.4 FlashCopy	225
7.11.5 PPRC	226
7.12 Geographically Dispersed Parallel Sysplex (GDPS)	227
Chapter 8. Support information	229
8.1 Key requirements information	230

8.2 zSeries environment support	230
8.2.1 Multiple Allegiance and I/O Priority Queuing	231
8.2.2 Parallel Access Volumes	231
8.2.3 PPRC	231
8.2.4 PPRC-XD	231
8.2.5 FlashCopy	231
8.2.6 FICON support	231
8.2.7 Control-Unit-Initiated Reconfiguration	232
8.2.8 Large Volume Support	232
8.3 z/OS support	232
8.3.1 Other related support products	234
8.4 z/VM support	235
8.4.1 Guest support	236
8.5 VSE/ESA support	236
8.6 TPF support	238
8.6.1 Control unit emulation mode	238
8.6.2 Multi Path Locking Facility	238
8.6.3 TPF support levels	238
8.7 Linux	239
8.8 Open systems environment support	240
8.8.1 Installation scripts	240
8.8.2 IBM Subsystem Device Driver	240
8.8.3 ESS Copy Services command-line interface (CLI)	241
8.8.4 Boot support	241
8.8.5 PPRC	241
8.8.6 PPRC-XD	241
8.8.7 FlashCopy	242
8.9 ESS Specialist	242
8.10 IBM TotalStorage Expert	242
Chapter 9. Installation planning and migration	245
9.1 Overview	246
9.2 Physical planning	246
9.2.1 Hardware configuration	246
9.2.2 Site requirements - dimensions and weight	247
9.2.3 Site requirements - power and others	247
9.3 Configuration planning	248
9.4 Installation and configuration	248
9.4.1 Physical installation	248
9.4.2 Configuration	248
9.5 Migration	249
9.6 Data migration in z/OS environments	250
9.7 Migrating data in z/VM	253
9.8 Migrating data in VSE/ESA	253
9.9 Data migration in UNIX environment	254
9.9.1 Migration methods	254
9.10 Migrating from SCSI to Fibre Channel	255
9.11 Migrating from ESCON to FICON	256
9.12 IBM migration services	259
9.12.1 Enhanced Migration Services - Piper	260
Appendix A. Feature codes	263
A.1 Overview	264

A.2 Major feature codes	264
A.2.1 Processors	264
A.2.2 Cache sizes	264
A.2.3 Host adapters	265
A.2.4 ESS Master Console	267
A.3 Disk storage configurations	268
A.3.1 Capacity range (physical capacity)	268
A.3.2 Disk eight-packs	268
A.4 Advanced Functions	270
A.4.1 ESS Advanced Functions	270
A.4.2 Capacity tier calculation	271
A.4.3 Ordering Advanced Functions	271
A.5 Additional feature codes	272
Related publications	275
IBM Redbooks	275
Other resources	275
Referenced Web sites	276
How to get IBM Redbooks	276
IBM Redbooks collections	276
Index	277

Figures

Ê	The team that wrote this book: Steve, Guenter, Gustavo, and Peter	xx
1-1	IBM's Seascape architecture - ESS Model 800	2
1-2	IBM TotalStorage Enterprise Storage Server for storage consolidation	4
1-3	IBM TotalStorage Enterprise Storage Server capabilities	5
1-4	Disaster recovery and availability	7
1-5	Storage Area Network (SAN)	9
1-6	ESCON and FICON host attachment components	10
1-7	SCSI and Fibre Channel host attachment components	12
2-1	IBM TotalStorage Enterprise Storage Server Model 800	20
2-2	ESS Model 800 base and Expansion Enclosures	21
2-3	Photograph of the ESS Model 800 base enclosure (front covers removed)	22
2-4	ESS Model 800 major components	23
2-5	ESS cages	24
2-6	ESS Model 800 disks	25
2-7	ESS Model 800 clusters	29
2-8	ESS device adapters	31
2-9	SSA loops	33
2-10	ESS Model 800 host adapters	35
2-11	ESCON host adapters	36
2-12	SCSI host adapters	37
2-13	Fibre Channel host adapters	39
2-14	FICON host adapters	41
2-15	ESS Model 800 power supplies	43
2-16	ESS Model 800 - other interfaces	45
2-17	ESS Master Console	46
3-1	ESS Model 800 design overview	50
3-2	Data availability design	51
3-3	ESS Model 800 availability features	53
3-4	ESS Model 800 maintenance strategy	54
3-5	Concurrent maintenance	57
3-6	Concurrent power maintenance	58
3-7	Sparing	59
3-8	Normal cluster operation	61
3-9	Cluster 1 failing - failover initiated	62
3-10	Cluster 1 recovered - failback is initiated	62
3-11	Initial rank setup	64
3-12	RAID 5 rank implementation	65
3-13	RAID 10 rank implementation	66
3-14	RAID 10 balanced configuration	67
3-15	RAID 5 and RAID 10 in the same loop	67
3-16	Logical Subsystems and device adapters mappings	69
3-17	Logical Subsystem and rank relationship	70
3-18	Host mapping	71
3-19	Architecture addressing characteristics	72
3-20	FB implementation for SCSI and FCP on the ESS	73
3-21	CKD implementation for ESCON and FICON on the ESS	75
3-22	CKD server view of the ESS	76
3-23	CKD LSS	77

3-24	SCSI server view of the ESS	78
3-25	FB LSS - SCSI attachment	79
3-26	Fibre Channel server view of ESS	80
3-27	FB LSS - Fibre Channel attachment	81
3-28	Data flow - host adapters	83
3-29	Data flow - read	84
3-30	Data flow - write	86
3-31	Cache - read	87
3-32	NVS - write	88
3-33	Sequential read	90
3-34	Sequential write	91
3-35	CKD accelerators	92
4-1	Configuration process	94
4-2	Physical configuration options	95
4-3	Logical configuration characteristics	97
4-4	ESS base enclosure— eight-pack installation sequence	98
4-5	ESS Expansion Enclosure — eight-pack installation sequence	99
4-6	Eight-pack logical loop configuration — with Expansion Enclosure	100
4-7	Block diagram of an ESS	101
4-8	DA pair - maximum loop configuration	103
4-9	Single-capacity DA pair loop configuration - RAID 5 ranks	104
4-10	Single capacity DA pair loop configuration - RAID 10 ranks	105
4-11	Mixed capacity DA pair loop configuration - RAID 10 and RAID 5 ranks	106
4-12	ESSNet setup	108
4-13	ESS Specialist — Storage Allocation window	110
4-14	Logical configuration terminology	111
4-15	Logical configuration process	112
4-16	Fibre Channel adapters	114
4-17	Configure Host Adapter Ports window — FICON connection	115
4-18	Logical Subsystem mapping	116
4-19	Configure LCU window	117
4-20	Disk groups and RAID ranks — initial setup	118
4-21	Disk group associations	119
4-22	RAID 10 followed by RAID 5 formatting	121
4-23	RAID 5 followed by RAID 10 formatting	121
4-24	Unbalanced RAID 10 arrays	122
4-25	Balanced LSSs	123
4-26	Configure Disk Group window	124
4-27	Fixed Block Storage window	125
4-28	Add CKD Volumes window	126
4-29	CKD logical device mapping	129
4-30	FB logical device mapping for SCSI attached hosts	131
4-31	FB logical device mapping for Fibre Channel-attached hosts	132
4-32	LSS ranks assignment	133
4-33	SCSI connectivity	134
4-34	SCSI daisy-chaining	135
4-35	Fibre Channel connectivity	136
4-36	ESCON connectivity example	138
4-37	Establishment of logical paths for ESCON attachment	139
4-38	FICON connectivity	140
4-39	Establishment of logical paths for FICON attachment	142
4-40	FICON connectivity example	143
4-41	Over-defined paths — system message	144

5-1	ESS Performance accelerators	148
5-2	RISC SMP processors (simplified diagram).	149
5-3	RAID 5 logical volume striping.	152
5-4	RAID 10 logical volume striping.	153
5-5	Disk drive	154
5-6	Subsystem Device Driver (SDD)	157
5-7	Performance measurement tools.	159
5-8	How the Expert communicates with the ESS.	162
6-1	zSeries specific performance features.	166
6-2	Traditional z/OS behavior	167
6-3	Parallel I/O capability using PAV	168
6-4	Parallel Access Volume (PAV) benefits	169
6-5	Potential performance impact of PAV	170
6-6	PAV base and alias addresses	170
6-7	PAV tuning.	171
6-8	Modify PAV Assignments window	172
6-9	Querying PAVs	173
6-10	Assignment of alias addresses	174
6-11	Dynamic PAVs in a sysplex.	175
6-12	Reassignment of dynamic PAV alias	176
6-13	Activation of dynamic alias tuning for the WLM.	177
6-14	No mixing of PAV types	178
6-15	Parallel I/O capability with Multiple Allegiance.	179
6-16	Multiple Allegiance.	180
6-17	Benefits of Multiple Allegiance for mixing workloads	181
6-18	I/O queuing	182
6-19	I/O Priority Queuing	183
6-20	CKD custom volumes	184
6-21	ESCON cache hit I/O operation sequence	187
6-22	ESCON cache miss I/O operation sequence.	188
6-23	FICON cache hit I/O operation sequence	189
6-24	FICON cache miss I/O operation sequence	190
7-1	ESS Copy Services for open systems	194
7-2	ESS Copy Services for zSeries	195
7-3	ESS Specialist Welcome window	196
7-4	ESS Copy Services Welcome window.	197
7-5	ESS Copy Services setup	200
7-6	FlashCopy point-in-time copy	201
7-7	Synchronous volume copy PPRC	204
7-8	PPRC volume states - synchronous mode	205
7-9	PPRC configuration options.	210
7-10	PPRC links.	211
7-11	PPRC logical paths	212
7-12	PPRC Extended Distance (PPRC-XD)	213
7-13	Duplex-pending XD volume state	214
7-14	PPRC-XD Basic operation.	214
7-15	Connectivity - distance and support considerations.	217
7-16	Concurrent Copy	219
7-17	Extended Remote Copy.	220
7-18	XRC implementation	221
7-19	XRC unplanned outage support.	222
7-20	CXRC performance and scalability	223
7-21	XRC FICON support	224

7-22	I Series and ESS PPRC implementation	226
8-1	IBM TotalStorage Expert Welcome window.	243
9-1	Several options - hardware components and advanced functions	247
9-2	Preparation for zSeries data migration	250
9-3	UNIX data migration methods	254
9-4	ESS configuration with ESCON adapters	257
9-5	Interim configuration with ESCON and FICON intermix	258
9-6	Target FICON ESS migration configuration.	259
9-7	Piper CKD migration concurrent or nonconcurrent	260
9-8	Piper FB migration concurrent or nonconcurrent.	261

Tables

- 2-1 Disk eight-pack effective capacity chart (gigabytes) 27
- 2-2 Fibre distances. 43
- 4-1 RAID 5 versus RAID 10 overheads compared to non-RAID 107
- 4-2 CKD logical device capacities 127
- 8-1 Supported zSeries and S/390 operating systems 230
- 8-2 Processor PSP buckets for FICON support. 232
- 8-3 Processor and Channel PSP buckets for z/OS and OS/390 233
- 8-4 Processor PSP buckets for z/VM and VM/ESA. 236
- 8-5 Processor PSP buckets for VSE/ESA 237
- A-1 ESS processor options 264
- A-2 ESS cache capacities (Gigabytes). 265
- A-3 ESS host adapters. 265
- A-4 2 Gb Fibre Channel/FICON cable feature codes. 266
- A-5 SCSI cable feature codes 267
- A-6 Disk eight-pack feature codes 268
- A-7 ESS Function Authorization fe9.9(a)9.9(t)0.2(u)9.9(r)6.4(e9.9(s)-119.2(.)24.6(.)24.4(.)24.4(.)24.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.



This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both: :

e (logo)® 	NUMA-Q®
Redbooks Logo™ 	NUMACenter™
AIX®	OS/390®
AS/400®	OS/400®
CICS®	PR/SM™
CUA®	Predictive Failure Analysis®
DB2®	PR/SM™
DFS™	pSeries™
DFSMS/MVS®	RACF®
DFSMS/VM®	RAMAC®
DFSMSdfp™	Redbooks™
DFSMSdss™	RETAIN®
DFSMSHsm™	RMF™
DFSORT™	RS/6000®
DYNIX®	S/390®
DYNIX/ptx®	Seascape®
Enterprise Storage Server™	Sequent®
Enterprise Systems Connection® Architecture®	SP™
ESCON®	StorWatch™
FICON™	System/390®
FlashCopy™	Tivoli®
IBM®	TotalStorage™
IMS™	VM/ESA®
IMS/ESA®	VSE/ESA™
iSeries™	xSeries™
Magstar®	z/OS™
MVS™	z/VM™
	zSeries™

The following terms are trademarks of International Business Machines Corporation and Lotus Development Corporation in the United States, other countries, or both:

Lotus®	Notes®	Word Pro®
--------	--------	-----------

The following terms are trademarks of other companies:

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

C-bus is a trademark of Corollary, Inc. in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

Other company, product, and service names may be trademarks or service marks of others.

Preface

This IBM Redbook describes the IBM TotalStorage Enterprise Storage Server Model 800, its architecture, its logical design, hardware design and components, advanced functions, performance features, and specific characteristics. The information contained in this redbook will be useful for those who need a general understanding of this powerful model of disk enterprise storage server, as well as for those looking for a more detailed understanding of how the ESS Model 800 is designed and operates.

In addition to the logical and physical description of the ESS Model 800, the fundamentals of the configuration process are also described in this redbook. This is all useful information for the IT storage person for proper planning and configuration when installing the ESS, as well as for the efficient management of this powerful storage subsystem.

Characteristics of the ESS Model 800 described in this redbook include the ESS copy functions: FlashCopy, Peer-to-Peer Remote Copy, PPRC Extended Distance, Extended Remote Copy, and Concurrent Copy. The performance features of the ESS are also explained, so that the user can better optimize the storage resources at the computing center.

Other characteristics of the IBM TotalStorage Enterprise Storage Server Model 800 described in this redbook include:

- ▶ New cluster SMP processors, with a Turbo feature option
- ▶ 2 GB non-volatile storage
- ▶ Double bandwidth CPI (common parts interconnect)
- ▶ 2 Gb Fibre Channel/FICON host adapters
- ▶ 64 GB cache option
- ▶ New, more powerful SSA device adapters
- ▶ RAID 10 and RAID 5 rank configuration
- ▶ Disk capacity intermix
- ▶ 18.2 GB, 36.4 GB, and 72.8 GB disks
- ▶ 10,000 and 15,000 rpm disks
- ▶ Disk speed intermix (rpm)
- ▶ 32 K cylinder large-volume support
- ▶ CUIR (control-unit-initiated reconfiguration)
- ▶ ESS Master Console
- ▶ Additional PPRC connectivity options

The team that wrote this redbook

This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization, San Jose Center.

Gustavo Castets is a Project Leader at the International Technical Support Organization, San Jose Center. He has co-authored four previous redbooks and teaches IBM classes worldwide on areas of disk storage systems. Before joining the ITSO, Gustavo worked in System Sales as a Field Technical Support Specialist. Gustavo has worked in Buenos Aires for more than 22 years in many IT areas for IBM Argentina.

Peter Crowhurst is a Consulting IT Specialist in the Technical Sales Support group within Australia. He has 25 years of experience in the IT industry in network planning and design, including nine years working in a customer organization as an Applications and Systems

Programmer. Peter joined IBM 16 years ago and has worked mainly in a large systems technical pre-sales support role for zSeries and storage products. He was an author of the original *Implementing the ESS* redbook, SG24-5420, produced in 1999.

Stephen Garraway is an IT Specialist in the Storage Management team at the Hursley Laboratory in the UK. He has 14 years of experience in storage management, mainly with VM and TSM, and has worked for IBM for 16 years. His areas of expertise include zSeries storage, particularly the IBM TotalStorage Enterprise Storage Server and the IBM TotalStorage Enterprise Tape Library, and more recently Storage Area Networks and pSeries storage.

Guenter Rebmann is a DASD support specialist in Germany. He has five years of experience working as a CE for S/390 customers. Since 1993, he has been a member of the EMEA DASD support group located in Mainz/Germany. His areas of expertise include all large-system DASD products, particularly the IBM TotalStorage Enterprise Storage Server.



The team that wrote this book: Steve, Guenter, Gustavo, and Peter

Thanks to the following people for their invaluable contributions to this project:

Bill Avila, from IBM San Jose
Andreas Baer, from IBM Mainz, Germany
Helen Burton, from IBM Tucson
Thomas Fiege, from IBM San Jose
Martin Hitchman, from IBM Hursley, UK
Nicholas Kum, from IBM San Jose
Cynthia Regens, from IBM Tucson
Richard Ripberger, from IBM Tucson
David Sacks, from IBM Chicago

Special thanks to the following people for their diligence in reviewing this book:

Gary Albert, from IBM Tucson
Charlie Burger, from IBM San Jose
Ron Chapman, from IBM Tucson
Jennifer Eaton, from IBM San Jose
Siebo Friesenborg, from IBM Dallas
Lee La Frese, from IBM Tucson
Phil Lee, from IBM Markham, Canada
Charles Lynn, from IBM Tucson
Michael Purcell, from IBM Tucson
Dave Reeve, from IBM UK
Brian Sherman, from IBM Markham, Canada

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbook dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You'll team with IBM technical professionals, Business Partners and/or customers.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you'll develop a network of contacts in IBM development labs, and increase your productivity and marketability.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our Redbooks to be as helpful as possible. Send us your comments about this or other Redbooks in one of the following ways:

- ▶ Use the online **Contact us** review redbook form found at:

ibm.com/redbooks

- ▶ Send your comments in an Internet note to:

redbook@us.ibm.com

- ▶ Mail your comments to:

IBM Corporation, International Technical Support Organization
Dept. QXXE Building 80-E2
650 Harry Road
San Jose, California 95120-6099



Introduction

This chapter introduces the IBM TotalStorage Enterprise Storage Server Model 800 and discusses some of the benefits that can be achieved when using it. Then some basic terms used in this book are also described.

1.1 The IBM TotalStorage Enterprise Storage Server

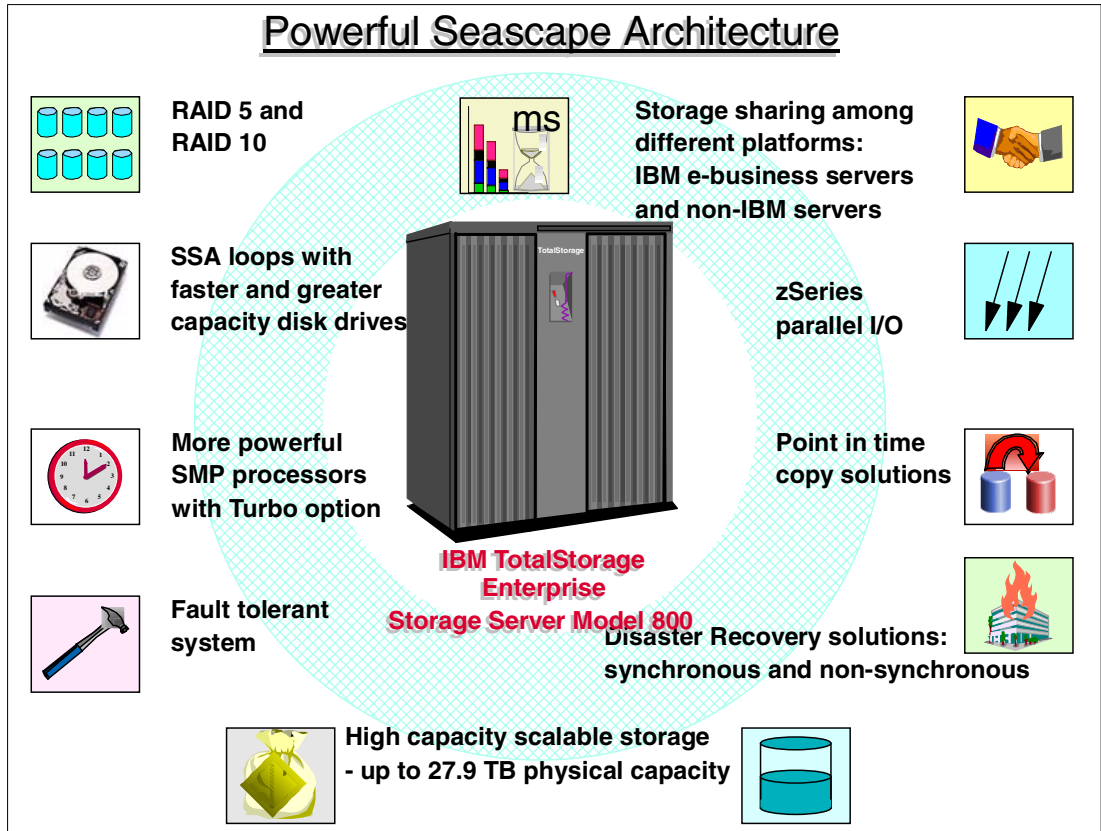


Figure 1-1 IBM's Seascope architecture - ESS Model 800

The IBM TotalStorage Enterprise Storage Server (ESS) is IBM's most powerful disk storage server, developed using IBM Seascope architecture. The ESS provides un-matchable functions for all the @server family of e-business servers, and also for the non-IBM (that is, Intel-based and UNIX-based) families of servers. Across all of these environments, the ESS features unique capabilities that allow it to meet the most demanding requirements of performance, capacity, and data availability that the computing business may require.

The Seascope architecture is the key to the development of IBM's storage products. Seascope allows IBM to take the best of the technologies developed by the many IBM laboratories and integrate them, producing flexible and upgradeable storage solutions. This Seascope architecture design has allowed the IBM TotalStorage Enterprise Storage Server to evolve from the initial E models to the succeeding F models, and to the recently announced 800 models, each featuring new, more powerful hardware and functional enhancements and always integrated under the same successful architecture with which the ESS was originally conceived.

The move to e-business presents companies with both extraordinary opportunities and significant challenges. A whole new world of potential customers, automated and streamlined processes, and new revenue streams are being fueled by e-business. Consequently, companies also face an increase of critical requirements for more information that is universally available online, around the clock, every day of the year.

To meet the unique requirements of e-business, where massive swings in the demands placed on your systems are common and continuous operation is imperative, you'll need very

high-performance, intelligent storage technologies and systems that can support any server application in your business, today and into the future. The IBM TotalStorage Enterprise Storage Server has set new standards in function, performance, and scalability in these most challenging environments.

1.2 The IBM TotalStorage Enterprise Storage Server Model 800

Since its initial availability with the ESS Models E10 and E20, and then with the succeeding F10 and F20 models, the ESS has been the storage server solution offering exceptional performance, extraordinary capacity, scalability, heterogeneous server connectivity, and an extensive suite of advanced functions to support customers' mission-critical, high-availability, multi-platform environments. The ESS set a new standard for storage servers back in 1999 when it was first available, and since then it has evolved into the F models and the recently announced third-generation ESS Model 800, keeping up with the pace of customers' needs by adding more sophisticated functions to the initial set, enhancing the connectivity options, and powering its performance features.

The IBM TotalStorage Enterprise Storage Server Model 800 provides significantly improved levels of performance, throughput, and scalability while continuing to exploit the innovative features introduced with its preceding E and F models such as Parallel Access Volumes, Multiple Allegiance, I/O Priority Queuing, the remote copy functions (synchronous and asynchronous), and the FlashCopy point-in-time copy function. Also the heterogeneous server support characteristic—for connectivity and remote copy functions—of previous models is continued with the ESS Model 800, in addition to the enhancement features that were introduced more recently with the F models, such as disk capacity intermix, ESS Master Console, S/390 CUIR, 32 K cylinder large-volume support, 72.8 GB disk drives, new 18.2 GB and 36.4 GB 15000 rpm disk drives, non-synchronous PPRC Extended Distance, and the additional connectivity options for the remote copy functions.

The new IBM TotalStorage Enterprise Storage Server Model 800 introduces important changes that dramatically improve the overall value of ESS in the marketplace and provide a strong base for strategic Storage Area Network (SAN) initiatives.

Among the enhancements introduced in the Model 800 are:

- ▶ 2 Gb Fibre Channel/FICON host adapter
- ▶ Up to 64 GB of cache
- ▶ New, more powerful SMP cluster processors with a Turbo feature option
- ▶ 2 GB non-volatile storage (NVS) with double the bandwidth
- ▶ A doubling of the bandwidth of the Common Platform Interconnect (CPI) RAID 10 array configuration capability

These hardware enhancements introduced by the new IBM TotalStorage Enterprise Storage Server Model 800 all combine to provide a balanced two-fold performance boost as compared to the predecessor F models, and up to two-and-a-half boost with the Turbo processor option.

1.3 Benefits

The new IBM TotalStorage Enterprise Storage Server Model 800 can help you achieve your business objectives in many areas. It provides a high-performance, high-availability subsystem with flexible characteristics that can be configured according to your requirements.

1.3.1 Storage consolidation

The ESS attachment versatility —and large capacity— enable the data from different platforms to be consolidated onto a single high-performance, high-availability box. Storage consolidation can be the first step towards server consolidation, reducing the number of boxes you have to manage and allowing you to flexibly add or assign capacity when it is needed. The IBM TotalStorage Enterprise Storage Server supports all the major operating systems platforms, from the complete set of IBM @server series of e-business servers and IBM NUMA-Q, to the non-IBM Intel-based servers and the different variations of UNIX based servers, as shown in Figure 1-2.

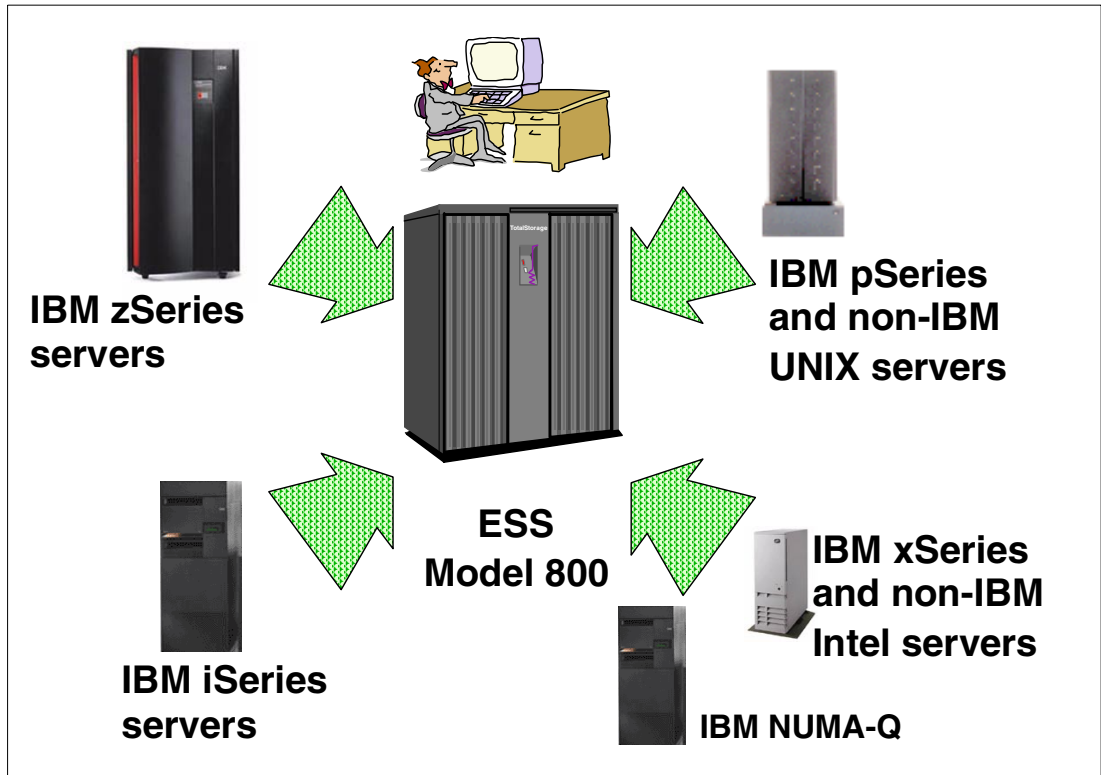


Figure 1-2 IBM TotalStorage Enterprise Storage Server for storage consolidation

With a total capacity of more than 27 TB, and a diversified host attachment capability —SCSI, ESCON, and Fibre Channel/FICON— the IBM TotalStorage Enterprise Storage Server Model 800 provides outstanding performance while consolidating the storage demands of the heterogeneous set of server platforms that must be dealt with nowadays.

1.3.2 Performance

The IBM TotalStorage Enterprise Storage Server is a storage solution with a design for high performance that takes advantage of IBM's leading technologies.

In today's world, you need business solutions that can deliver high levels of performance continuously every day, day after day. You also need a solution that can handle different workloads simultaneously, so you can run your business intelligence models, your large databases for enterprise resource planning (ERP), and your online and Internet transactions alongside each other. Some of the unique features that contribute to the overall high-performance design of the IBM TotalStorage Enterprise Storage Server Model 800 are shown in Figure 1-3 on page 5.

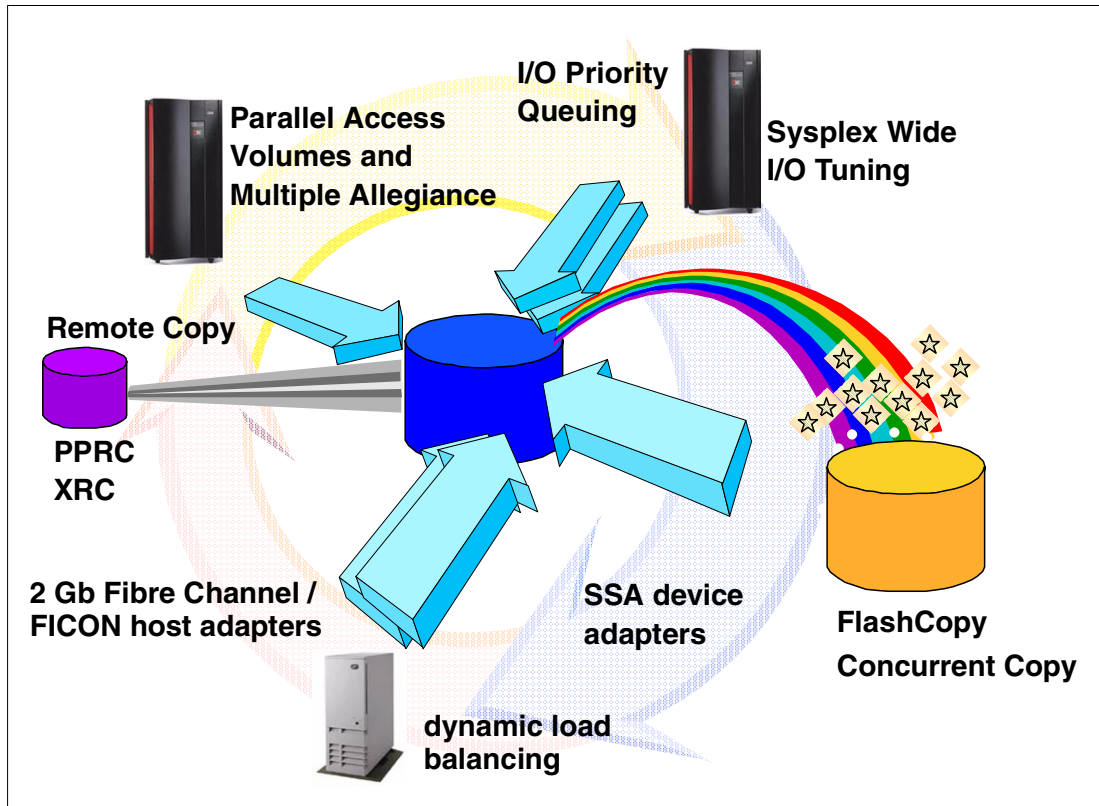


Figure 1-3 IBM TotalStorage Enterprise Storage Server capabilities

Third-generation hardware - ESS Model 800

The IBM TotalStorage Enterprise Storage Server Model 800 integrates a new generation of hardware from top to bottom, allowing it to deliver unprecedented levels of performance and throughput. Key features that characterize the performance enhancements of the ESS Model 800 are:

- ▶ The ESS Model 800 generally is capable of delivering twice the throughput of its predecessor Model F20.
- ▶ With the optional Turbo feature, it is capable of providing 2.5 times the throughput of its predecessor Model F20, for increased scalability and response times.
- ▶ 64 GB cache supports much larger system configurations and increases hit ratios, driving down response times.
- ▶ Double the internal bandwidth provides high sequential throughput for digital media, business intelligence, data warehousing, and life science application.
- ▶ Larger NVS with twice the bandwidth allows greater scalability for write-intensive applications.
- ▶ Third-generation hardware provides response time improvements of up to 40% for important database applications.
- ▶ 2 Gb Fibre Channel/FICON host adapters provide doubled performance sustained and instantaneous throughput for both open systems and zSeries environments.
- ▶ RAID 10 can provide up to 75% greater throughput for selected database workloads compared to equal physical capacity configured as RAID 5. While most typical workloads will experience excellent response times with RAID 5, some cache-unfriendly applications

and some applications with high random write content can benefit from the performance offered by RAID 10.

- ▶ 15,000 rpm drives provide up to 80% greater throughput per RAID rank and 40% improved response time as compared to 10,000 rpm drives. This allows driving the workloads to significantly higher access densities, while also experiencing improved response times.

All this performance boost is built upon the reliable and proven ESS hardware architecture design and unique advanced features.

Efficient cache management and powerful back-end

The ESS is designed to provide the highest performance, for the different type of workloads, even when mixing dissimilar workload demands. For example, zSeries servers and open systems put very different workload demands on the storage subsystem. A server like the zSeries typically has an I/O profile that is very cache-friendly, and takes advantage of the cache efficiency. On the other hand, an open system server does an I/O that can be very cache-unfriendly, because most of the hits are solved in the host server buffers. For the zSeries type of workload, the ESS has the option of a large cache (up to 64 GB) and —most important — it has efficient cache algorithms. For the cache unfriendly workloads, the ESS has a powerful back-end, with the SSA high performance disk adapters providing high I/O parallelism and throughput for the ever-evolving high-performance hard disk drives.

Plus the IBM TotalStorage Enterprise Storage Server Model 800 is introducing new more powerful hardware features that double the performance of its predecessor F Model:

- ▶ New more powerful SSA device adapters
- ▶ Double CPI (Common Platform Interconnect) bandwidth
- ▶ Larger cache option (64 GB)
- ▶ Larger NVS (2 GB non-volatile storage) and bandwidth
- ▶ New, more powerful SMP dual active controller processors, with a Turbo feature option
- ▶ 2 Gb Fibre Channel/FICON server connectivity, doubling the bandwidth and instantaneous data rate of previous host adapters

Sysplex I/O management

In the zSeries Parallel Sysplex environments, the z/OS Workload Manager (WLM) controls where work is run and optimizes the throughput and performance of the total system. The ESS provides the WLM with more sophisticated ways to control the I/O across the sysplex. These functions, described in detail later in this book, include parallel access to both single-system and shared volumes and the ability to prioritize the I/O based upon WLM goals. The combination of these features significantly improves performance in a wide variety of workload environments.

Parallel Access Volume (PAV) and Multiple Allegiance

Parallel Access Volume and Multiple Allegiance are two distinctive performance features of the IBM TotalStorage Enterprise Storage Server for the zSeries users, allowing them to reduce device queue delays, which means improving throughput and response time.

I/O load balancing

For selected open system servers, the ESS in conjunction with the Subsystem Device Driver (SDD), a pseudo device driver designed to support multipath configurations, provides dynamic load balancing. Dynamic load balancing helps in the elimination of data-flow bottlenecks by distributing the I/O workload over the multiple active paths, thus contributing to improve the I/O throughput and response time of the open system server.

2 Gb Fibre Channel/FICON host adapters

As the amount of data and transactions grow, so does the traffic over the storage area networks (SAN). As SANs migrate to 2 Gb technologies to cope with this increased amount of data transit, so does the IBM TotalStorage Enterprise Storage Server Model 800 with its 2 Gb host adapters. These host adapters double the bandwidth of the previous adapters, thus providing more throughput and performance for retrieving and storing users' data.

1.3.3 Data protection

Many design characteristics and advanced functions of the IBM TotalStorage Enterprise Storage Server Model 800 contribute to protect the data in an effective manner.

Fault-tolerant design

The IBM TotalStorage Enterprise Storage Server is designed with no single point of failure. It is a fault-tolerant storage subsystem, which can be maintained and upgraded concurrently with user operation. Some of the functions that contribute to these attributes of the ESS are shown in Figure 1-4.

RAID 5 or RAID 10 data protection

With the IBM TotalStorage Enterprise Storage Server Model 800, there now is the additional option of configuring the disk arrays in a RAID 10 disposition (mirroring plus striping) in addition to the RAID 5 arrangement, which gives more flexibility when selecting the redundancy technique for protecting the users' data.

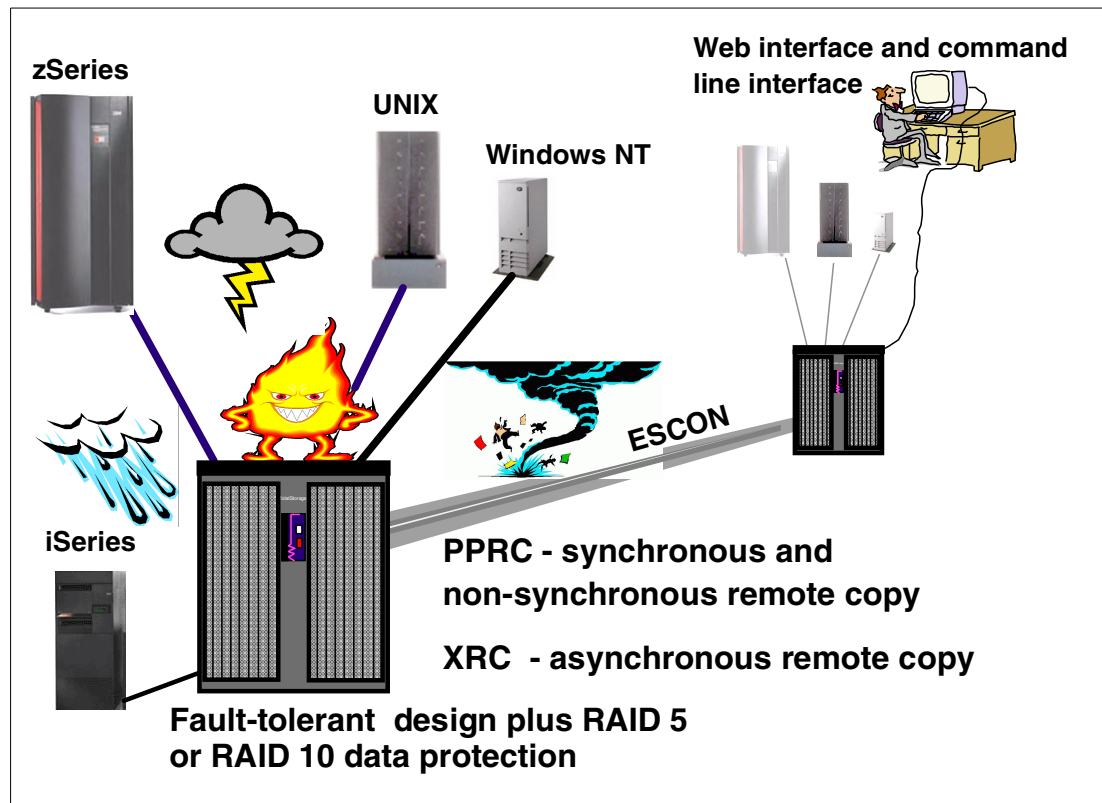


Figure 1-4 Disaster recovery and availability

Remote copy functions

The IBM TotalStorage Enterprise Storage Server Model 800 provides a set of remote copy functions (illustrated in Figure 1-3 on page 5 and Figure 1-4 on page 7) that allow you to more flexibly plan your business continuance solution.

Peer-to-Peer Remote Copy (PPRC)

The Peer-to-Peer Remote Copy (PPRC) function is a hardware-based solution for mirroring logical volumes from a primary site (the application site) onto the volumes of a secondary site (the recovery site). PPRC is a remote copy solution for the open systems servers and for the zSeries servers.

Two modes of PPRC are available with the IBM TotalStorage Enterprise Storage Server Model 800:

- ▶ PPRC synchronous mode, for real-time mirroring between ESSs located up to 103 km apart
- ▶ PPRC Extended Distance (PPRC-XD) mode, for non-synchronous data copy over continental distances

PPRC can be managed using a Web browser to interface with the ESS Copy Services Web user interface (WUI). PPRC can also be operated using commands for selected open systems servers that are supported by the ESS Copy Services command-line interface (CLI), and for the z/OS and OS/390 environments using TSO commands.

PPRC channel extension, DWDM, and connectivity options

PPRC flexibility is further enhanced with support for additional network connectivity options when using channel extenders. PPRC is supported over all the network technologies that are currently supported by the CNT UltraNet Storage Director or the INRANGE 9801 Storage Networking System, including Fibre Channel, Ethernet/IP, ATM-OC3, and T1/T3. Also the Cisco ONS 15540 DWDM (Dense Wave Division Multiplexer) and the Nortel Networks OPTera Metro 5300 DWDM are supported for PPRC connectivity.

These new options support the exploitation of existing or new communication infrastructures and technologies within metropolitan networks and the WAN to help optimize cost, performance, and bandwidth.

Extended Remote Copy (XRC)

Extended Remote Copy (XRC) is a combined hardware and software remote copy solution for the z/OS and OS/390 environments. The asynchronous characteristics of XRC make it suitable for continental distance implementations.

Point-in-Time Copy function

Users still need to take backups to protect data from logical errors and disasters. For all environments, taking backups of user data traditionally takes a considerable amount of time. Usually backups are taken outside prime shift because of their duration and the consequent impact to normal operations. Databases must be closed to create consistency and data integrity, and online systems are normally shut down.

With the IBM TotalStorage Enterprise Storage Server Model 800, the backup time has been reduced to a minimal amount of time when using the FlashCopy function. FlashCopy creates an instant point-in-time copy of data, and makes it possible to access both the source and target copies immediately, thus allowing the applications to resume with minimal disruption.

For all server platforms, FlashCopy can be controlled using a Web browser by means of the ESS Copy Services Web user interface of the IBM TotalStorage Enterprise Storage Server. Under z/OS, FlashCopy can also be invoked using DFSMSdss, and TSO commands.

For selected open systems servers the IBM TotalStorage Enterprise Storage Server also provides the ESS Copy Services command-line interface (CLI) for invocation and management of FlashCopy functions through batch processes and scripts.

1.3.4 Storage Area Network (SAN)

The third-generation IBM TotalStorage Enterprise Storage Server Model 800 continues to deliver on its SAN strategy, initiated with its predecessors E and F models. As SANs migrate to 2 Gb technology, then storage subsystems must exploit this more powerful bandwidth. Keeping pace with the evolution of SAN technology, the IBM TotalStorage Enterprise Storage Server Model 800 is introducing new 2 Gb Fibre Channel/FICON host adapters for native server connectivity and SAN integration.

These new 2 Gb Fibre Channel/FICON host adapters, which double the bandwidth and instantaneous data rate of the previous adapters available with the F Model, have one port with an LC connector for full-duplex data transfer over long-wave or short-wave fiber links. These adapters support the SCSI-FCP (Fibre Channel Protocol) and the FICON upper-level protocols.

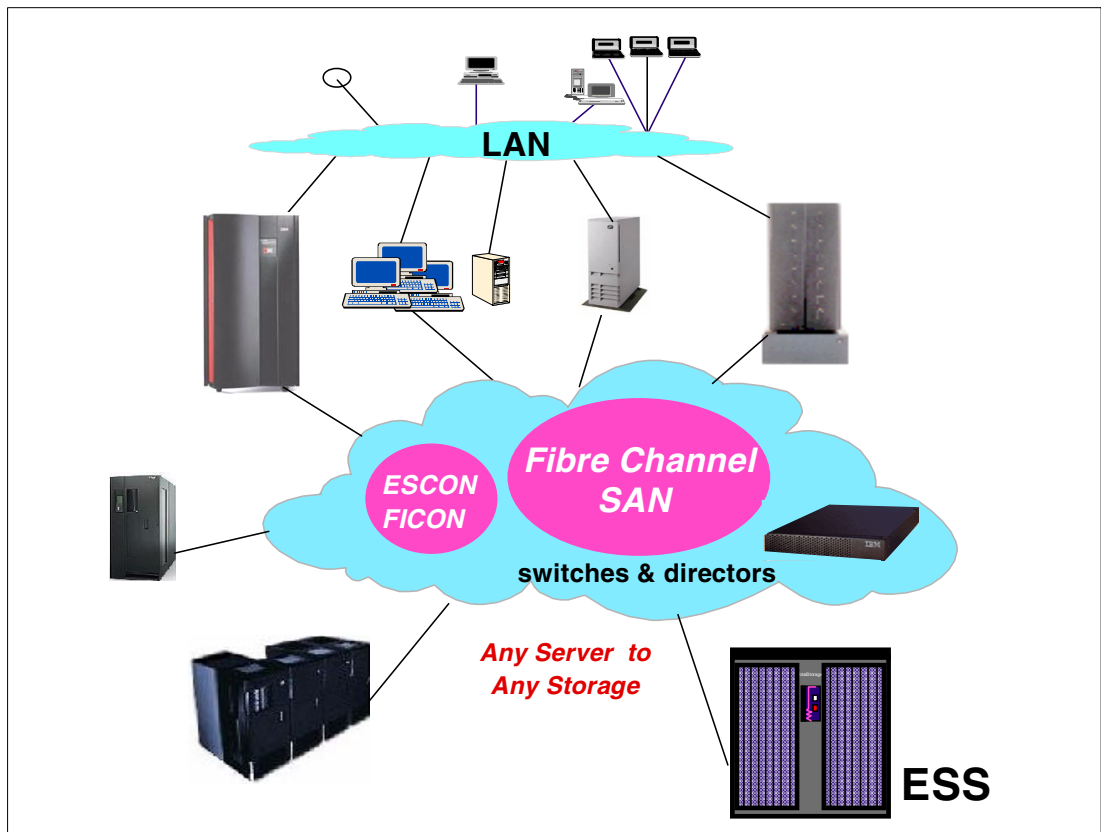


Figure 1-5 Storage Area Network (SAN)

Fabric support now includes the following equipment:

- ▶ IBM TotalStorage SAN switches IBM 3534 Model F08 and IBM 2109 Models F16, S08 and S16
- ▶ McDATA “Enterprise to Edge” Directors (IBM 2032 Model 064) for 2 Gb FICON and FCP attachment (up to 64 ports)
- ▶ McDATA 16 and 12 port switches for FCP attachment (IBM 2031)

- ▶ INRANGE FC/9000 Director for FCP attachment (up to 64 and 128 ports) and FICON attachment (up to 256 ports) — IBM 2042 Models 001 and 128

The ESS supports the Fibre Channel/FICON intermix on the INRANGE FC/9000 Fibre Channel Director and the McDATA ED-6064 Enterprise Fibre Channel Director. With Fibre Channel/FICON intermix, both FCP and FICON upper-level protocols can be supported within the same director on a port-by-port basis. This new operational flexibility can help users to reduce costs with simplified asset management and improved asset utilization.

The extensive connectivity capabilities make the ESS the unquestionable choice when planning the SAN solution. For the complete list of the ESS fabric support, please refer to:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

For a description of the IBM TotalStorage SAN products, please refer to:

<http://www.storage.ibm.com/ibmsan/products/sanfabric.html>

1.4 Terminology

Before starting to look at the hardware, architecture, and configuration characteristics of the IBM TotalStorage Enterprise Storage Server Model 800, let's review some of the terms more commonly used throughout this book.

1.4.1 Host attachment

Figure 1-6 and Figure 1-7 on page 12 illustrate the elements involved in the attachment of the storage server to the host — the host attachment components.

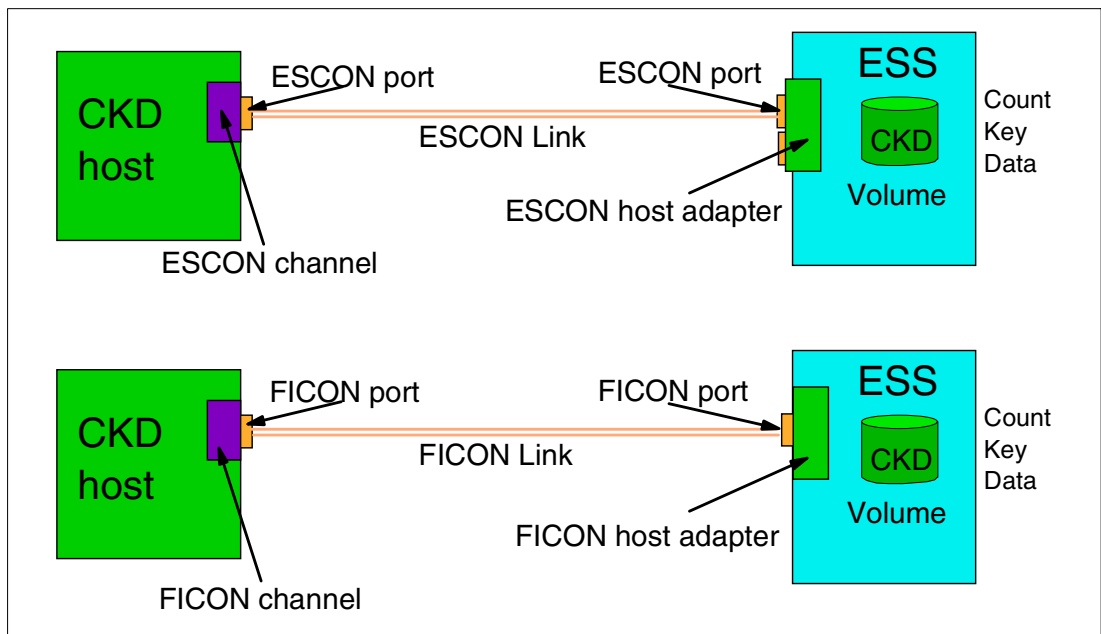


Figure 1-6 ESCON and FICON host attachment components

ESCON channel

The *ESCON channel* is the hardware feature on the zSeries and S/390 servers that controls data and command operations over the ESCON link — it is the *host I/O interface*. An ESCON

channel is usually installed on an ESCON channel card, which may contain up to four ESCON channels, depending upon host type.

ESCON host adapter

The *ESCON host adapter* (HA) is the physical component of the ESS used to attach the host ESCON I/O interfaces and ESCON Director ports. The ESCON host adapter connects to an ESCON channel by means of an ESCON link and accepts the CCWs (channel command words) from the host system. The ESS ESCON host adapters have two ports to connect ESCON links.

ESCON port

The *ESCON port* is the physical interface into the ESCON channel. An ESCON port has an ESCON connector interface. You have an ESCON port wherever you plug in an ESCON link.

ESCON link

An *ESCON link* is the fiber connection between the zSeries server ESCON channel and the ESS ESCON host adapter port. An ESCON link can also exist between a zSeries processor and an ESCON Director (fiber switch), and between an ESCON Director and the ESS (or other ESCON capable devices).

FICON channel

The *FICON channel* is the hardware feature on the zSeries and on the IBM 9672 G5 and G6 servers that controls data and command operations over the FICON link. It is the *host I/O interface*. A FICON channel is usually installed on a FICON channel card, which contains up to two FICON channels.

FICON host adapter

The *FICON host adapter* (HA) is the physical component of the ESS used to attach the host FICON I/O interfaces and FICON Director ports. The FICON host adapter connects to a FICON channel by means of the FICON link and accepts the CCWs (channel command words) from the host system. The ESS FICON host adapter, which in fact is a Fibre Channel/FICON adapter card that can be configured either for FICON or for FCP use, has one 2 Gb port to connect the FICON link.

FICON port

The *FICON port* is the physical interface into the FICON channel. The ESS FICON port connects the FICON link using an LC connector.

FICON link

A *FICON link* is the fiber connection between the zSeries server and the storage server. A FICON link can also exist between a zSeries processor and a FICON Director (switch), and between a FICON switch and the ESS (or other FICON-capable devices).

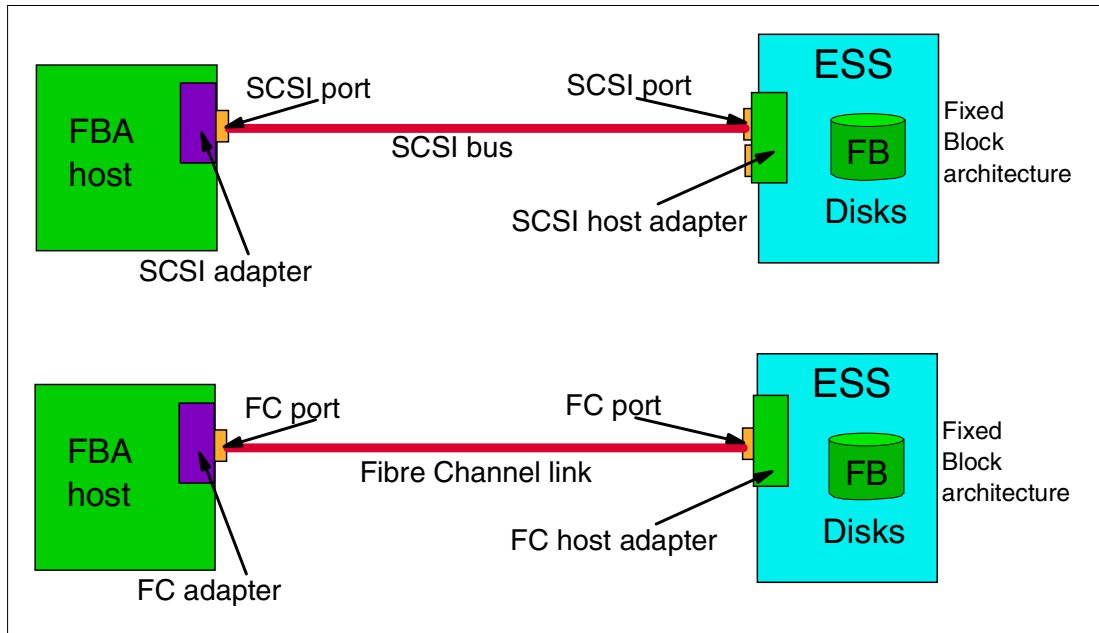


Figure 1-7 SCSI and Fibre Channel host attachment components

SCSI adapter

A *SCSI adapter* is a card installed in a host system to interface with the connected devices. It is a *host I/O interface* card. It connects to the SCSI bus through a SCSI connector. There are different versions of SCSI, some of which can be supported by the same adapter. The protocols that are used on the SCSI adapter (the command set) can be either SCSI-2 or SCSI-3.

SCSI bus

The *SCSI bus* is the physical path — cables — linking all the devices that are chained to the same SCSI adapter (daisy chaining). Each device on the bus is connected to the next one by a SCSI cable and the last device on the bus has a terminator.

SCSI port

A *SCSI port* is the physical interface into which you connect a SCSI cable. The physical interface varies, depending on what level of SCSI is supported.

SCSI host adapter

The SCSI *host adapter* (HA) is the physical component of the ESS used to attach the SCSI host I/O interfaces. The SCSI host adapter is connected to the SCSI bus and accepts the SCSI commands that are sent by the host system. The ESS SCSI host adapter has two ports for SCSI bus connection.

SCSI

SCSI (Small Computer Systems Interface) is the protocol that the SCSI adapter cards use. Although SCSI protocols can be used on Fibre Channel (then called FCP), most people mean the parallel interface when they say SCSI.

Fibre Channel adapter

A *Fibre Channel adapter* (FC adapter) is a card installed in a host system to interface with the connected devices. It is a *host I/O interface*. The FC adapter allows data to be transferred

over fiber links at very high speeds (2 Gb) and over greater distances (10 km) than SCSI. According to its characteristics and its configuration, they allow the server to participate in different connectivity topologies (point-to-point, fabric, arbitrated loop).

Fibre Channel

Fibre Channel (FC) is a technology standard that allows data to be transferred from one node to another at high speeds. This standard has been defined by a consortium of industry vendors and has been accredited by the American National Standards Institute (ANSI). The word *Fibre* in Fibre Channel takes the French spelling rather than the traditional spelling of fiber, as in fiber optics, because the interconnection between nodes are not necessarily based on fiber optics. Fibre Channel on the ESS is always based on fiber optics.

Fibre Channel (FC) is capable of carrying traffic using various protocols — IP traffic, FICON traffic, FCP (SCSI) traffic— all at the same level in the standard FC transport.

Fibre Channel host adapter

The Fibre Channel *host adapter* (HA) is the physical component of the ESS used to attach the servers' Fibre Channel I/O interfaces and SAN fabric ports. The ESS Fibre Channel host adapter connects to the server Fibre Channel I/O adapter by means of the Fibre Channel link and accepts the upper-layer commands (more than one protocol is supported by the Fibre Channel standard) from the host system. The ESS Fibre Channel host adapter, which in fact is a Fibre Channel/FICON adapter card, can be configured either for FICON or for FCP use. It has one 2 Gb port for fiber connection.

FCP

FCP is an acronym for *Fibre Channel Protocol*. When mapping SCSI to the Fibre Channel transport (FC-4 Upper Layer), then we have FCP. This is SCSI over Fibre Channel.

1.4.2 Data architecture

In this section we review the terms commonly used when discussing the different ways in which data is organized, viewed, or accessed.

CKD

Count-key-data (CKD) is the disk data architecture used by zSeries and S/390 servers. In this data organization, the *data* field stores the user data. Also because the data records can be variable in length, they all have an associated *count* field that indicates the user data record size. Then the *key* field is used to enable a hardware search on a key. However, this is not generally used for most data anymore. ECKD is a more recent version of CKD that uses an enhanced S/390 channel command set.

The commands used in the CKD architecture for managing the data and the storage devices are called channel command words (CCWs). These are equivalent to the SCSI commands.

DASD

DASD is an acronym for *Direct Access Storage Device*. This term is common in the zSeries and iSeries environments to designate a volume. This volume may be a physical disk or, more typically, a logical disk, which is usually mapped over multiple physical disks.

Fixed Block architecture (FBA or FB)

The SCSI implementation uses a *Fixed Block architecture*, that is, the data (hence the logical volumes) is mapped over fixed-size blocks or sectors. With an FB architecture, the location of any block can be calculated to retrieve that block. The concept of tracks and cylinders also

exists, because on a physical disk we have multiple blocks per track, and a cylinder is the group of tracks that exists under the disk heads at one point in time without doing a seek.

The FCP implementation (SCSI over Fibre Channel) uses this same arrangement.

Hard disk drive (HDD) — disk drive module (DDM)

The *HDD* is the primary non-volatile storage medium that is used to store data within the disk storage server. These are the round-flat rotating plates coated with magnetic substances that store the data on their surface. User data is mapped, in different ways, upon the HDDs, which can be arranged in RAID arrays (Redundant Array of Independent Disks) for data protection.

The *DDMs* are the hardware-replaceable units that hold the HDDs. Many times, both terms are used interchangeably and may also be called a *disk drive* in a short form.

Logical disk

See *logical volume* next.

Logical volume

A *logical volume* is the storage medium associated with a logical disk and typically resides on one or more HDDs. For the ESS, the logical volumes are defined at logical configuration time. For CKD servers, the logical volume size is defined by the device emulation mode and model (3390 or 3380 track format, emulated 3390 model, or custom volume). For FB hosts, the size is 100 MB to the maximum capacity of a rank. For the IBM *e*server iSeries servers, the size corresponds to the 9337 or 2105 emulated volume models. The AIX operating system views a logical volume as a logical disk or a hard disk (hdisk), an AIX term for storage space.

Logical device

The disk drives in the ESS can be configured to conform host-addressable *logical devices*, where the logical volumes are then allocated. These setups are seen by the host as its logical devices and will be pointed using an addressing scheme that depends on the attachment setup. For FICON and ESCON attachments, it will be the ESCON or FICON unit address of a 3390 or 3380 emulated device. For open systems with SCSI attachment, it will be the SCSI target, LUN assigned to a logical device. For open systems with Fibre Channel attachment, it will be the Fibre Channel adapter, LUN assigned to a logical device.

Logical unit

A *logical unit* is the Small Computer System Interface (SCSI) term for a logical device.

Logical unit number (LUN)

The *LUN* is the SCSI term for the field in an identifying message that is used to select a logical unit on a given target. The LUNs are the virtual representation of the logical devices as seen and mapped from the host system.

For the FB hosts, the ESS LUN sizes can be configured in increments of 100 MB. This increased granularity of LUN sizes enables improved storage management efficiencies.

SCSI ID

A *SCSI ID* is the unique identifier assigned to a SCSI device that is used in protocols on the SCSI interface to identify or select the device. The number of data bits on the SCSI bus determines the number of available SCSI IDs. A wide interface has 16 bits, with 16 possible IDs. A SCSI device is either an initiator or a target.

1.4.3 Server platforms

Following are the terms used to generically designate or refer to the different server platforms and operating systems.

CKD server

The term *CKD server* is used to refer, in a generic way, to the zSeries servers and the rest of the S/390 servers. These servers connect to the ESS using ESCON or FICON host attachment features. For these environments, the data in the ESS is organized according to the CKD architecture. These servers run the z/OS, OS/390, MVS/ESA, z/VM, VM/ESA, VSE/ESA, Linux, and TPF family of operating systems.

ESCON host

An *ESCON host* is a CKD server that uses ESCON channels to connect to the ESS (ESCON host attachment).

FB server

The term *FB server* is used to refer, in a generic way, to the hosts that attach via SCSI or Fibre Channel (FCP) facilities to the ESS. For these servers, the data in the ESS is organized according to the Fixed Block architecture characteristics. These servers run the Windows NT, Windows 2000, Novell NetWare, DYNIX/ptx, IBM AIX, OS/400 operating systems and the non-IBM variants of UNIX.

Fibre Channel host

A *Fibre Channel host* is a server that uses Fibre Channel I/O adapters to connect to the ESS

FICON host

A *FICON host* is a CKD server that uses FICON channels to connect to the ESS (FICON host attachment).

Intel-based servers

The term *Intel-based server* (or Intel servers) is used, in a generic way, to refer to all the different makes of servers that run on Intel processors. This is a generic term that includes servers running Windows NT and Windows 2000, as well as Novell NetWare, DYNIX/ptx, and Linux. These servers are the IBM Netfinity and IBM PC Server families of servers, the IBM NUMA-Q servers, the most recent e-business xSeries family of servers, and the various non-IBM makes of servers available on the market that run using Intel processors.

iSeries

The term *iSeries* is used to refer to the iSeries family of IBM enterprise e-business servers. The iSeries is the successor to the AS/400 family of processors. These servers run the OS/400 operating system.

Mainframe

The term *mainframe* is used to refer, in a generic way, to the zSeries family of IBM enterprise e-business processors, to the previous IBM 9672 G5 and G6 (Generation 5 and 6, respectively) processors, and also to the rest of previous S/390 processors. The zSeries servers can run under the new z/Architecture, while the 9672 and previous server models run under the System/390 (S/390) architecture.

Open systems

The term *open systems* (or *open servers*) is used, in a generic way, to refer to the systems running Windows NT, Windows 2000, Novell NetWare, DYNIX/ptx, Linux, as well as the systems running IBM AIX, IBM OS/400 and the many variants of non-IBM UNIX operating systems.

pSeries

The term *pSeries* is used to refer to the IBM @server pSeries family of IBM enterprise e-business servers. The pSeries are the successors to the RS/6000 and RS/6000 SP family of processors. These servers run the AIX operating system.

SCSI host

A *SCSI host* is an FB server that uses SCSI I/O adapters to connect to the ESS.

UNIX servers

The term *UNIX servers* is used, in a generic way, to refer to the servers that run the different variations of the UNIX operating system. This includes the IBM @server pSeries family of IBM enterprise e-business servers and the RS/6000 and the RS/6000 SP family of servers running IBM AIX. It also includes the non-IBM servers running the various versions of UNIX, for example the Sun servers running Solaris, the HP9000 Enterprise servers running HP-UX, the Compaq Alpha servers running Open VMS, or Tru64 UNIX.

xSeries

The term *xSeries* is used to refer to the IBM @server xSeries family of IBM enterprise e-business servers. The xSeries are the new servers in the IBM Netfinity and PC Server family of servers. These servers run operating systems on Intel processors.

z/Architecture

z/Architecture is the IBM 64-bit real architecture implemented in the new IBM @server zSeries family of IBM enterprise e-business servers. This architecture is an evolution from the previous IBM S/390 architecture.

z/OS

The IBM *z/OS* operating system is highly integrated with the z/Architecture microcode and hardware implementation of the IBM @server zSeries family of IBM enterprise e-business processors. It is the evolution of the OS/390 operating system.

z/VM

Building upon the solid VM/ESA base, the succeeding *z/VM* delivers enhancements for the new hardware technologies such as 64-bit addressing, FICON channels, high-speed communication adapters, and advanced storage solutions.

zSeries

The term *zSeries* is used to refer to the IBM @server zSeries family of IBM enterprise e-business servers. The zSeries servers, with their z/Architecture, are the architectural evolution of the S/390 servers (IBM 9672 G5 and G6 processors, and previous S/390 processor models). These new servers run the z/OS, z/VM, OS/390, VM/ESA, VSE/ESA, and TPF operating systems.

1.4.4 Other terms

The following terms are also commonly used throughout this book.

Array

A *disk array* is a group of disk drive modules (DDMs) that are arranged in a relationship, for example, a RAID 5 or a RAID 10 array. For the ESS, the arrays are built upon the disks of the disk eight-packs.

Back end

A *back end* consists of all the hardware components of the storage server that are connected and functionally relate, from the cluster processors and CPI down to the SSA device adapters and DDMs. See also *front end*.

Cluster

A *cluster* is a partition of a storage server that is capable of performing all functions of a storage server. When a cluster fails in a multiple-cluster storage server, any remaining clusters in the configuration can take over the processes of the cluster that fails (failover).

Controller image — logical control unit (LCU)

The term *controller image* is used in the zSeries (and S/390) environments to designate the LSS, or Logical Subsystems, that are accessed with ESCON or FICON host I/O interfaces. It is also referred to as a *logical control unit* (LCU). One or more controller images exist in a physical controller. Each image appears to be an independent controller, but all of them share the common set of hardware facilities of the physical controller. In the ESS, the CKD LSSs are viewed as logical control units by the connected S/390 operating systems. The ESS can emulate 3990-3, 3990-6, or 3990-3 TPF controller images.

Destage

Destage is the process of writing user data updates from cache to the disk arrays.

Disk eight-pack

The physical storage capacity of the ESS is materialized by means of the disk eight-packs. These are sets of eight DDMs that are installed in pairs in the ESS. Two disk eight-packs provide for two disk groups —four DDMs from each disk eight-pack. These disk groups can be configured as either RAID 5 or RAID 10 ranks.

Disk group

A *disk group* is a set of eight DDMs belonging to a pair of disk eight-packs —four DDMs from each one— where a RAID-protected rank is going to be configured.

Front end

A *front end* consists of all the hardware components of the storage server that are connected and functionally relate, from the cluster processors and CPI up to the host adapters (the host interface). See also *back end*.

Rank

This is how the ESS represents the disk arrays, whether configured as RAID 5 or RAID 10.

Staging

Staging is the process of moving data from an offline or low-priority device back to an online or higher priority device. In the ESS, this is the movement of data records from the disk arrays to the cache.



Hardware

This chapter describes the physical hardware components of the IBM TotalStorage Enterprise Storage Server Model 800 (ESS).

2.1 IBM TotalStorage Enterprise Storage Server Model 800

The IBM TotalStorage Enterprise Storage Server is an IBM Seascape architecture disk system implementation for the storage and management of the enterprise data. It is a solution that provides the outboard intelligence required by Storage Area Network (SAN) solutions, off-loading key functions from host servers and freeing up valuable processing power for applications.

2.1.1 Hardware characteristics

The IBM TotalStorage Enterprise Storage Server Model 800 (ESS) is a third generation high-performance, high-availability, and high-capacity disk storage subsystem. Figure 2-1 summarizes its main characteristics.

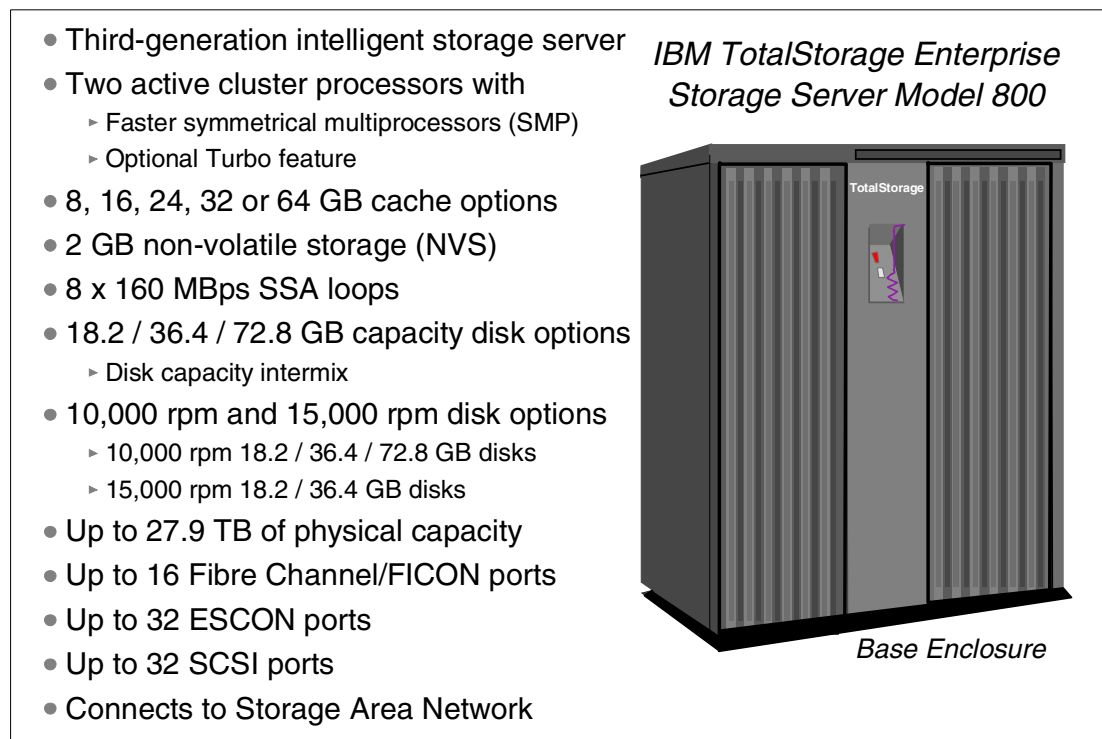


Figure 2-1 IBM TotalStorage Enterprise Storage Server Model 800

The IBM TotalStorage Enterprise Storage Server Model 800 integrates a new generation of hardware from top to bottom, including new SMP processors (with an optional Turbo feature), 64 GB of cache, 2 GB non-volatile storage (NVS), increased internal bandwidth, and 2 Gb Fibre Channel / FICON host adapters. This new hardware, when combined with the new RAID 10 support and 15,000 rpm drives, enables the ESS Model 800 to deliver unprecedented levels of performance and throughput.

The ESS Model 800 supports up to 27.9 TB of physical capacity that can be configured as RAID 5 or RAID 10, or a combination of both. RAID 5 remains the price/performance leader, with excellent performance for most user applications. RAID 10 can offer better performance for selected applications. Price, performance, and capacity can further be optimized to meet specific application and business requirements through the intermix of 18.2, 36.4, and 72.8 GB drives operating at 10,000 and 15,000 rpm.

Yet with all this, the fundamental design of the ESS remains unchanged. With stable and proven technology, the ESS Model 800 has comprehensive 24 x 7 availability, with a design that minimizes single points of failure by providing component redundancy. The ESS Model 800 also maintains the advanced functions that deliver business continuance solutions: FlashCopy, PPRC, PPRC Extended Distance, and XRC. It also maintains the zSeries superior performance features: FICON, Parallel Access Volumes (PAV), Multiple Allegiance, and I/O Priority Queuing.

The ESS also connects to Storage Area Networks (SANs), and by means of a gateway, it can also participate in Network Attached Storage (NAS) environments. As a comprehensive SAN storage solution, the ESS provides the management flexibility to meet the fast-paced changing requirements of today's business.

The ESS Model 800 hardware characteristics summarized in Figure 2-1 on page 20 are explained in detail in the following sections of this chapter.

Base enclosure

The ESS Model 800 consists of a base enclosure and also can attach an Expansion Enclosure for the larger-capacity configurations.

The base enclosure provides the rack and packaging that contain the host adapters, the cluster processors, cache and NVS, and the device adapters. It has two three-phase power supplies and supports 128 disk drives in two cages (cages are described later in 2.5, "ESS cages" on page 24).

Expansion Enclosure

The Expansion Enclosure provides the rack and packaging for an additional 256 disk drives. It also contains its own set of power supplies.

2.2 ESS Expansion Enclosure

<p>Base enclosure</p> <ul style="list-style-type: none">Two three-phase power suppliesUp to 128 disk drives in two cagesFeature #2110 for Expansion Enclosure attachmentAlso contains host adapters, cluster processors, cache and NVS, and the SSA device adapters <p>Expansion Enclosure</p> <ul style="list-style-type: none">Two three-phase power suppliesUp to 256 disk drives in four cages for additional capacity
--

Figure 2-2 ESS Model 800 base and Expansion Enclosures

The Expansion Enclosure rack attaches to the ESS Model 800 (feature 2110 of the 2105-800) and uses two three-phase power supplies (refer to Figure 2-2). Two power-line cords are

required for the Expansion Enclosure, and they have the same requirements as the ESS base rack line cords. The expansion rack is the same size as the base enclosure.

Up to four ESS cages (ESS cages are described in 2.5, “ESS cages” on page 24) can be installed in the Expansion Enclosure rack, and this gives a maximum capacity of 256 disk drives for the expansion rack. This brings the total disk drive capacity of the ESS, when configured with the Expansion Enclosure, to 384.

2.3 Photograph of the ESS Model 800

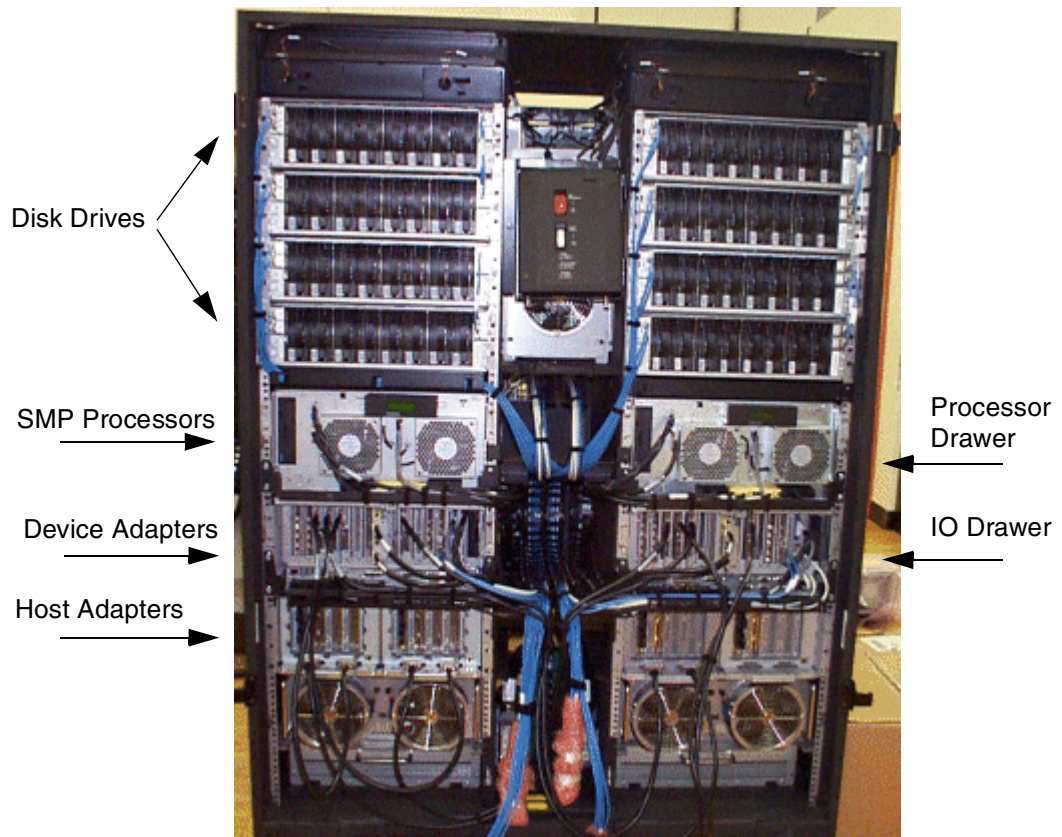


Figure 2-3 Photograph of the ESS Model 800 base enclosure (front covers removed)

Figure 2-3 shows a photograph of an ESS Model 800 with the front covers removed. At the top of the frame are the disk drives, and immediately under them are the processor drawers that hold the cluster SMP processors. Just below the processor drawers are the I/O drawers that hold the SSA device adapters that connect to the SSA loops (blue cables in Figure 2-3). Just below the I/O drawers are the host adapter bays that hold the host adapters. At the bottom of the frame are the AC power supplies and batteries.

The ESS in this photo has two cages holding the disk drives (DDMs). If the capacity of this ESS was 64 or fewer disk drives, then the top right side of this ESS would have an empty cage in it. The photo clearly shows the two clusters, one on each side of the frame.

Between the two cages of DDMs is an operator panel that includes an emergency power switch, local power switch, power indicator lights, and message/error indicator lights.

The height of the ESS is 70.7 inches (1.796 meters) without its top cover, the width is 54.4 inches (1.383 meters), and the depth is 35 inches (0.91 meters). Using a raised floor environment is recommended to provide increased air circulation and easier cable access from the front.

For larger configurations, the ESS base enclosure attaches to an Expansion Enclosure rack that is the same size as the base ESS, and stands next to the ESS base frame.

2.4 ESS major components

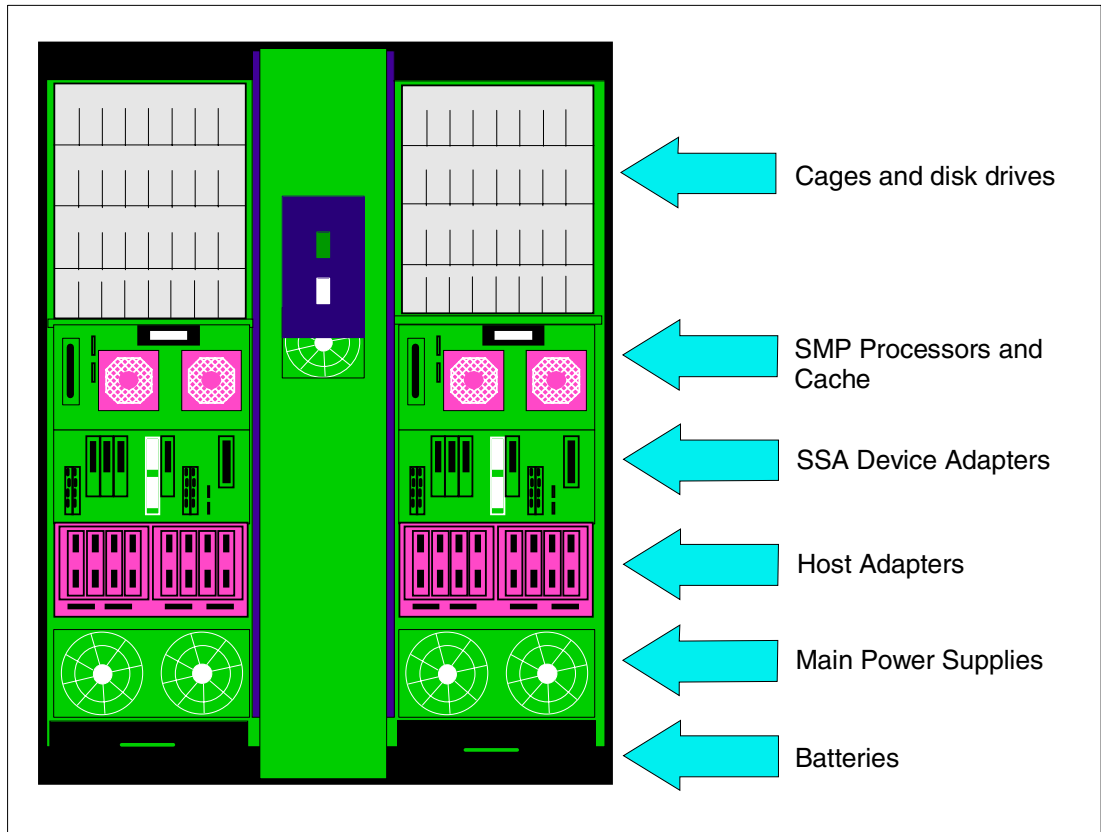


Figure 2-4 ESS Model 800 major components

The diagram in Figure 2-4 shows an IBM TotalStorage Enterprise Storage Server Model 800 and its major components. As you can see, the ESS base rack consists of two clusters, each with its own power supplies, batteries, SSA device adapters, processors, cache and NVS, CD drive, hard disk, floppy disk and network connections. Both clusters have access to any host adapter card, even though they are physically spread across the clusters.

At the top of each cluster is an ESS cage. Each cage provides slots for up to 64 disk drives, 32 in front and 32 at the back.

In the following sections, we will look at the ESS major components in detail.

2.5 ESS cages

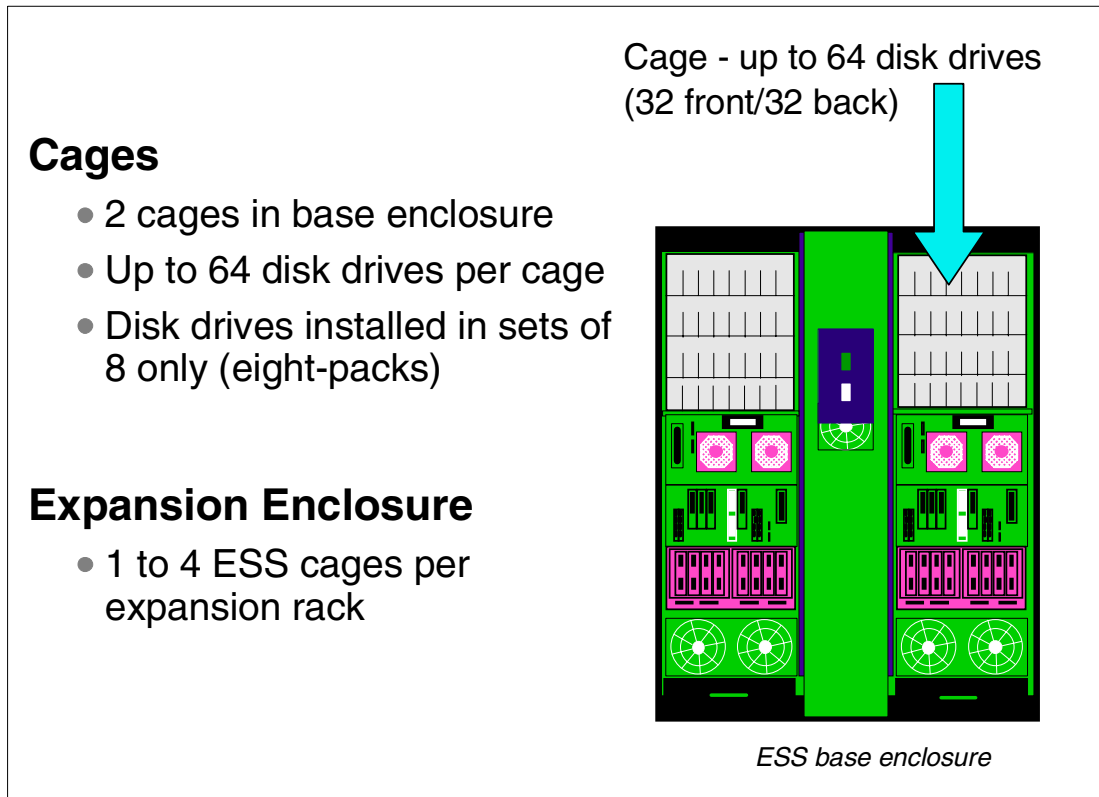


Figure 2-5 ESS cages

The high-capacity ESS cages are supplied as part of a disk mounting kit and provide some common functions to the large set of disks they accommodate. A total of six cages can be installed in an ESS (configured with the Expansion Enclosure feature), which provides accommodation for up to 384 disk drive modules (DDMs). The ESS base enclosure always comes with two cages (disk eight-pack mounting kit), as Figure 2-5 illustrates.

Disks are installed in the cages in groups of eight. These are called disk eight-packs. The cages (disk eight-pack mounting kits consisting of sheet metal cages, power supplies, fans, cables, etc.) are required for every eight disk eight-packs. The cage provides the power connections for each eight-pack that comes packaged into a case and slides into a slot in the cage.

The minimum number of eight-packs is 4 (32 DDMs) after which they can be added in sets of two. The front of the left cage is filled up first from bottom to top, then the back, again filled from bottom to top. Once the left cage is full, then the right cage is filled up in the same order. Empty slots have a protective flap that controls the airflow over the disks.

Expansion Enclosure

The Expansion Enclosure rack can hold from one to four cages. Similar to the cages in the base enclosure, each cage in the Expansion Enclosure rack can contain up to 64 disk drives in sets of eight. This allows an additional capacity of 256 disk drives in four cages when the ESS Expansion Enclosure is attached.

The expansion rack is populated with disk eight-packs in a similar manner to the base frame, that is bottom to top, front of cage, then the back, left cage then right. See also 4.6, “Expansion Enclosure” on page 99.

2.6 ESS disks

- **Eight-packs**
 - Set of eight similar capacity/rpm disk drives packed together
 - Installed in the ESS cages
 - Initial minimum configuration is four eight-packs
 - Upgrades are available in increments of two eight-packs
 - Maximum of 48 eight-packs per ESS with expansion
- **Disk drives**
 - 18.2 GB 15,000 rpm or 10,000 rpm
 - 36.4 GB 15,000 rpm or 10,000 rpm
 - 72.8 GB 10,000 rpm
- **Eight-pack conversions**
 - Capacity and/or RPMs
- **Step Ahead capacity on demand option**

Figure 2-6 ESS Model 800 disks

With a number of disk drive sizes and speeds available, including intermix support, the ESS provides a great number of capacity configuration options. Figure 2-6 illustrates the options available when configuring the disk drives.

2.6.1 ESS disk capacity

The maximum number of disk drives supported within the IBM TotalStorage Enterprise Storage Server Model 800 is 384 — with 128 disk drives in the base enclosure and 256 disk drives in the expansion rack. When configured with 72.8 GB disk drives, this gives a total physical disk capacity of approximately 27.9 TB (see Table 2-1 on page 27 for more details).

The minimum available configuration of the ESS Model 800 is 582 GB. This capacity can be configured with 32 disk drives of 18.2 GB contained in four eight-packs, using one ESS cage. All incremental upgrades are ordered and installed in pairs of eight-packs; thus the minimum capacity increment is a pair of similar eight-packs of either 18.2 GB, 36.4 GB, or 72.8 GB capacity.

2.6.2 Disk features

The ESS is designed to deliver substantial protection against data corruption, not just relying on the RAID implementation alone. The disk drives installed in the ESS are the latest state-of-the-art magneto resistive head technology disk drives that support advanced disk functions such as disk error correction codes (ECC), Metadata checks, disk scrubbing and predictive failure analysis. These functions are described in more detail in 3.2.2, “Data protection” on page 52.

2.6.3 Disk eight-packs

The ESS eight-pack is the basic unit of capacity within the ESS base and expansion racks (refer to Figure 2-6 on page 25). As mentioned before, these eight-packs are ordered and installed in pairs. Each eight-pack can be configured as a RAID 5 rank (6+P+S or 7+P) or as a RAID 10 rank (3+3+2S or 4+4).

The IBM TotalStorage ESS Specialist will configure the eight-packs on a loop with spare DDMs as required. Configurations that include drive size intermixing may result in the creation of additional DDM spares on a loop as compared to non-intermixed configurations (see 3.7, “Sparing” on page 58 for details).

Currently there is the choice of three different new-generation disk drive capacities for use within an eight-pack:

- ▶ 18.2 GB/ 15,000 rpm disks
- ▶ 36.4 GB/ 15,000 rpm disks
- ▶ 72.8 GB/ 10,000 rpm disks

Also available is the option to install eight-packs with:

- ▶ 18.2 GB/ 10,000 rpm disks or
- ▶ 36.4 GB/ 10,000 rpm disks

The eight disk drives assembled in each eight-pack are all of the same capacity. Each disk drive uses the 40 MBps SSA interface on each of the four connections to the loop.

It is possible to mix eight-packs of different capacity disks and speeds (rpm) within an ESS, as described in the following sections.

2.6.4 Disk eight-pack capacity

Eight-packs in the ESS can be of different capacities. Additionally, these groups of eight disk drives once installed in the ESS can be configured either as RAID 5 or RAID 10 arrays.

RAID 5 as implemented on the ESS offers the most cost-effective performance/capacity trade-off options for the ESS internal disk configurations, because it optimizes the disk storage capacity utilization. RAID 10 offers higher potential performance for selected applications, but requires considerably more disk space. Because of its disk-to-disk mirroring, RAID 10 eight-packs can have (approximately) between 40% and 65% of the *effective capacity* of RAID 5 eight-packs, depending upon whether or not there are spare disks in the arrays.

Table 2-1 on page 27 should be used as a guide for determining the capacity of a given eight-pack, after consulting 2.6.5, “Disk intermixing” on page 27 to understand the limitations. This table shows the capacities of the disk eight-packs when configured as RAID ranks. These capacities are the *effective capacities* available for user data.

Table 2-1 Disk eight-pack effective capacity chart (gigabytes)

Disk Size	Physical Capacity (raw capacity)	Effective usable capacity (2)			
		RAID 10		RAID 5 (3)	
		3 + 3 + 2S Array (4)	4 + 4 Array (5)	6+P+S Array (6)	7 + P Array (7)
18.2	145.6	52.50	70.00	105.20	122.74
36.4	291.2	105.12	140.16	210.45	245.53
72.8	582.4	210.39	280.52	420.92	491.08

Notes:

1. A gigabyte (GB) equals one billion (10⁹) bytes when referring to disk capacity.
2. *Effective capacity* represents the approximate portion of the disk eight-pack that is usable for customer data. All available capacity may not be fully utilized due to overheads on logical devices and/or a loss of capacity due to a configured set of logical devices.
3. In RAID 5 configurations, the parity information utilizes the capacity of one disk, but is actually distributed across all the disks within a given disk eight-pack.
4. Array consists of three data drives mirrored to three copy drives. The remaining two drives are used as spares.
5. Array consists of four data drives mirrored to four copy drives.
6. Array consists of six data drives and one parity drive. The remaining drive is used as a spare.
7. Array consists of seven data drives and one parity drive.

Physical capacity

The physical capacity (or raw capacity) of the ESS is calculated by multiplying each disk eight-pack installed in the ESS by its respective physical capacity value (see Table 2-1) and then summing the values.

Effective capacity

The effective capacity of the ESS is the capacity available for user data. The combination and sequence in which eight-packs are purchased, installed, and logically configured when using the ESS Specialist will determine the effective capacity of the ESS. This includes considering spare DDM allocation and other overheads (refer to Chapter 4, "Configuration" on page 93 for these considerations).

2.6.5 Disk intermixing

In the SSA loops of the ESS, it is possible to intermix:

1. 18.2 GB, 36.4 GB, and 72.8 GB disk eight-packs
2. 10,000 rpm and 15,000 rpm disk eight-packs
3. RAID 5 and RAID 10

Limitations apply as explained later in this section. For an ESS that includes an intermix of drive sizes, the disk eight-packs will be installed in sequence from highest capacity to lowest capacity (see also 4.5, "Base enclosure" on page 98 and 4.6, "Expansion Enclosure" on page 99 for a description of the physical installation sequence within the ESS enclosures).

Disk capacity intermix considerations

Disk eight-packs with different capacity drives can be installed within the same ESS.

In an SSA loop, a *spare pool* consisting of two disk drives is created for each drive size on the SSA loop. The spares are reserved from either two 6+P arrays (RAID 5) or one 3+3 array (RAID 10). Refer to 4.22, “RAID 5 and RAID 10 rank intermixing” on page 120 for more information.

Disk speed (RPM) intermix considerations

An ESS can have eight-packs that differ in their drive speed, but not if their drive capacity is similar. For example 36.4 GB, 15,000 rpm can be intermixed with 72.8 GB, 10,000 rpm eight-packs in the same ESS.

The ESS supports the intermix of 10,000 rpm and 15,000 rpm disk eight-packs within the same ESS subject to the following limitation:

Within an ESS, for a given capacity, the 15,000 rpm disk eight-pack cannot be intermixed with 10,000 rpm disk eight-pack of the same capacity. For example 18.2 GB, 15,000 rpm cannot be intermixed with 18.2 GB, 10,000 rpm within the same ESS.

Statement of general direction

IBM plans to support the intermix of 15,000 rpm drives with lower rpm drives of the same capacity within an ESS Model 800.

All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Availability, prices, ordering information, and terms and conditions will be provided when the product is announced for general availability.

2.6.6 Disk conversions

There is the ability to exchange an installed eight-pack with an eight-pack of greater capacity, or higher rpm, or both. As well as enabling you to best exploit the intermix function, the capacity conversions are particularly useful for increasing storage capacity at sites with floor-space constraints that prohibit the addition of Expansion Enclosures.

The eight-pack conversions must be ordered in pairs and are subject to the disk intermix limitations discussed previously in 2.6.5, “Disk intermixing” on page 27.

2.6.7 Step Ahead option

The Step Ahead capacity-on-demand program enables the installation of capacity ahead of the customer requiring it (the capacity is purchased at a later time). Step Ahead provides two additional eight-packs of the same capacity, pre-installed in the ESS. The same disk intermix rules, as explained previously in “Disk intermixing” on page 27, apply with the Step Ahead eight-packs.

The ESS cache uses ECC (error checking and correcting) memory technology to enhance reliability and error correction of the cache. ECC technology can detect single and double bit errors and correct all single bit errors. Memory scrubbing, a built-in hardware function, is also performed and is a continuous background read of data from memory to check for correctable errors. Correctable errors are corrected and rewritten to cache.

To protect against loss of data on a write operation, the ESS stores two copies of written data, one in cache and the other in NVS.

2.7.3 Non-volatile storage (NVS)

NVS is used to store a second copy of write data to ensure data integrity, should there be a power failure or a cluster failure and the cache copy is lost. The NVS of cluster 1 is located in cluster 2 and the NVS of cluster 2 is located in cluster 1. In this way, in the event of a cluster failure, the write data for the failed cluster will be in the NVS of the surviving cluster. This write data is then destaged at high priority to the disk arrays. At the same time, the surviving cluster will start to use its own NVS for write data, ensuring that two copies of write data are still maintained. This ensures that no data is lost even in the event of a component failure.

The ESS Model 800 has a 2 GB NVS. Each cluster has 1 GB of NVS, made up of four cards. Each pair of NVS cards has its own battery-powered charger system that protects data even if power is lost on the entire ESS for up to 72 hours.

A more detailed description of the NVS use is described in 3.28, “NVS and write operations” on page 88.

2.8 Device adapters

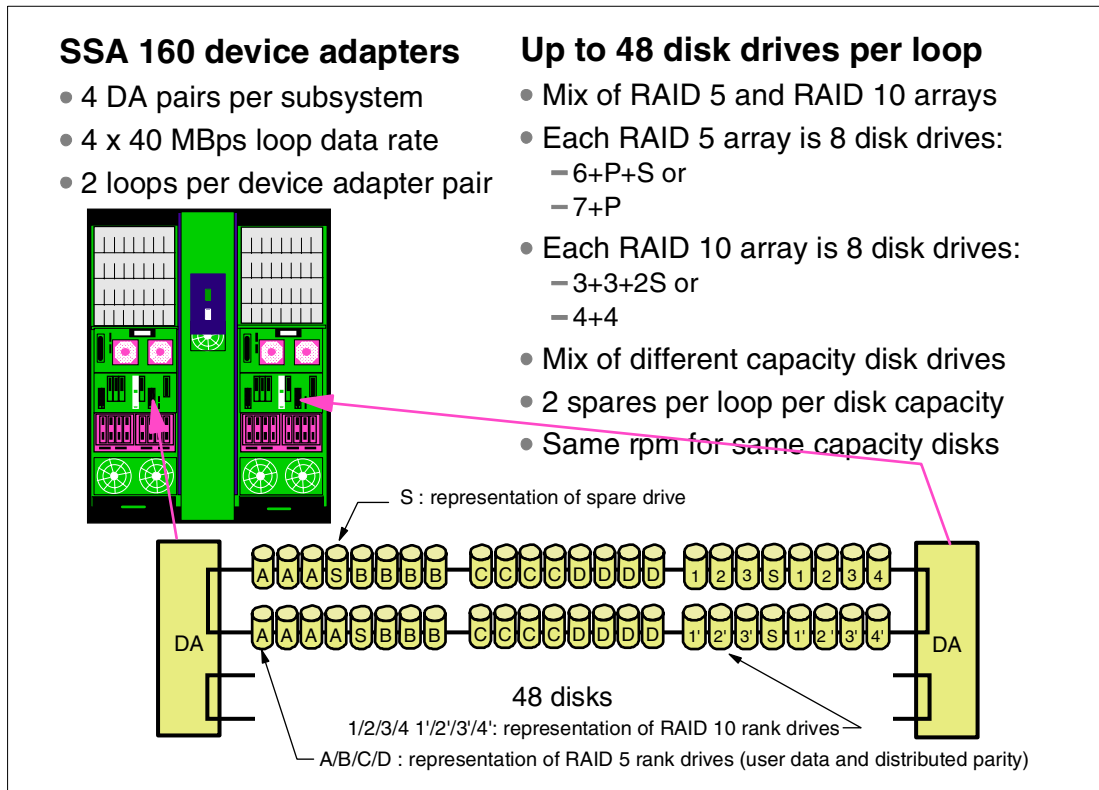


Figure 2-8 ESS device adapters

Device adapters (DA) provide the connection between the clusters and the disk drives (refer to Figure 2-8). The IBM TotalStorage Enterprise Storage Server Model 800 implements faster SSA (Serial Storage Architecture) device adapters than its predecessor models.

2.8.1 SSA 160 device adapters

The IBM TotalStorage Enterprise Storage Server Model 800 uses the latest SSA160 technology in its device adapters (DA). With SSA 160, each of the four links operates at 40 MBps, giving a total nominal bandwidth of 160 MBps for each of the two connections to the loop. This amounts to a total of 320 MBps across each loop (see 2.9.3, “Spatial reuse” on page 34). Also, each device adapter card supports two independent SSA loops, giving a total bandwidth of 320 MBps per adapter card. There are eight adapter cards, giving a total nominal bandwidth capability of 2560 MBps. See 2.9, “SSA loops” on page 33 for more on the SSA characteristics.

One adapter from each pair of adapters is installed in each cluster as shown in Figure 2-8. The SSA loops are between adapter pairs, which means that all the disks can be accessed by both clusters. During the configuration process, each RAID array is configured by the IBM TotalStorage ESS Specialist to be normally accessed by only one of the clusters. Should a cluster failure occur, the remaining cluster can take over all the disk drives on the loop.

RAID 5 and RAID 10 are managed by the SSA device adapters. RAID 5 and RAID 10 are explained in detail in 3.10, “RAID Data protection” on page 63.

2.8.2 Disk drives per loop

Each loop supports up to 48 disk drives, and each adapter pair supports up to 96 disk drives. There are four adapter pairs supporting up to 384 disk drives in total.

Figure 2-8 on page 31 shows a logical representation of a single loop with 48 disk drives (RAID ranks are actually split across two eight-packs for optimum performance). In the figure you can see there are six RAID arrays: four RAID 5 designated A to D, and two RAID 10 (one 3+3+2 spare and one 4+4).

2.8.3 Hot spare disks

The ESS requires that a loop have a minimum of two spare disks to enable sparing to occur. The sparing function of the ESS is automatically initiated whenever a DDM failure is detected on a loop and enables regeneration of data from the failed DDM onto a hot spare DDM.

A hot DDM *spare pool* consisting of two drives, created with two 6+P+S arrays (RAID 5) or one 3+3+2S array (RAID 10), is created for each drive size on an SSA loop. Therefore, if only one drive size is installed on a loop, only two spares are required. The hot sparing function is managed at the SSA loop level. SSA will spare to a larger capacity DDM on the loop in the very uncommon situation that no spares are available on the loop for a given capacity.

Two additional spares will be created whenever a new drive size is added to an SSA loop. Thus, a unique *spare pool* is established for each drive capacity on the SSA loop.

Figure 2-8 on page 31 shows arrays A and B both have spare disks that are used across the loop in case of a disk failure. When the failed disk is replaced, it becomes the new spare. Over time the disks (data and spare) in the RAID ranks on the loop become mixed. So it is not possible to remove an eight-pack without affecting the rest of the data in the loop.

If a loop has only one disk capacity size installed, then both RAID formats will share the spares. In Figure 2-8 on page 31, there are two RAID 10 arrays consisting of a pair of eight-packs of a different drive size from the RAID 5 eight-packs. Therefore, two additional hot spare DDMs will be reserved by the 3+3+2S arrays.

2.9 SSA loops

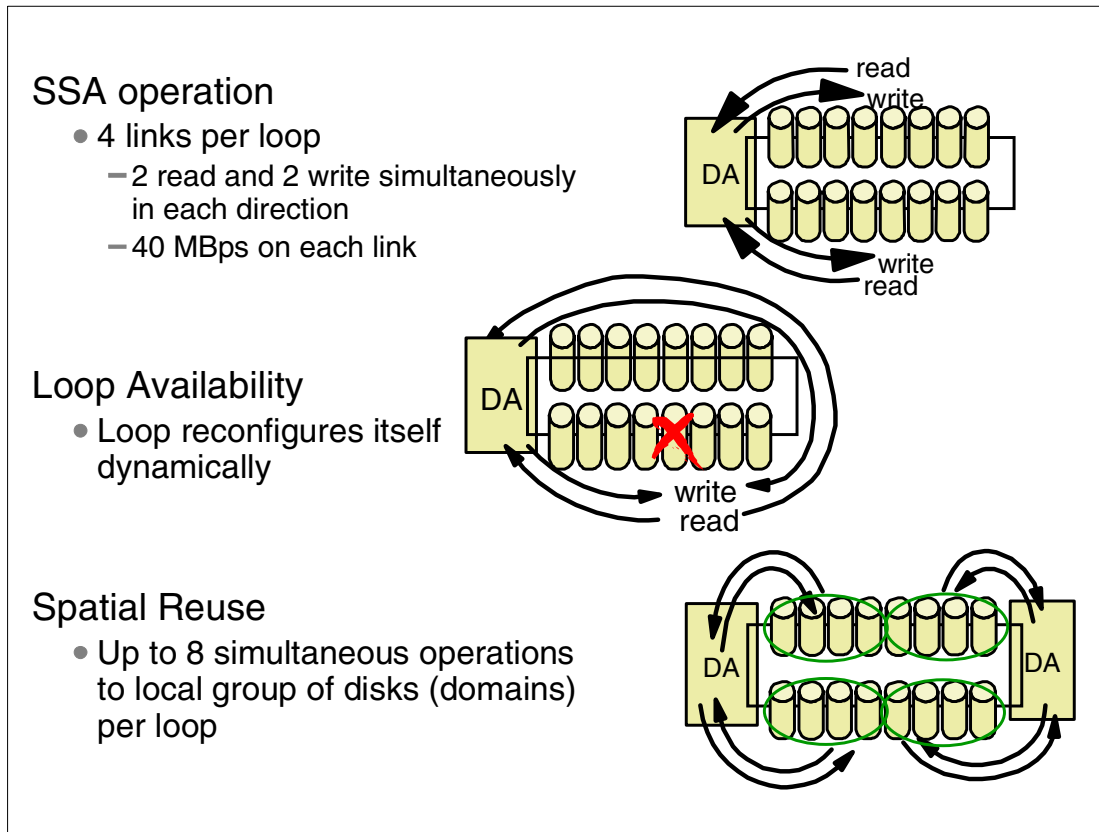


Figure 2-9 SSA loops

The IBM TotalStorage Enterprise Storage Server Model 800 uses the latest SSA160 technology in its device adapters (DA).

2.9.1 SSA operation

Serial Storage Architecture (SSA) is a high performance serial-connection technology for disk drives. SSA is a full-duplex loop-based architecture, with two physical read paths and two physical write paths to every disk drive attached to the loop (refer to Figure 2-9). Data is sent from the adapter card to the first disk drive on the loop and then passed around the loop by the disk drives until it arrives at the target disk. Unlike bus-based designs, which reserve the whole bus for data transfer, SSA only uses the part of the loop between adjacent disk drives for data transfer. This means that many simultaneous data transfers can take place on an SSA loop, and it is one of the main reasons that SSA performs so much better than SCSI. This simultaneous transfer capability is known as *spatial reuse*.

Each read or write path on the loop operates at 40 MBps, providing a total loop bandwidth of 160 MBps.

2.9.2 Loop availability

The loop is a self-configuring, self-repairing design that allows genuine hot-plugging. If the loop breaks for any reason, then the adapter card will automatically reconfigure the loop into two single loops. In the ESS, the most likely scenario for a broken loop is if the actual disk

drive interface electronics should fail. If this should happen, the adapter card will dynamically reconfigure the loop into two single loops, effectively isolating the failed disk drive.

If the disk drive were part of a RAID 5 array, the adapter card would automatically regenerate the missing disk drive (using the remaining data and parity disk drives) to the spare. If it were part of a RAID 10 array, the adapter card would regenerate the data from the remaining mirror image drive.

Once the failed disk drive has been replaced, the loop will automatically be re-configured into full-duplex operation, and the replaced disk drive will become a new spare.

2.9.3 Spatial reuse

Spatial reuse allows domains to be set up on the loop. A domain means that one or more groups of disk drives “belong” to one of the two adapter cards, as is the case during normal operation. The benefit of this is that each adapter card can talk to its domains (or disk groups) using only part of the loop. The use of domains allows each adapter card to operate at maximum capability, because it is not limited by I/O operations from the other adapter. Theoretically, each adapter card could drive its domains at 160 MBps, giving 320 MBps throughput on a single loop! The benefit of domains may diminish slightly over time, due to disk drive failures causing the groups to become intermixed, but the main benefits of spatial reuse will still apply.

The spatial reuse feature sets SSA apart from other serial link technologies and from SCSI. With spatial reuse, SSA loops can achieve effective bandwidths well in excess of any individual link bandwidth. This is particularly true for RAID arrays where data transfer activity can also occur within the disk arrays and device adapters level, independently of transfers to the clusters processors and cache.

If a cluster or device adapter should fail, the remaining cluster device adapter will own all the domains on the loop, thus allowing full data access to continue.

2.10 Host adapters

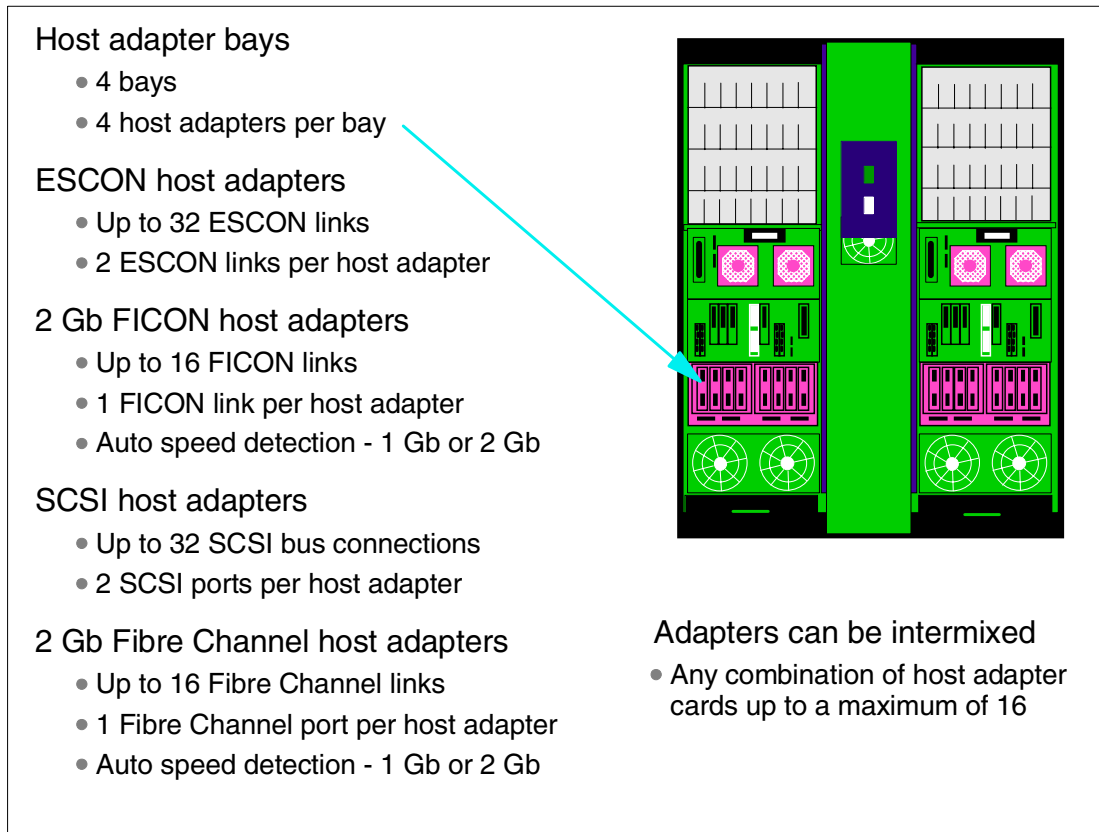


Figure 2-10 ESS Model 800 host adapters

The IBM TotalStorage Enterprise Storage Server has four host adapter (HA) bays, two in each cluster. Each bay supports up to four host adapter cards. Each of these host adapter cards can be for FICON, ESCON, SCSI, or Fibre Channel server connection. Figure 2-10 lists the main characteristics of the ESS host adapters.

Each host adapter can communicate with either cluster. To install a new host adapter card, the bay must be powered off. For the highest path availability, it is important to spread the host connections across all the adapter bays. For example, if you have four ESCON links to a host, each connected to a different bay, then the loss of a bay for upgrade would only impact one out of four of the connections to the server. The same would be valid for a host with FICON connections to the ESS.

Similar considerations apply for servers connecting to the ESS by means of SCSI or Fibre Channel links. For open system servers the Subsystem Device Driver (SDD) program that comes standard with the ESS, can be installed on the connecting host servers to provide multiple paths or connections to handle errors (path failover) and balance the I/O load to the ESS. See 5.8, “Subsystem Device Driver” on page 157.

The ESS connects to a large number of different servers, operating systems, host adapters, and SAN fabrics. A complete and current list is available at the following Web site:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

2.11 ESCON host adapters

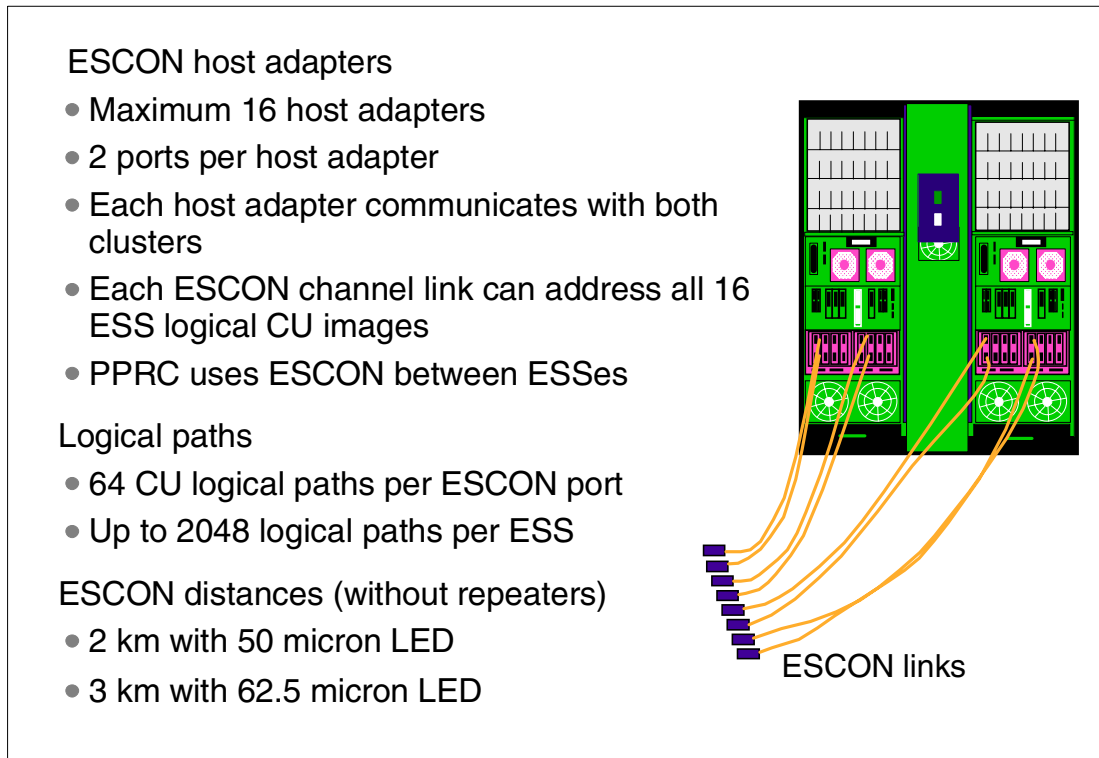


Figure 2-11 ESCON host adapters

The ESS can connect up to 32 ESCON channels (see Figure 2-11), two per ESCON host adapter. Each ESCON host adapter is connected to both clusters. The ESS emulates up to 16 of the 3990 logical control units (LCUs). Half of the LCUs (even numbered) are in cluster 1, and the other half (odd-numbered) are in cluster 2. Because the ESCON host adapters are connected to both clusters, each adapter can address all 16 LCUs. More details on the CKD logical structure are presented in 3.16, “ESS Implementation - CKD” on page 75.

Logical paths

An ESCON link consists of two fibers, one for each direction, connected at each end by an ESCON connector to an ESCON port. Each ESCON adapter card supports two ESCON ports or links, and each link supports 64 logical paths. With the maximum of 32 ESCON ports, the maximum number of logical paths per ESS is 2048.

ESCON distances

For connections without repeaters, the ESCON distances are 2 km with 50 micron multimode fiber, and 3 km with 62.5 micron multimode fiber. The ESS supports all models of the IBM 9032 ESCON directors that can be used to extend the cabling distances.

PPRC and ESCON

The PPRC remote copy function —control unit to control unit— uses ESCON connections between the two participating ESSs. For synchronous PPRC implementations, you can extend the distance at which you can operate the ESS up to 103 km. This distance requires at least two pairs of Dense Wavelength Division Multiplexers (DWDM) that can transport multiple protocols over the same fiber optic link. Each pair can be separated by up to 50 km (31 miles) of fiber. The link can be done using dark fiber (that is, leased from a common

carrier such as a telephone company or cable TV operator) as long as it meets the 2029 attachment criteria.

Even greater distances can be achieved when using nonsynchronous PPRC Extended Distance (PPRC-XD) and channel extender technology as provided by the Inrange 9801 Storage Networking System or CNT UltraNet Storage Director. The actual distance achieved is typically limited only by the capabilities of your network and the channel extension technologies. More information is found in 7.8, “PPRC connectivity” on page 216.

ESCON supported servers

ESCON is used for attaching the ESS to the IBM S/390 and zSeries servers. The most current list of supported servers is listed at the Web site:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

This document should be consulted regularly, because it has the most u-to-date information on server attachment support.

2.12 SCSI host adapters

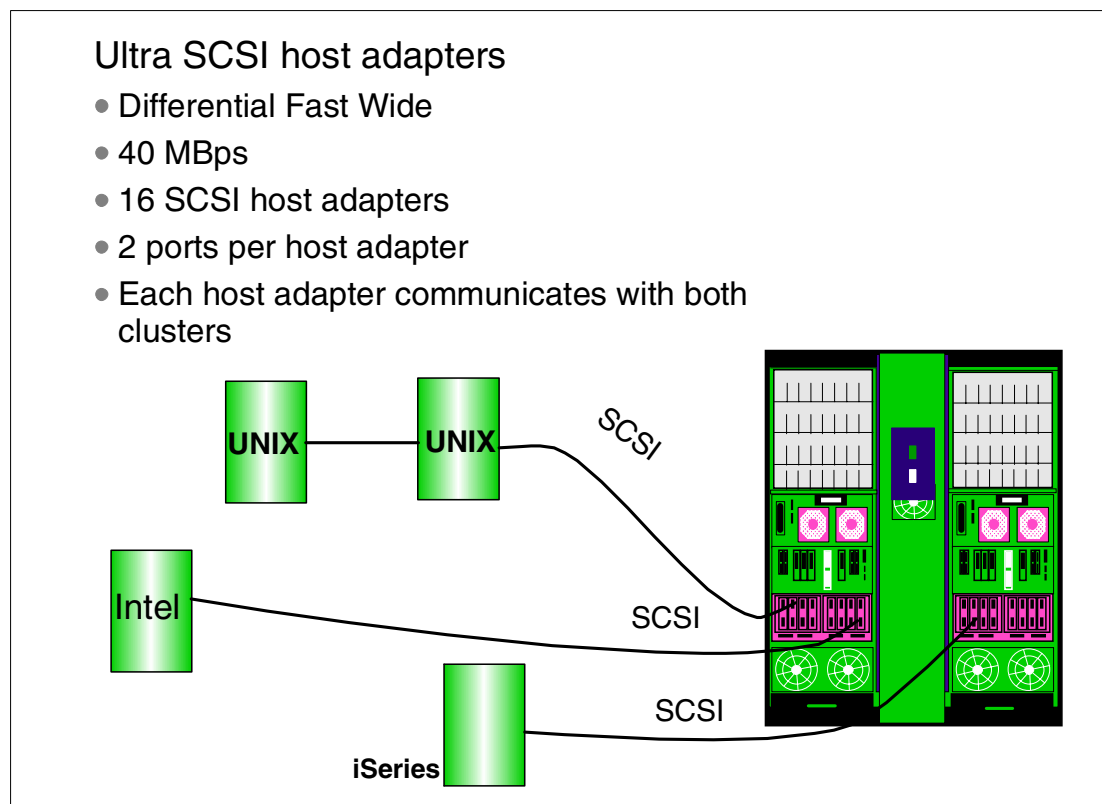


Figure 2-12 SCSI host adapters

The IBM TotalStorage Enterprise Storage Server provides Ultra SCSI interface with SCSI-3 protocol and command set for attachment to open systems (refer to Figure 2-12). This interface also supports SCSI-2.

Each SCSI host adapter supports two SCSI port interfaces. These interfaces are Wide Differential and use the VHDCI (Very High Density Connection Interface). These SCSI cables can be ordered from IBM.

SCSI supported servers

The current list of host systems supported by the ESS SCSI interface is listed at the Web site:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

This document should be consulted regularly, because it has the most up-to-date information on server attachment support. Servers supported include:

- ▶ IBM RS/6000, IBM RS/6000 SP, and the pSeries family of IBM @servers
- ▶ IBM AS/400 and the iSeries family of the IBM @servers
- ▶ Compaq Alpha servers
- ▶ HP9000 Enterprise servers
- ▶ SUN servers with Solaris
- ▶ Intel-based servers: IBM and non-IBM supported servers

SCSI targets and LUNs

The ESS SCSI interface supports 16 target SCSI IDs (the host requires one initiator ID for itself, so this leaves 15 targets for the ESS definitions) with up to 64 logical unit numbers (LUNs) per target (the SCSI-3 standard). The number of LUNs actually supported by the host systems varies from 8 to 32, and it is a characteristic of the server. Check with your host server supplier on the number supported by any specific level of driver or machine.

See 3.15, “ESS Implementation - Fixed block” on page 73 for more details on SCSI attachment characteristics.

2.13 Fibre Channel host adapters

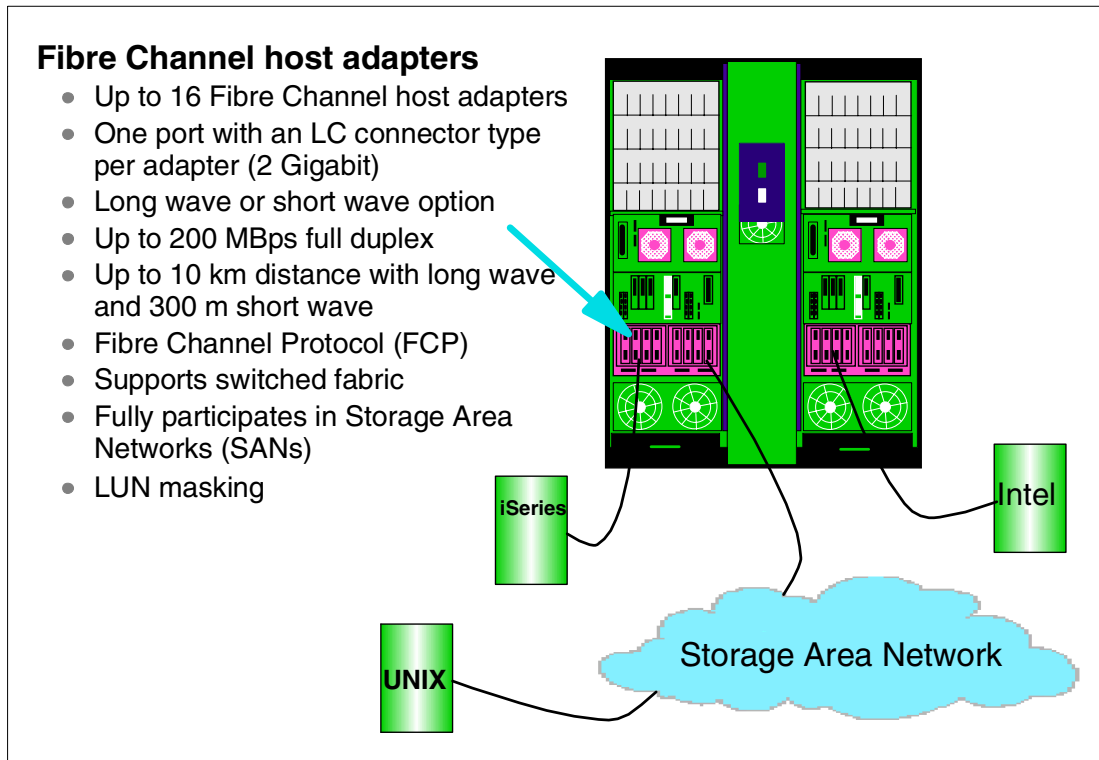


Figure 2-13 Fibre Channel host adapters

Fibre Channel is a technology standard that allows data to be transferred from one node to another at high speeds (up to 200 MBps) and greater distances (up to 10 km).

It is the very rich connectivity options of the Fibre Channel technology that has resulted in the Storage Area Network (SAN) implementations. The limitations seen on SCSI in terms of distance, performance, addressability, and connectivity are overcome with Fibre Channel and SAN.

The ESS with its Fibre Channel host adapters provides FCP (Fibre Channel Protocol, which is SCSI traffic on a serial fiber implementation) interface, for attachment to open systems that use Fibre Channel adapters for their connectivity (refer to Figure 2-13).

The ESS supports up to 16 host adapters, which allows for a maximum of 16 Fibre Channel ports per ESS. Each Fibre Channel host adapter provides one port with an LC connector type. An LC to SC “female” 2-meter cable can be ordered with the ESS to enable connection of the adapter port to an existing cable infrastructure (see the cable option features in Table A-4 on page 266).

As SANs migrate to 2 Gb technology, your storage should be able to exploit this bandwidth. The ESS Fibre channel adapters operate at up to 2 Gb. The adapter auto-negotiates to either 2 Gb or 1 GB link speed, and will operate at 1 Gb unless both ends of the link support 2 Gb operation.

Each Fibre Channel port supports a maximum of 128 host login IDs.

There are two types of host adapter cards you can select: *long wave* (feature 3024) and *short wave* (feature 3025). With long-wave laser, you can connect nodes at distances of up to

10 km (non-repeated). With short-wave laser, you can connect at distances of up to 300m. See 2.15, "Fibre distances" on page 42 for more details. The distances can be extended if using a SAN fabric.

Note: The Fibre Channel/FICON host adapter is an adapter card that supports both FICON or SCSI-FCP, but not simultaneously; the protocol to be used is configurable on an adapter-by-adapter basis.

When equipped with the Fibre Channel/FICON host adapters, configured for a Fibre Channel interface, the ESS can participate in all three topology implementations of Fibre Channel:

- ▶ Point-to-point
- ▶ Switched fabric
- ▶ Arbitrated loop

For detailed information on Fibre Channel implementations using the ESS, refer to *Implementing Fibre Channel Attachment on the ESS*, SG24-6113.

Fibre Channel supported servers

The current list of servers supported by the Fibre Channel attachment is listed at the Web site:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

This document should be consulted regularly, because it has the most up-to-date information on server attachment support. Some of the servers supported include:

- ▶ IBM RS/6000, IBM RS/6000 SP, and the pSeries family of IBM @servers
- ▶ IBM @server iSeries
- ▶ Compaq Alpha servers
- ▶ HP9000 Enterprise servers
- ▶ SUN servers
- ▶ IBM NUMA-Q
- ▶ Intel-based servers: IBM and non-IBM supported servers

iSeries

The IBM @servers iSeries Models 820, 830, 840 and 270, with the Fibre Channel Disk adapter card (fc 2766, shortwave), and running OS/400 Version 5.1 or higher, attach via Fibre Channel to the ESS.

Fibre Channel distances

The type of ESS Fibre Channel host adapter ordered, whether short wave or long wave, and the physical characteristics of the fiber used to establish the link, will determine the maximum distances for connecting nodes to the ESS. See 2.15, "Fibre distances" on page 42 for additional information.

Storage Area Network

The ESS is well suited to being at the heart of a Storage Area Network, because it supports a large number of heterogeneous hosts and can be connected using a wide range of switch and director products.

Being SAN enabled, the ESS can fully participate in current and future SAN implementations, such as LAN-less or server-free backup solutions. For a more detailed description of these implementations, please refer to *Introduction to Storage Area Network, SAN*, SG24-5470.

2.14 FICON host adapters

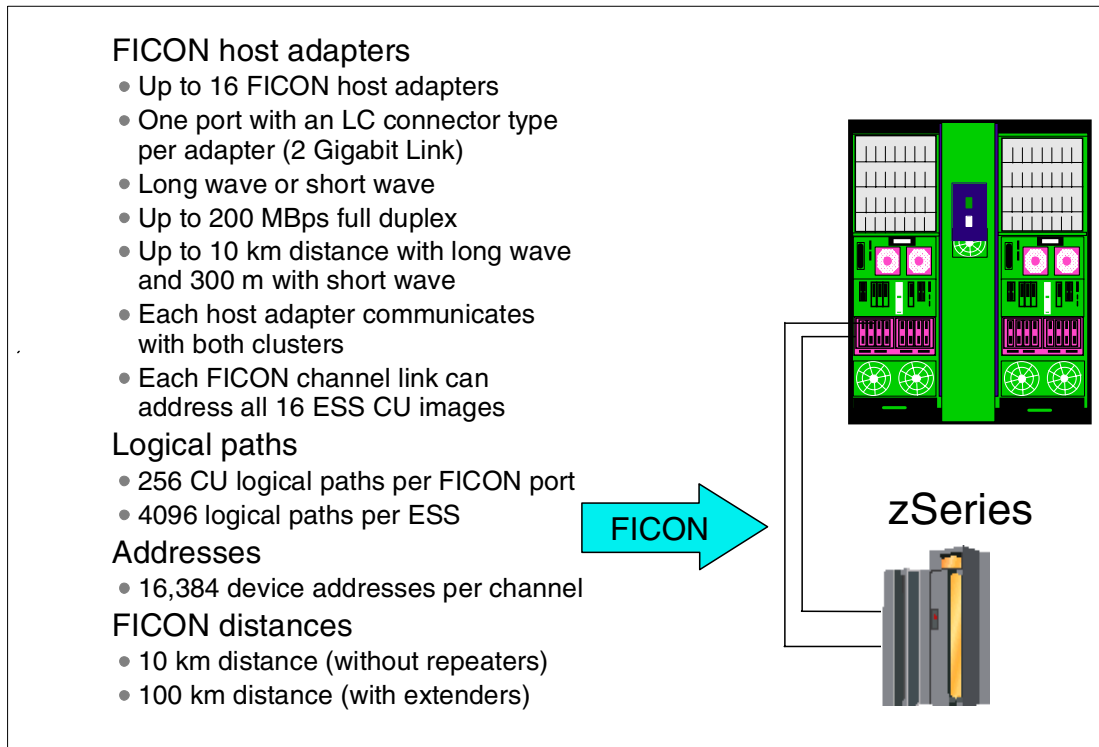


Figure 2-14 FICON host adapters

FICON (Fiber Connection) is based on the standard Fibre Channel architecture, and therefore shares the attributes associated with Fibre Channel. This includes the common FC-0, FC-1, and FC-2 architectural layers, the 100 MBps bidirectional (full-duplex) data transfer rate, and the point-to-point distance capability of 10 kilometers. The ESCON protocols have been mapped to the FC-4 layer, the Upper Level Protocol (ULP) layer, of the Fibre Channel architecture. All this provides a full-compatibility interface with previous S/390 software and puts the zSeries servers in the Fibre Channel industry standard.

FICON goes beyond ESCON limits:

- ▶ Addressing limit, from 1024 device addresses per channel to up to 16,384 (maximum of 4096 devices supported within one ESS).
- ▶ Up to 256 control unit logical paths per port.
- ▶ FICON channel to ESS allows multiple concurrent I/O connections (the ESCON channel supports only one I/O connection at one time).
- ▶ Greater channel and link bandwidth: FICON has up to 10 times the link bandwidth of ESCON (1 Gbps full-duplex, compared to 200 MBps half duplex). FICON has up to more than four times the effective channel bandwidth.
- ▶ FICON path consolidation using switched point-to-point topology.
- ▶ Greater unrepeated fiber link distances (from 3 km for ESCON to up to 10 km, or 20 km with an RPQ, for FICON).

These characteristics allow more powerful and simpler configurations. The ESS supports up to 16 host adapters, which allows for a maximum of 16 Fibre Channel/FICON ports per machine (refer to Figure 2-14).

Each Fibre Channel/FICON host adapter provides one port with an LC connector type. The adapter is a 2 Gb card and provides a nominal 200 MBps full-duplex data rate. The adapter will auto-negotiate between 1 Gb and 2 Gb, depending upon the speed of the connection at the other end of the link. For example, from the ESS to a switch/director, the FICON adapter can negotiate to 2 Gb if the switch/director also has 2 Gb support. The switch/director to host link can then negotiate at 1 Gb.

There are two types of host adapter cards you can select: long wave (feature 3024) and short wave (feature 3025). With long-wave laser, you can connect nodes at distances of up to 10 km (without repeaters). With short-wave laser, you can connect distances of up to 300m. See 2.15, “Fibre distances” on page 42 for more details. These distances can be extended using switches/directors.

As an alternative to the 31 meter cables, an LC to SC “female” 2-meter cable can be ordered with the ESS to enable connection of the adapter port to existing cable infrastructure (see the cable option features in Table A-4 on page 266).

Note: The Fibre Channel/FICON host adapter is an adapter card that supports both FICON or SCSI-FCP, but not simultaneously; the protocol to be used is configurable on an adapter-by-adapter basis.

Topologies

When configured with the FICON host adapters, the ESS can participate in point-to-point and switched topologies. The supported switch/directors are INRANGE FC9000 (IBM 2042) or McDATA ED-6064 (IBM 2032). The most updated list of supported switches and directors that you should consult is found at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

FICON supported servers

FICON is used for attaching the ESS to the IBM zSeries family of servers, and to the IBM S/390 processors 9672 G5 and G6 family of servers. These environments benefit from FICON's exceptional characteristics.

2.15 Fibre distances

The distance characteristics when using fiber on FICON or Fibre Channel implementations are common to both and are a function of the ESS host adapter cards being used (long wave or short wave) and of the fiber being used (9, 50, or 62.5 micro-meters).

Table 2-2 on page 43 shows the supported maximum distances for the Fibre Channel/FICON host adapters of the ESS. As can be seen in the table, the distance is dependent on the host adapter card and the type of fiber. The quality of the fiber also determines the distance. Remember that distances up to 100 km can be supported with the appropriate fabric components. Refer to *Fiber Optic Link Planning*, GA23-0367 for detailed considerations on fiber distances.

Table 2-2 Fibre distances

Adapter	Transfer Rate	Cable Type	Distance
FC 3024 (long wave)	1 Gb	9 micron single mode	10 km
	2 Gb	9 micron single mode	10 km
	1 Gb	50 or 62.5 micron multimode	550 m
FC 3025 (short wave)	1 Gb	62.5 micron multimode (200 MHz per km)	300 m
	2 Gb	62.5 micron multimode (200 MHz per km)	150 m
	1 Gb	62.5 micron multimode (160 MHz per km)	250 m
	2 Gb	62.5 micron multimode (160 MHz per km)	120 m
	1 Gb	50 micron multimode	500 m
	2 Gb	50 micron multimode	300 m

Note: The Fibre Channel / FICON (long wave) ESS host adapter (fc 3024) operating at 1 Gb can be used with existing 50 and 62.5 micron ESCON cables, when the Mode Conditioning Patch cables are used. The Mode Conditioning Patch cables are not supported at 2 Gb.

The *IBM TotalStorage ESS Introduction and Planning Guide*, GC26-7444, should be consulted for more details.


2.16 Power supplies

Power supplies

- Dual power cords
 - Base frame
 - Expansion Enclosure
- Three phase power only
 - Base and Expansion Enclosures

Power redundancy

- N+1
 - Cages, DC-DC power, Cooling
- 2N
 - Processor, I/O, and HA Bay drawers
 - AC supply for base and enclosure
- Battery
- Mirrored power to electronics drawers



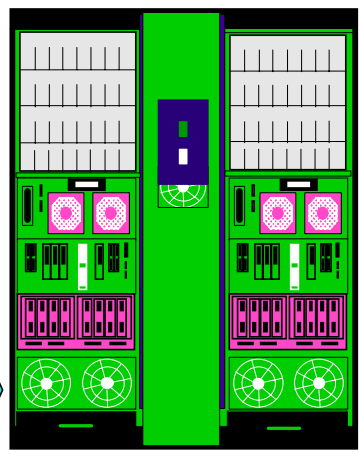


Figure 2-15 ESS Model 800 power supplies

The IBM TotalStorage Enterprise Storage Server is a fault-tolerant subsystem and has dual power cords to protect against power loss. The power units are:

- ▶ N+1, so a failure in a single active power unit has no effect and the ESS will continue normal operations
- ▶ 2N — two active supplies, each capable of providing the full requirements in case the other one fails

In either implementation, the failing power unit can then be replaced nondisruptively. Likewise, the fans can be replaced nondisruptively should they fail.

2.16.1 Power characteristics

As Figure 2-15 on page 43 illustrates, the two ESS enclosures (base and expansion) each require two three-phase supplies. The voltage ranges available are 200 to 240 volts, or 380 to 480 volts. You have a selection of 30, 50 or 60 Amp cables with a variety of connectors from which to choose. The ESS requires a maximum of 6.4 kVA for a full base frame and up to 13.8 kVA for a full 27.9 TB.

The ESS has two bulk external power supplies, each with their own line cord. In the event of the failure of one power input (each line cord should be attached to an independent power source), the remaining power input is sufficient for the ESS to continue normal operation.

The power to each electronics drawer within the cluster is mirrored (2N), thus giving dual power supply to the processors, cache, NVS, device adapters, and host adapter bays and cards.

2.16.2 Battery backup

The ESS has a battery (one per cluster) with sufficient capacity to provide power for a fully configured ESS for a minimum of five minutes if power from both line cords is lost.

When power to both line cords is lost, the ESS remains operational for the first 50 seconds. If the power failure persists beyond the first 50 seconds, the ESS will begin an orderly shutdown:

- ▶ The ESS stops all host data access and initiates the destaging of all modified data in cache to disk.
- ▶ When all modified data destaging has completed, an orderly shutdown commences that includes shutdown of the operating system in each cluster, then power off.
- ▶ Once the orderly shutdown has started, it cannot be interrupted.

If power is restored to either line cord before the 50 seconds elapses, then normal ESS operation continues.

2.17 Other interfaces

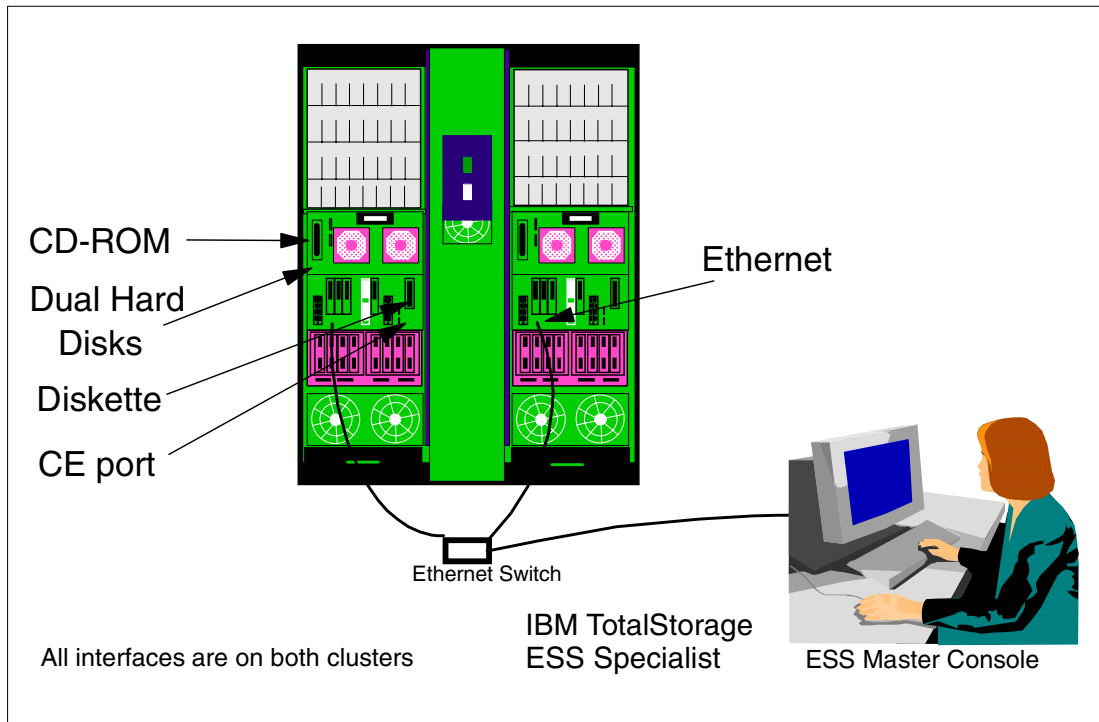


Figure 2-16 ESS Model 800 - other interfaces

Each cluster of the ESS has external interfaces that allow for Licensed Internal Code (LIC) installation/update and offloading of information (refer to Figure 2-16).

The CD-ROM drive can be used to load the Licensed Internal Code when LIC levels need to be upgraded. Both clusters have a CD-ROM drive, diskette drive, and dual hard disk drives that are used to store both the current level of LIC and a new level, or the current active level and the previous level.

The CE port is used by the IBM System Support Representative (SSR) to connect the CE Mobile Solution Terminal (MOST). This allows the SSR to set up and test the ESS, and to perform upgrades and repair operations, although most of these SSR activities are performed from the ESS Master Console (see 2.18, “ESS Master Console” on page 46).

The customer interface is through an Ethernet connection (10/100BaseT) from the IBM TotalStorage ESS Specialist (ESS Specialist) running on the ESS to a customer-supplied Web browser (Netscape or Internet Explorer). This interface allows you to configure the host adapters, RAID ranks, and assign capacity to the various hosts. Details of how to configure the ESS are in Chapter 4, “Configuration” on page 93.

The two clusters come from the factory connected together by a simple Ethernet cable. During installation, the IBM SSR representative connects both clusters to an Ethernet switch. This way, the two clusters can communicate with each other as well as with the ESS Master Console. The switch has a port available to enable users to interconnect their network and the ESS network to provide access to the ESS Specialist from an external browser.

The ESS contains two service processors, one in each cluster, that monitor each cluster and handle power-on and re-IML of the RISC processors.

The diskette drives are used to back up to an external medium the ESS configuration data and other information that is used by the IBM SSR during some upgrade activities. Two copies of the configuration data are stored internally on the dual hard disks (one pair of disks in each cluster).

2.18 ESS Master Console

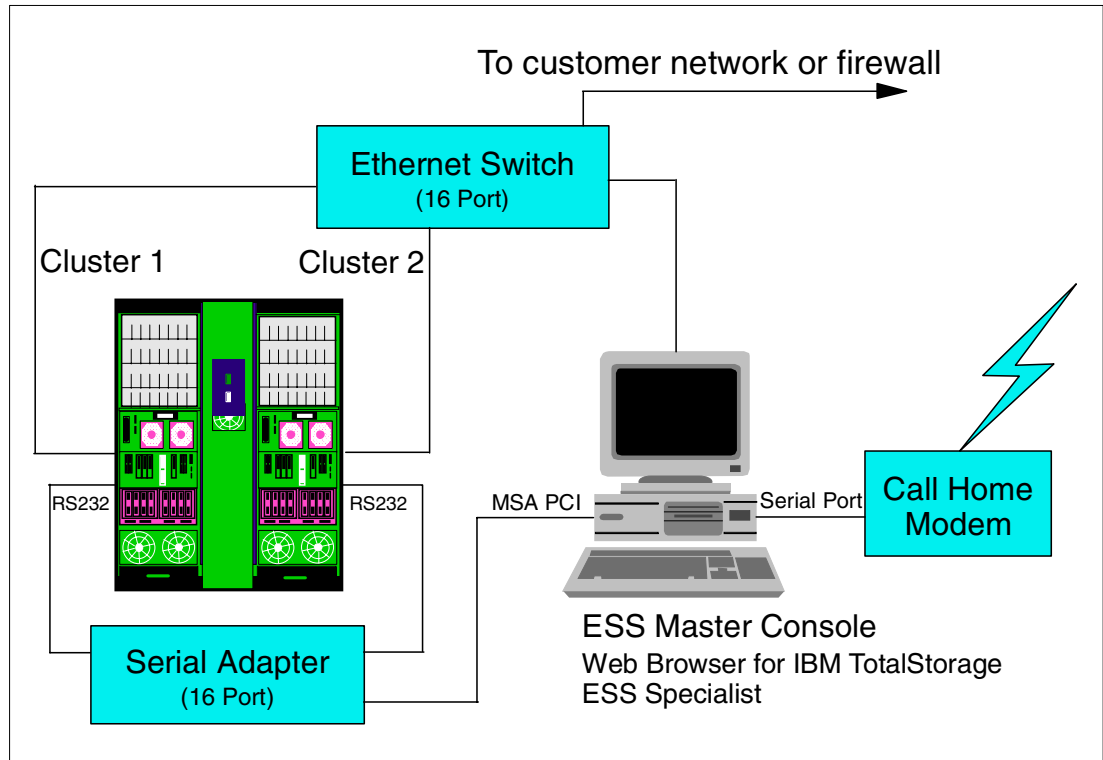


Figure 2-17 ESS Master Console

The IBM TotalStorage Enterprise Storage Server Master Console (ESS Master Console) is a PC running Linux and some specialized applications, which can connect up to seven ESS subsystems via an Ethernet LAN — making the ESSNet. The ESS Master Console allows data transfer to and from each ESS in order to perform Call Home, Remote Support and Service, in addition to giving access to the ESS Specialist (via a Web browser) for the logical configuration tasks.

The ESS Master Console feature consists of a dedicated console (processor, modem, monitor, keyboard, and multiport serial adapter) and networking components (switch and Ethernet cables). One ESS Master Console (feature 2717) must be installed with the first ESS installed. Six additional ESS machines can utilize this console by using the remote support cables (feature 2716).

Call Home

The Call Home portion of the console has the capability to receive data from the ESS and send it to different IBM catcher systems. Generally data is used for error notification or machine status, repair action support, trace data offload, performance analysis, statistics, and MRPD information.

Remote service and support

The console provides a graphical interface for remote service call-in. The local IBM SSR can also perform service functions, simultaneously if needed, on multiple ESSs. The console also manages other service-oriented functions, such as LIC download and activation for multiple ESSs.

Security considerations for the user's data portion of the ESS in relation to the Call Home and Remote Service Support features of the ESS Master Console are discussed in 3.4.1, "Call Home and remote support" on page 54.

Configuration tasks

The ESS Master Console is preloaded with a browser (Netscape Navigator) that provides access to the ESS Specialist and the ESS Copy Services via the Web interface, to enable configuration of the ESS.

2.18.1 ESS local area network

The IBM ESSNet (Enterprise Storage Server Network) is a dedicated local area network connecting up to seven IBM TotalStorage Enterprise Storage Servers. The ESS Master Console is connected to the ESSNet as illustrated in Figure 2-17 on page 46.

Where more than seven ESSs are used, their ESSNets may be interconnected creating an expanded single ESSNet. Alternatively, one or more ESSNets may be connected to an enterprise wide-area network enabling control of ESSs from a central location, regardless of where they may be located.

The ESSNet is a self-contained Ethernet LAN. An Ethernet 10/100BaseT switch is provided as part of the Master Console feature (fc 2717). The IBM System Support Representative will attach the LAN cables (one per cluster) from the ESS to the Ethernet switch. The switch is then connected to the ESS Master Console.

Attachment to the customer's local area network permits access to the ESS Specialist outside the immediate area of the ESS. But it also has the potential of increasing the number of people who can access, view, administer, and configure the ESS.

If you want to attach your company LAN to the ESSNet LAN, you will need to provide the required TCP/IP information to the IBM System Support Representative (the ESSNet needs to be configured with TCP/IP addresses that are recognized as part of your IP network). The IBM SSR will connect your LAN cable to the ESSNet switch and enter the required TCP/IP information.

The ESS can be installed in a secured-LAN, limited-access environment. IBM recommends that the ESS network be established so that it is limited to those requiring access to the ESS. It is not recommended that it be installed on the enterprise intranet nor the worldwide Internet. Installing the ESSNet behind a firewall is one method of ensuring ESS security.

Physical setup

The ESSNet components are supplied with the ESS Master Console (feature 2717) of the IBM TotalStorage Enterprise Storage Server. This feature, besides providing the console, also provides the modem, switch, cables, and connectors required to attach the first ESS to the telephone system. For additional ESSs, feature 2716 provides the cables to connect the second through seventh ESS.

The ESSNet will require:

- ▶ Two power outlets for the console and monitor
- ▶ One power outlet for the Ethernet switch
- ▶ One remote support modem

If the ESSNet is to be connected to the customer's LAN, Ethernet cables need to be obtained. No cable is provided to attach the ESSNet to the customer's network.



Architecture

This chapter describes the IBM TotalStorage Enterprise Storage Server Model 800 architecture. The logical structures that make up the ESS architecture are presented, such as the Logical Subsystem (LSS), and the logical view that the host servers have of the ESS are also discussed. This chapter illustrates the data flow across the components of the ESS, for both read and write operations. Also described in this chapter are the disk sparing, the RAID 5 and RAID 10 rank implementations of the ESS, and the availability features of the ESS for failure handling and maintenance.

3.1 Overview

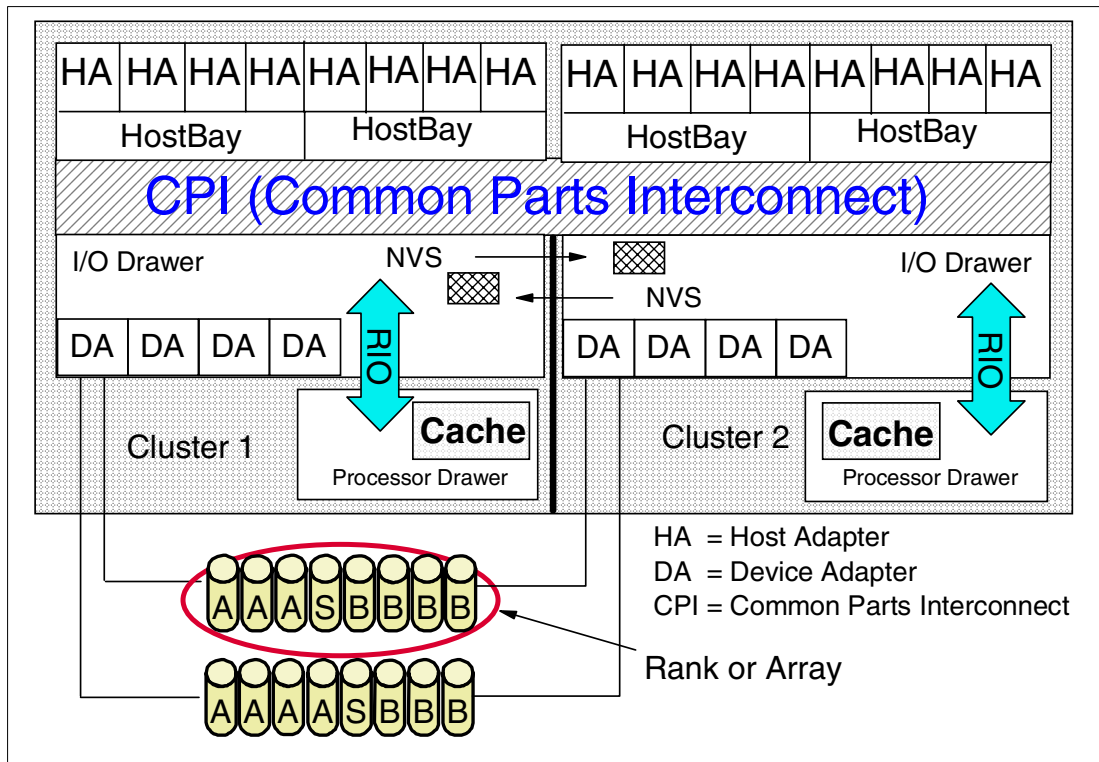


Figure 3-1 ESS Model 800 design overview

A simplified diagram of the IBM TotalStorage Enterprise Storage Server Model 800 architecture is presented in Figure 3-1. At the top we have up to 16 host adapters (HAs) in four host adapter bays (HBs). Each host adapter can have either one Fiber Channel/FICON port, or two ESCON ports, or two SCSI ports. Also, each host adapter is connected to both clusters through the Common Parts Interconnect (CPI) buses, so that either cluster can handle I/O from any host adapter. With the ESS Model 800, the bandwidth of the CPI has been doubled for a higher and faster data throughput.

Each cluster consists of a processor drawer and an I/O drawer (illustrated in Figure 2-3 on page 22). Together, the I/O drawers provide 16 PCI slots for logic cards, eight for device adapter cards (DA) and eight for NVS adapter cards (NA cards, are not detailed in the simplified diagram of Figure 3-1). Each NA card has NVS memory and a CPI bus to the host adapter. The PCI buses in the I/O drawers are connected via a Remote I/O (RIO) bridge to the processor drawer. The processor drawer of each cluster includes a SMP processor (the standard or the optional turbo feature), up to 32 GB cache (per cluster) and two mirrored SCSI drives. The I/O drawer includes the SSA device adapter cards (DA cards), IOA cards (including 1 GB NVS per cluster), the NVS batteries, and the SCSI adapter for the internal SCSI drives.

The ESS Model 800 includes new SSA device adapter cards for increased performance. The NVS for Cluster 1 data is in Cluster 2, and vice versa. Within each cluster we have four device adapters (DA). They always work in pairs, and the disk arrays are attached through SSA loops to both DAs in a pair. The disk arrays can be configured as RAID 5 ranks, or as RAID 10 ranks. All these components are described in further detail in the following sections of this chapter.

3.2 Data availability architecture

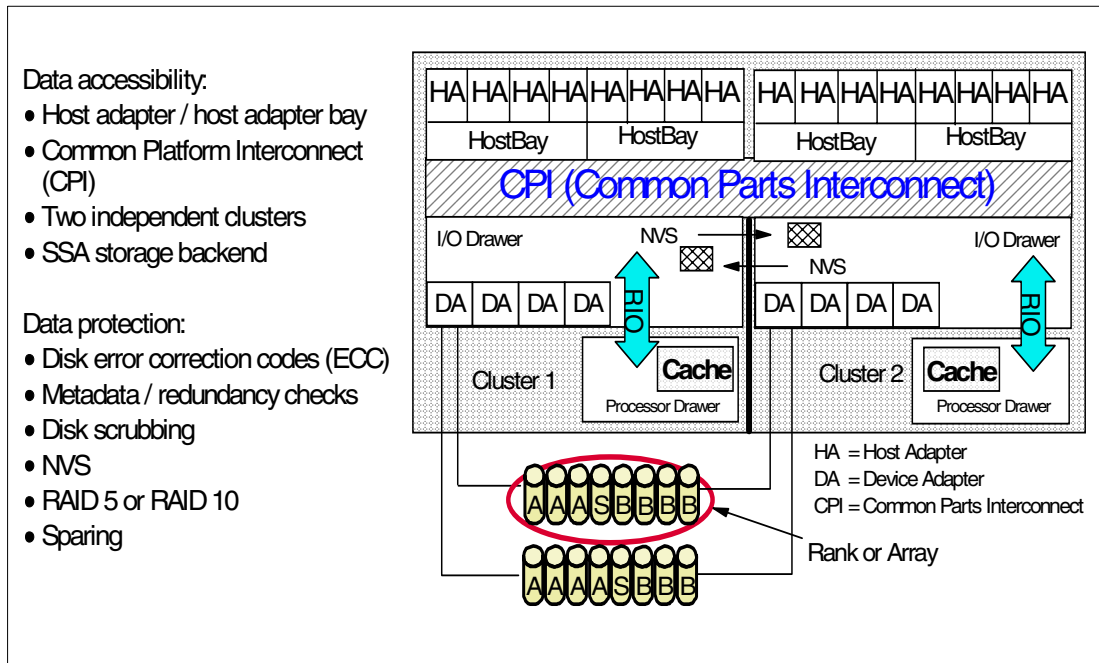


Figure 3-2 Data availability design

Today's typical business environment relies heavily upon IT, and the data stored on storage subsystems represents a crucial company asset. The IT department must ensure that continuous access to data is possible, that data is protected against corruption or loss, and that the system is resilient enough to tolerate hardware components outages.

The IBM TotalStorage Enterprise Storage Server Model 800 has implemented an extensive set of features to prevent data loss or loss of access to the data, with minimum or no impact on performance.

3.2.1 Data accessibility

The architecture of the ESS provides several mechanisms to ensure continuous access to the user data:

1. Host adapters/host adapter bays.

The connection of the ESS to the application servers is done by means of the ESS host adapters (HA), four of which reside in each of the four host adapters bays (HBs). This allows connections to the host in a way that at least one path can remain active in the event of a host adapter or a host adapter bay failure (refer to 4.9.3, "Host adapters" on page 101 for configuration recommendations). This characteristic also helps when planning for availability in a SAN environment.

2. Common Platform Interconnect (CPI).

The CPI connects the four host adapter bays, thus all the host adapters, to both clusters I/O drawers. This maintains the connectivity between the host adapters and the storage back end (disks) in case of a host adapter bay or cluster failure.

3. Two independent clusters.

During normal operation of the ESS, both clusters are active performing I/O operations and accessing the stored data. In the event that one cluster has a non-recoverable failure,

then all the activities are transferred to the other cluster (failover), thus ensuring continuous data availability. See 3.8, “Cluster operation: failover/failback” on page 60 for details on the failover process.

4. Powerful Serial Storage Architecture back end.

In the back end, the ESS uses the fault-tolerant SSA architecture, implemented with the device adapters and loops. This SSA architecture, together with the RAID 5 and RAID 10 implementations, give continuous availability and access of the disk data.

3.2.2 Data protection

The ESS is designed to ensure data integrity further beyond the RAID 5 and RAID 10 implementations:

1. Disk Error Checking and Correction.

All disk drives used in the ESS support error checking and correction (ECC). ECC consists of multiple bytes appended to each disk sector by the disk hardware (a sector is the smallest addressable unit of space on a disk drive, sometimes called a *block*). ECC is used by the disk hardware to detect and correct selected bit errors in a sector, so that the correct data can be returned in response to read requests. ECC is actually just one of multiple technologies used by disk drives in the ESS to protect data integrity. When the problem is not solved by this first layer of recovery techniques, then the ESS will use its RAID protection capabilities to regenerate data.

2. Metadata checks.

When application data enters the ESS, special codes or metadata, also known as *redundancy checks*, are appended to that data by an ESS host adapter. This metadata remains associated with the application data as it is transferred throughout the ESS. The metadata is checked by various internal components to validate the integrity of the data as it moves throughout the disk system. It is also checked by the host adapter before the data is sent to the host in response to a read I/O request. Further, the metadata also contains information used as an additional level of verification to confirm that data being returned to the host is coming from the desired disk volume.

3. Disk scrubbing.

The ESS will periodically read all sectors on a disk. (This is designed to occur without any interference to application performance.) If ECC-correctable bad bits are identified, the bits are corrected immediately by the ESS. This reduces the possibility of multiple bad bits accumulating in a sector beyond the ability of ECC to correct them. If a sector contains data that is beyond ECC's ability to correct, then RAID is used to regenerate the data and write a new copy on a spare sector on the disk.

4. Non-volatile storage (NVS).

The ESS, like most cached disk storage systems, protects the data that the applications write by keeping one copy in cache and a second copy in NVS until the data is destaged to the RAID-protected disks. This NVS copy is important because having only a single copy of new/changed data in cache (not yet destaged to disk) would make that data vulnerable to loss in the event of, for example, an ESS power failure.

3.3 ESS availability features

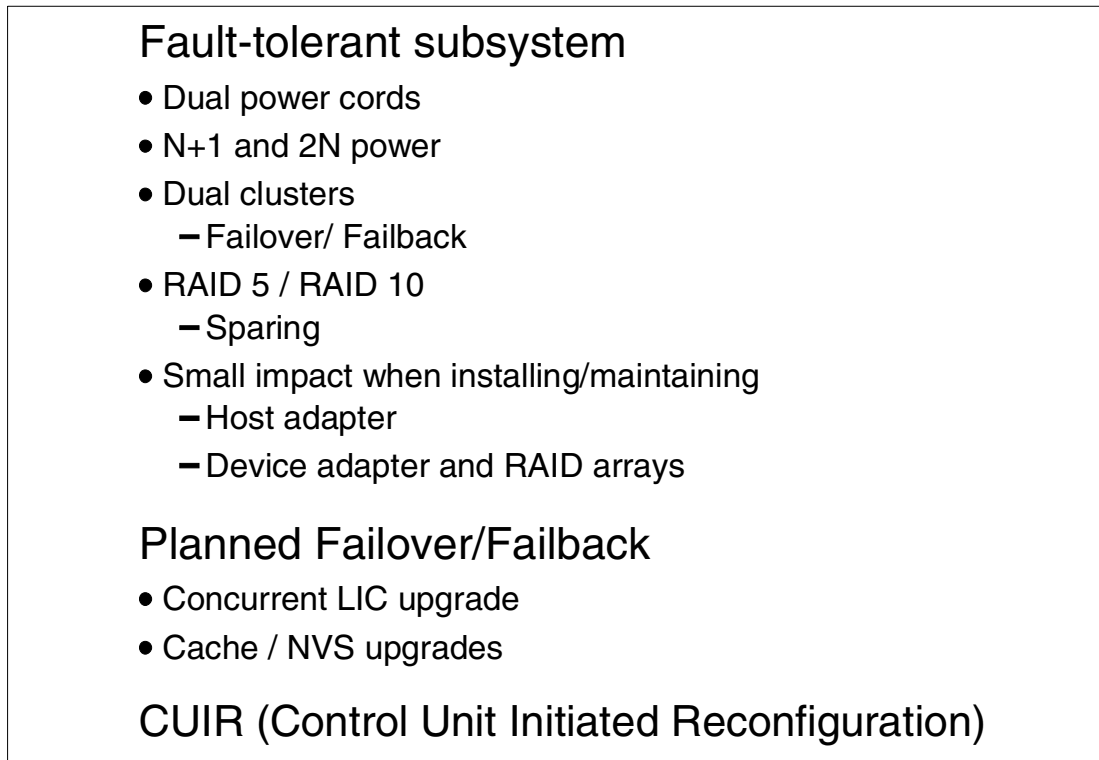


Figure 3-3 ESS Model 800 availability features

As summarized in Figure 3-3, there are many design characteristics that the IBM TotalStorage Enterprise Storage Server has integrated for continuous availability of the data.

3.3.1 Fault-tolerant subsystem

Several features and functions make the IBM TotalStorage Enterprise Storage Server Model 800 a fault-tolerant subsystem:

- ▶ Dual power supplies are (2N), each with its own line cord. Each of the two power supplies is capable of powering the whole subsystem.
- ▶ DC power supplies are (2N) for the processor drawers, I/O drawers, and the host adapter bays. All these components are powered by two DC power supplies, where one can fail without affecting the operation.
- ▶ Disk arrays in the ESS Model 800 are protected by RAID 5 or RAID 10 against data loss in case of a failing DDM.

3.3.2 Cluster failover/failback

The ESS consists of two active clusters, where under normal operation each is accessing its set of disk drives. Clusters are independent, but each cluster can operate all host connections and access all the disks in case the other cluster fails. The failover/failback function is used to handle both unplanned failures and planned upgrades or configuration changes, eliminating most planned outages and thus providing continuous availability.

3.3.3 CUIR

Control Unit Initiated Reconfiguration (CUIR) prevents volumes access loss in S/390 environments due to wrong path handling. This function automates channel path quiesce and resume actions in S/390 environments, in support of selected ESS service actions.

3.4 Maintenance strategy

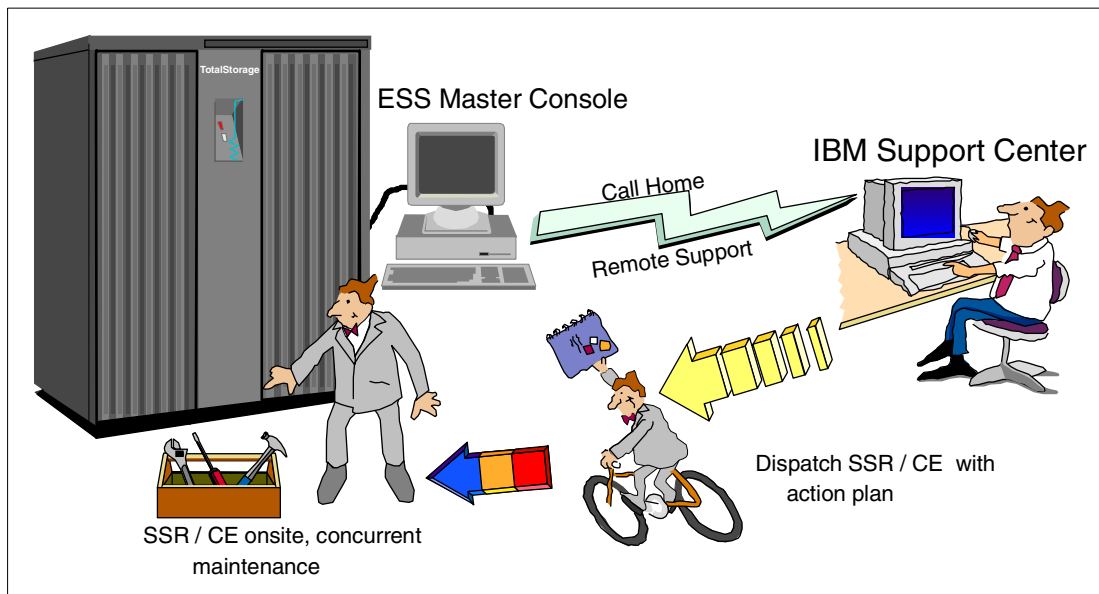


Figure 3-4 ESS Model 800 maintenance strategy

The elements in play in the IBM TotalStorage Enterprise Storage Server maintenance strategy are illustrated in Figure 3-4. An important part of the maintenance strategy is the capability of the ESS to place a *Call Home* in case of a failure, as well as the possibility of receiving *remote support*. These two features are key for quick and efficient maintenance. Both, the dial-in and the *Call Home* functions go through the ESS Master Console (described in 2.18, “ESS Master Console” on page 46).

3.4.1 Call Home and remote support

The *Call Home* feature of the ESS enables it to contact the IBM Support Center directly in case of a failure. The advantage of the *Call Home* feature is that the user has a 7-day/24-hour sentinel. The IBM Support Center receives a short report about a failure. With that information, the IBM Support Center is able to start analyzing the situation by using several databases for more detailed error information. If required, the IBM Support Center will be able to connect to the ESS, in case additional error logs, traces, or configuration information are needed for a more accurate failure analysis.

The capability of dialing into the machine also allows the IBM Support Center to help the user with configuration problems, or the restart of a cluster after a failover, as well as for periodic health checks. The *Remote Support* capability is a password-protected procedure, which is defined by the user and entered by the IBM System Support Representative (SSR) at installation time.

Note: The Remote Support function is able to connect the ESS Model 800 only for analysis and maintenance activities in case of a problem. There is *no* way to access any user data residing on the array disk drives.

If the ESS Master Console (the ESSNet) is integrated into the user LAN, then the ESS is able to e-mail problem notifications to selected users, as well as page or send SNMP alerts (see 4.14, “ESSNet setup” on page 108).

User’s data security considerations

The ESS and the ESS Master Console implementations are designed to ensure maximum security in a network environment. Some of the more relevant considerations are the following:

- ▶ An onsite IBM System Support Representative (IBM SSR) only has access to the *Service* menu functions and must be directly connected via his service terminal of the ESS Master Console to the ESS, that is, the IBM SSR must be physically present where the ESS and the ESS Master Console are located.
- ▶ Remote access to the user’s network is not possible via a modem. TCP/IP functions to access the network have been removed from the ESS and from the ESS Master Console. The remote connection into the ESS Master Console via modem and into the ESS via a serial connection is a *remote terminal* session that does not support TCP/IP functions.
- ▶ No user data can be transferred from the ESS to an attached ESS Master Console.
- ▶ The authentication schemes utilized for connecting into the ESS and into the ESS Master Console are independent from each other, that is, successfully authenticating with the ESS Master Console still requires an additional authentication for each attached ESS.
- ▶ There are two levels of remote access to the ESS:
 - The *Support Level* access authorization option is set by an onsite IBM SSR. The standard support access password is part of the Call Home Record sent to IBM. An optional remote access password is known only by the onsite IBM SSR and the user. User *support* can only view the machine’s configuration, settings, and logs.
 - The *Product Engineering (PE) Level* access requires a password to be supplied by the user or an onsite IBM SSR. The password expires after seven days. PE Level access has a higher authority than Support Level access has, and can change the ESS configuration and settings. In order to get privileged access to the ESS, PE Level access must overcome an expiring challenge/key password scheme.
- ▶ There are two levels of remote access to the ESS Master Console:
 - The *Support Level* access authorization option is set by an onsite IBM SSR. The standard support access password is part of the IBM internal support structure. The standard support access password can be changed to be unique for each ESS. The Support Level access does not allow viewing items. Nothing on the ESS Master Console or the ESS can be changed.
 - *Product Engineering (PE) Level* access requires authorization by the user or an onsite IBM SSR. The password used is a challenge/key password and expires after 48 hours. PE is a privileged user and can change the ESS Master Console’s configuration and settings.
- ▶ Any local or remote system/resource access attempt to the ESS Master Console is logged, whether it was successful or not.
- ▶ Only a privileged user can change the security settings of the ESS Master Console.

Product Engineering data can be sent to IBM using the Internet and IBM's anonymous FTP server. This option is intended for those users who desire significantly faster data offload times than are possible using the ESS Master Console's modem connection, but this process should only be used when there is a network firewall between the ESSNet and the Internet.

The responsibility for the network security is the task of the user's network administrator. It is the responsibility of the user to provide a secure connection between the ESSNet and the Internet. Typically this means providing a network firewall between the ESSNet and the Internet that supports some form of outbound FTP firewall service while blocking all other inbound forms of network access.

3.4.2 SSR/CE dispatch

After failure analysis using the remote support facility, then the IBM Support Center is able to start an immediate SSR/CE dispatch and parts order (FRU) if the reported problem requires any onsite action. The IBM SSR/CE will get a first-action plan that will most likely solve the situation on site. That action plan is based on the analysis of the collected error data, additional database searches, and if required, development input. All this occurs without any intervention by the user and helps to solve most problems in a very short time frame without customer impact or intervention. Often, the problem diagnosis and the first-action plan are available before the customer is aware of an existing problem and before being informed by the IBM Support Center.

3.4.3 Concurrent maintenance

An onsite SSR/CE is able to run nearly all maintenance actions concurrently with the user's operation at the IBM TotalStorage Enterprise Storage Server. This is possible due to the fault-tolerant architecture of the ESS. Procedures such as cluster Failover/Failback (explained in 3.8, "Cluster operation: failover/failback" on page 60) will allow a service representative to run service, maintenance, and upgrades concurrently, if configured properly.

3.5 Concurrent logic maintenance

The architecture of the ESS Model 800 allows nearly all maintenance actions such as repairs, code upgrades and capacity upgrades to be done concurrently. Figure 3-5 on page 57 provides details about the logic maintenance boundaries of the ESS. These boundaries are the basis for IBM concurrent maintenance strategy.

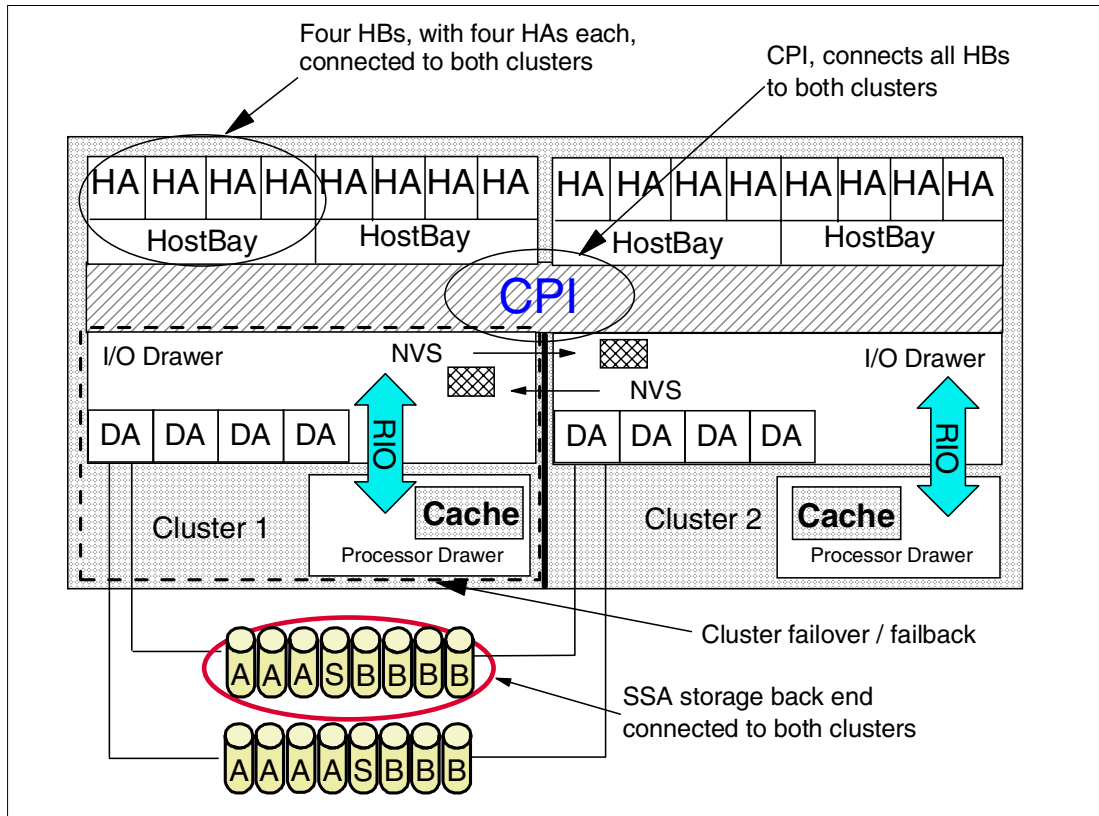


Figure 3-5 Concurrent maintenance

Concurrent maintenance actions

Concurrent means that, while an IBM System Support Representative (SSR) is working on the ESS, users can continue running all applications on the ESS. All logic components are concurrently replaceable. Some of them will even allow hot plugging. The following list indicates the components that are concurrently replaceable and upgradeable:

- ▶ Cluster logic

All components belonging to the processor drawer and I/O drawer, such as DA cards, IOA cards, cache memory, NVS and others, can be maintained concurrently using the failover/failback procedures. The cluster logic also manages the concurrent LIC load.

- ▶ Disk drives

The disk drives can be maintained concurrently, and their replacement is hot-pluggable. This is also valid for SSA loop cables.

- ▶ Code upgrade

The Licensed Internal Code, the control program of the ESS, is designed in such a way that an update to a newer level will take place while the machine is operational using the failover/failback procedure.

- ▶ Concurrent upgrades

The ESS is upgradeable with host adapter cards, cache size, DA cards, and disk drives. Whenever these upgrades are performed, they will run concurrently. In some cases the failover/failback procedure is used. The upgrade of a host adapter card will impact other cards within the same host adapter bay.

3.6 Concurrent power maintenance

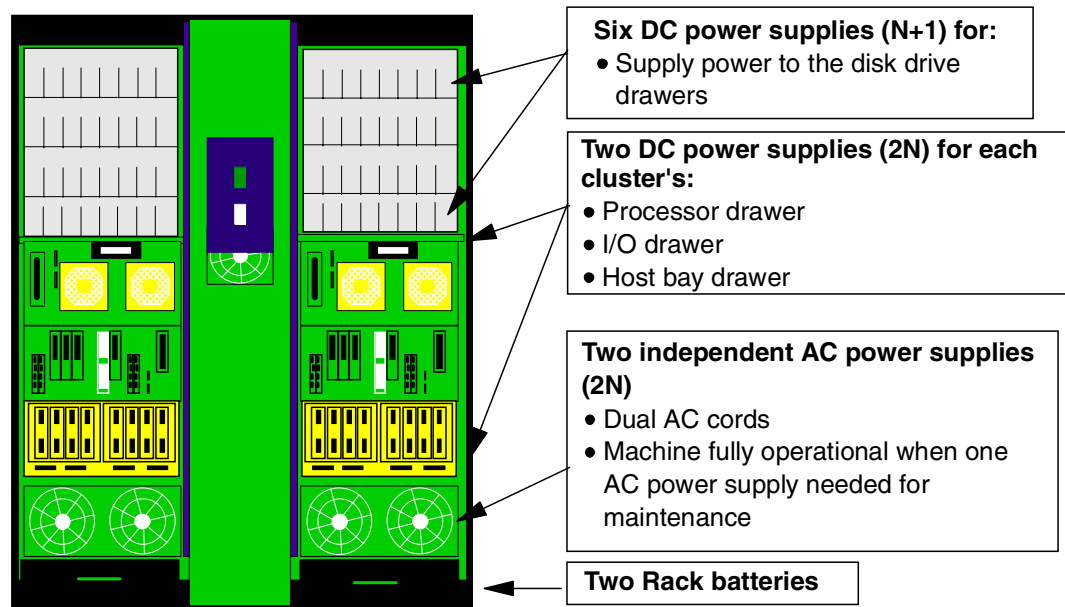


Figure 3-6 Concurrent power maintenance

Figure 3-6 shows the main units that power the ESS. All maintenance actions required in the power area are concurrent, including replacement of failed units as well as any upgrades. The three power areas in the ESS are:

► DC power supplies

DC power requirements for the disk drives is provided by an N+1 concept, where N supplies are enough for normal functioning. This will ensure, in case of outage of one of the DC power supplies, that an IBM System Support Representative (SSR) is able to replace the failed part concurrently.

Power requirements for each cluster's drawers (processor, I/O, and host bays) are provided by a 2N concept. This means that each drawer has two active power supplies, each with enough capacity to provide the full power requirement of the drawer in case the other power supply fails.

► Dual AC distribution

The ESS is a dual AC power machine (thus requiring dual power cords), also with the 2N concept implemented for its AC power requirements. This allows an IBM SSR to replace or upgrade either of the AC supplies.

► Rack batteries

Two rack batteries have been integrated in the racks to allow a controlled destage of cache data to the disk drives and a controlled power-down of the rack in case of a power loss. The IBM SSR will be able to replace them concurrently.

3.7 Sparring

In the disk subsystems with RAID implementations there are spare disk drives, so that in the event of a disk failure, data is rebuilt onto a spare disk drive.

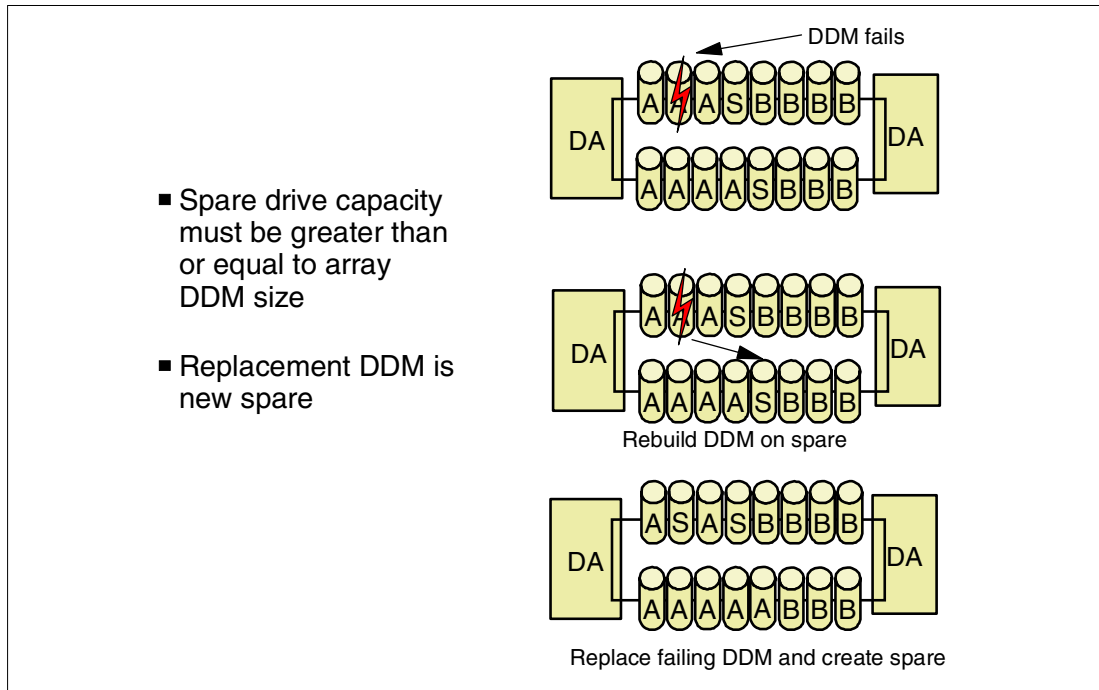


Figure 3-7 Sparring

3.7.1 Sparring in a RAID rank

Figure 3-7 illustrates how sparring is handled within a RAID array. The principle of sparring is the same for all arrays, whether they are configured as RAID 5 ranks or RAID 10 ranks.

Every SSA loop in the ESS has a minimum of two spare disks per capacity; this is the *spare pool*. For a configuration with an intermix of different capacity eight-packs in the same loop, two more spare disks will be reserved in the loop for each of the different capacities. Because it is possible to have up to six eight-packs per SSA loop and each pair of eight-packs could be of a different capacity from the other four, this means that there could be up to six spare disks per loop (two for each of the DDM capacities).

The diagram in Figure 3-7 shows an SSA loop with two RAID ranks and two spare drives. When a disk drive (DDM) fails, the SSA adapter recreates the missing data. If it is a RAID 5 rank, then the SSA adapter rebuilds the data reading the corresponding track on each of the other (data and parity) disk drives of the array, and then recalculates the missing data. If it is a RAID 10 rank, then the SSA adapter rebuilds the data reading it from the mirror disk. The SSA device adapter will, at the same time as normal I/O access, rebuild the data from the failed disk drive on one of the spare disks on the loop.

Two important characteristics of the ESS sparring are that the sparring process is automatically initiated by the ESS, without needing any user action, and that sparring can also be initiated by the ESS based on error thresholds being reached. Both RAID 5 and RAID 10 reconstruction are performed in the background and will generally have no noticeable impact upon performance, although a sparring operation will be finished sooner on RAID 10 than on RAID 5.

Once the rebuild has completed, the original spare is now part of the RAID rank holding user data, while the failed disk drive becomes the new spare (floating spare) once it has been replaced.

3.7.2 Replacement DDM is new spare

Once data has been rebuilt on a spare, it remains there. The replacement disk drive always becomes the new spare, thus minimizing data movement overheads (because there is no requirement to move data back to an original location).

Considerations

The eight-packs of a given capacity that are configured first in the loop will get the spare spool (for that capacity) reserved on them. The initially reserved spare disks can be used by any of the other arrays (of the same capacity) in the loop. Also because the spare disk drive “floats” across the arrays, the RAID rank will not always map onto the same eight disk drives in which it was initially held.

This means that over a period of time the initial relationship of arrays and eight-packs that hold them will change, and some of the disk drives that make up the array will be on different eight-packs in the loop from the ones where they were initially. For this reason, individual eight-packs cannot be removed from a loop without a significant disruption to all the arrays on the loop. You have to back up all the data on the loop, then delete and re-define all the arrays once the eight-pack has been removed.

3.7.3 Capacity intermix sparing

As previously mentioned, with eight-packs of different capacity in one loop we have two spares minimum for each capacity. For example, if there are three different eight-pack capacities in one loop, then there will be a minimum of six spare drives.

Normally when a DDM of a certain capacity fails, it is reconstructed on a spare drive of its same capacity within the loop.

In the uncommon situation where multiple quasi-simultaneous/same capacity DDMs fail (on a loop with an intermixed configuration), then sparing will follow a different rule. When a third DDM of the same capacity fails before one of the previous failed DDMs is replaced and no spare for this capacity is available, the disk will be spared to the next higher capacity spare disk. The excess space in this disk is flagged as unusable. If this happens, the failing DDM that was spared to the higher capacity drive must be replaced by a DDM with the higher capacity.

Example: If a 18.2 GB DDM is failing and spared to a 36.4 GB DDM, half of the 36.4 GB is flagged as unusable. The failing 18.2 GB DDM must be replaced by a 36.4 GB FRU DDM to become a 36.4 GB spare disk.

3.8 Cluster operation: failover/failback

Under normal operation, both ESS clusters are actively processing I/O requests. This section describes the failover and failback procedures that occur between the ESS clusters when an abnormal condition has affected one of them.

3.8.1 Normal operation before failover

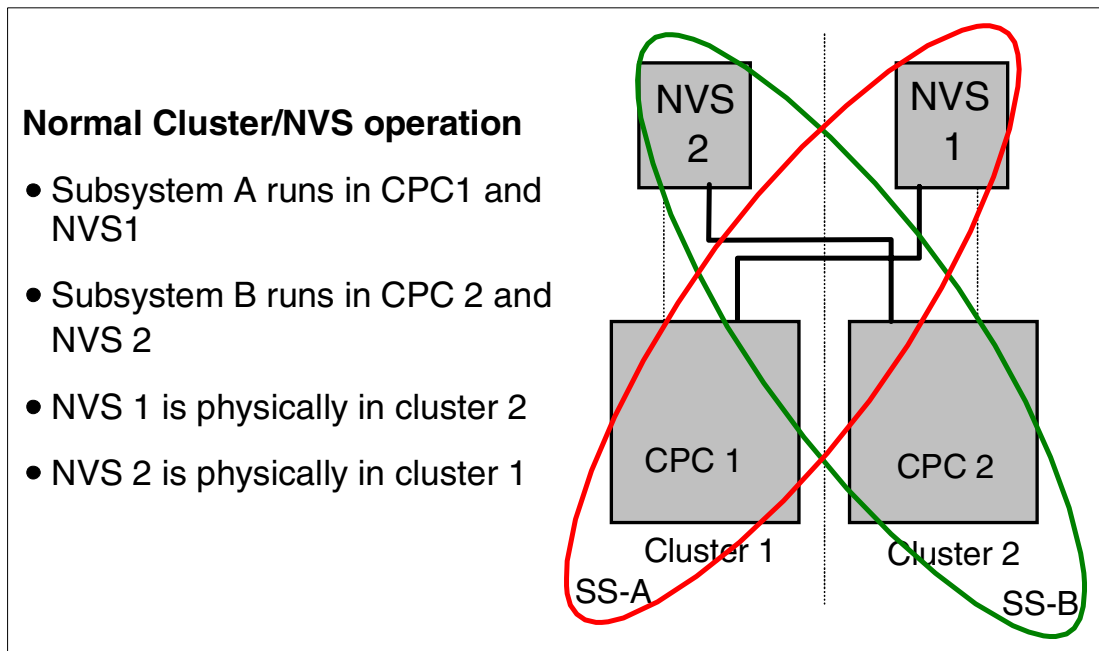


Figure 3-8 Normal cluster operation

The normal setup of the clusters is shown in Figure 3-8. For the purposes of showing how a cluster failover is handled, we use the following terminology:

- ▶ Subsystem A (SS-A): these are functions that normally run in CPC 1 and use NVS 1.
- ▶ Subsystem B (SS-B): these are functions that normally run in CPC 2 and use NVS 2.

The host adapters are connected to both clusters, and the device adapters in each cluster can access all the RAID ranks. In case of a failover, this will allow the remaining cluster to run both subsystems within the one CPC and NVS.

But during normal operation, the two subsystems will be handling different RAID ranks and talking to different host adapters and device adapters.

3.8.2 Failover

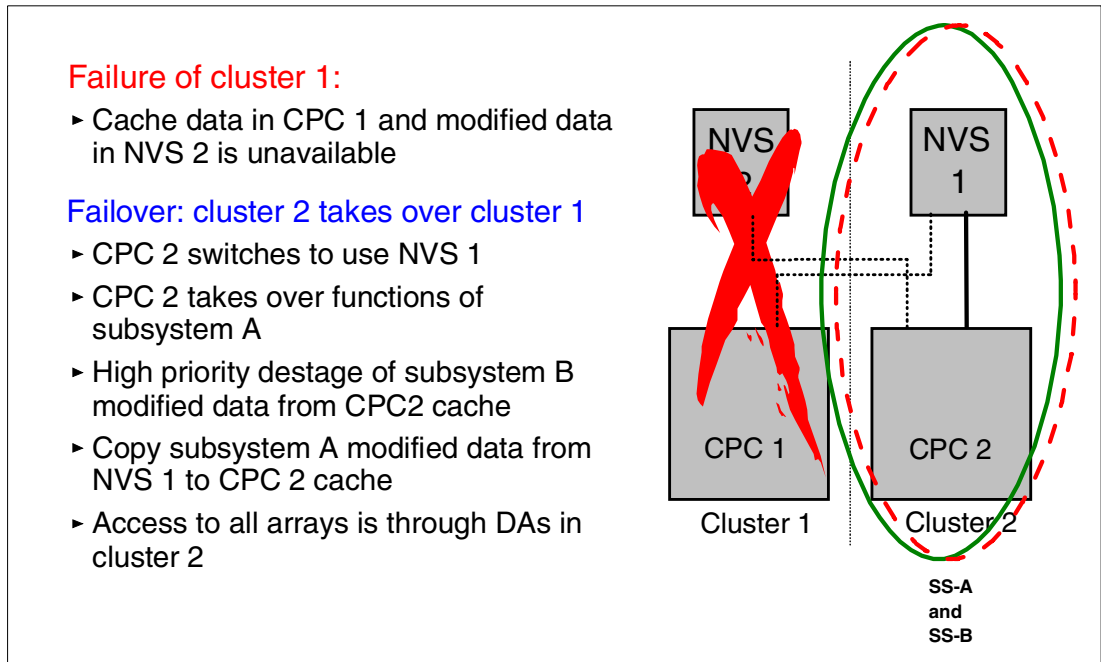


Figure 3-9 Cluster 1 failing - failover initiated

In case one cluster in the ESS is failing, as shown into Figure 3-9, the remaining cluster takes over all of its functions. The RAID arrays, because they are connected to both clusters, can be accessed from the remaining device adapters. Since we have only one copy of data, any modified data that was in cluster 2 in the diagram is destaged, and any updated data in NVS 1 is also copied into the cluster 2 cache. Cluster 2 can now continue operating using NVS-1.

3.8.3 Failback

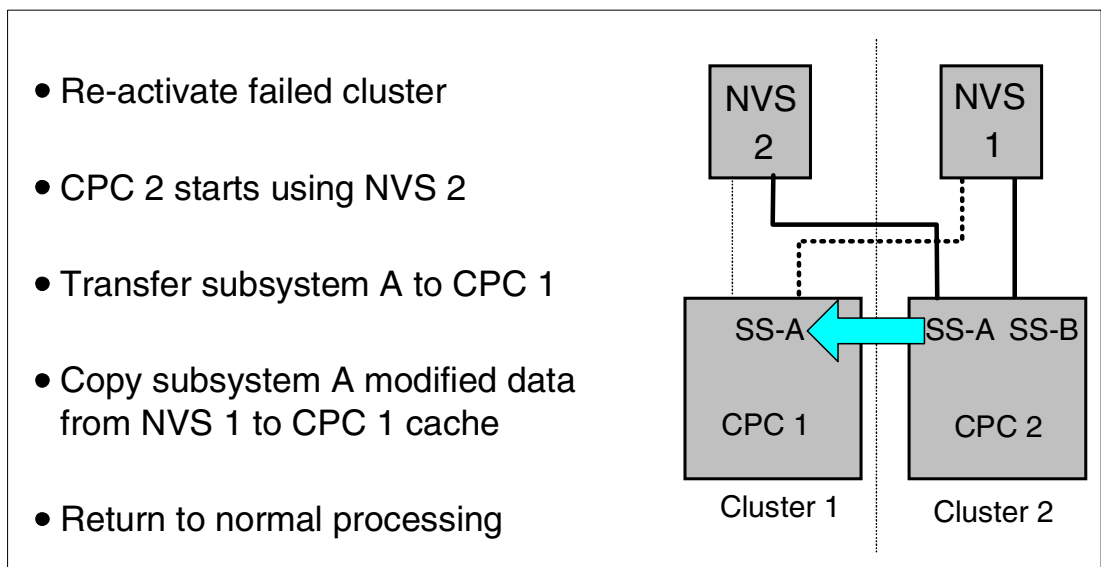


Figure 3-10 Cluster 1 recovered - failback is initiated

When the failed cluster has been repaired and restarted, the failback process is activated. CPC2 starts using its own NVS, and the subsystem function SS-A is transferred back to CPC1. Normal operations with both clusters active then resume.

3.9 CUIR

Control Unit Initiated Reconfiguration (CUIR) is available for ESS Model 800 when operated in the z/OS and z/VM environments. The CUIR function automates channel path quiesce/resume actions to minimize manual operator intervention during selected ESS service or upgrade actions.

CUIR allows the IBM Service Support Representative (IBM SSR) to request that all attached system images set all paths associated with a particular service action offline. System images with the appropriate level of software support will respond to such requests by varying off the affected paths, and either notifying the ESS subsystem that the paths are offline, or that it cannot take the paths offline. CUIR reduces manual operator intervention and the possibility of human error during maintenance actions, at the same time reducing the time required for the maintenance. This is particularly useful in environments where there are many systems attached to an ESS.

For more detailed information in CUIR use and implementation, refer to the IBM redbook *IBM TotalStorage Enterprise Storage Server: Implementing the ESS in Your Environment*, SG24-5420-01.

3.10 RAID Data protection

The IBM TotalStorage Enterprise Storage Server Model 800 disk drives arrays are configured in RAID (Redundant Array of Independent Disks) implementations. In this section, the ESS different RAID implementations are described. The information in this section can be complemented with information from Chapter 4, “Configuration” on page 93.

3.10.1 RAID ranks

The basic unit where data is stored in the ESS is the DDM (disk drive module). Disk drives for the ESS Model 800 are available with capacities of 18.2 GB, 36.4 GB or 72.8 GB. Physically eight DDMs (of the same capacity) are grouped together in an eight-pack, and these eight-packs are installed in pairs on the SSA loops. One SSA loop can hold up to six eight-packs (three pairs), which means that the maximum number of 48 DDMs can be found in a loop. The raw or physical capacity installed on a loop will depend on the eight-packs (their capacity) that are installed on that loop. For detailed information about array capacities, refer to 2.6.4, “Disk eight-pack capacity” on page 26.

Logically eight DDMs (out of an eight-pack pair) are grouped as an ESS array (rank). As illustrated in Figure 3-11 on page 64, initially four DDMs of the first eight-pack and the four DDMs of the second eight-pack make up the rank. This initial correspondence will change with time after initial configuration, due to the floating spare characteristic of the ESS (explained in 3.7, “Sparing” on page 58).

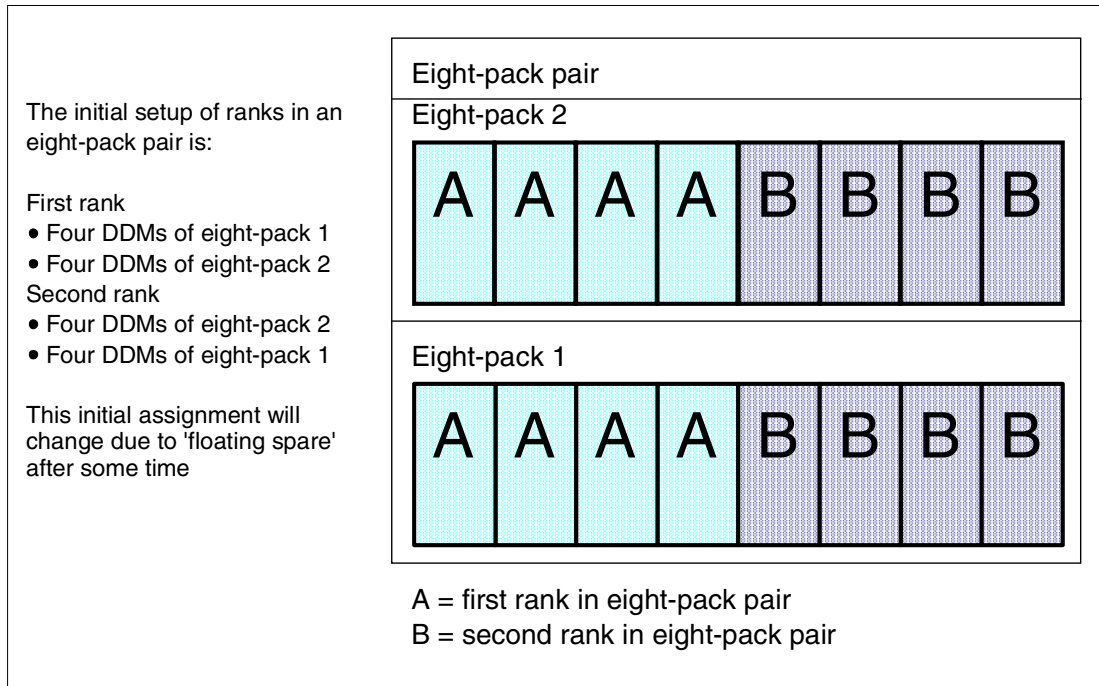


Figure 3-11 Initial rank setup

A RAID rank (or RAID array) is owned by one Logical Subsystem (LSS) only, either an FB LSS or a CKD (S/390) LSS (LSSs are described in “Logical Subsystems” on page 68). During the logical configuration process, the decision is made on which type of RAID rank the array will be. With the ESS Model 800, the ranks can be configured as RAID 5 or RAID 10.

Then each rank is formatted as a set of logical volumes (LV). The number of LVs in a rank depends on the capacity of the disk drives in the array (18.2 GB, 36.4 GB, or 72.8 GB), and the size of the LUNs (for FB attachment) or the emulated 3390 DASDs (for CKD attachment). The logical volumes are also configured during the logical configuration procedure. When configured, the LVs are striped across all the data disks and then mirrored (when it has been defined as a RAID 10 rank) or striped across all data disks in the array along with the parity disk (floating) if it has been defined as RAID 5 rank.

3.10.2 RAID 5 rank

One of the two possible RAID implementations in an ESS Model 800 rank is RAID 5. The ESS RAID 5 implementation consists of eight DDMs: a set of 6 or 7 disks for user data, plus a parity disk for reconstruction of any of the user data disks should one become unusable. In fact there is no dedicated physical parity disk, but a floating parity disk striped across the rest of the data disks in the array. This prevents any possibility of the parity disk becoming an I/O hot spot.

Because the ESS architecture for maximum availability is based on two spare drives per SSA loop (and per capacity), if the first two ranks that are configured in a loop are defined as RAID 5 then they will be defined by the ESS with six data disks plus one parity disk plus one spare disk (6+P+S). This will happen for the first two ranks of each capacity installed in the loop if configured as RAID 5.

Once the two spares per capacity rule is fulfilled, then further RAID 5 ranks in the loop will be configured by the ESS as seven data disks and one parity disk (7+P). Figure 3-12 on page 65

illustrates the two arrangements of disks that are done by the ESS when configuring RAID 5 ranks.

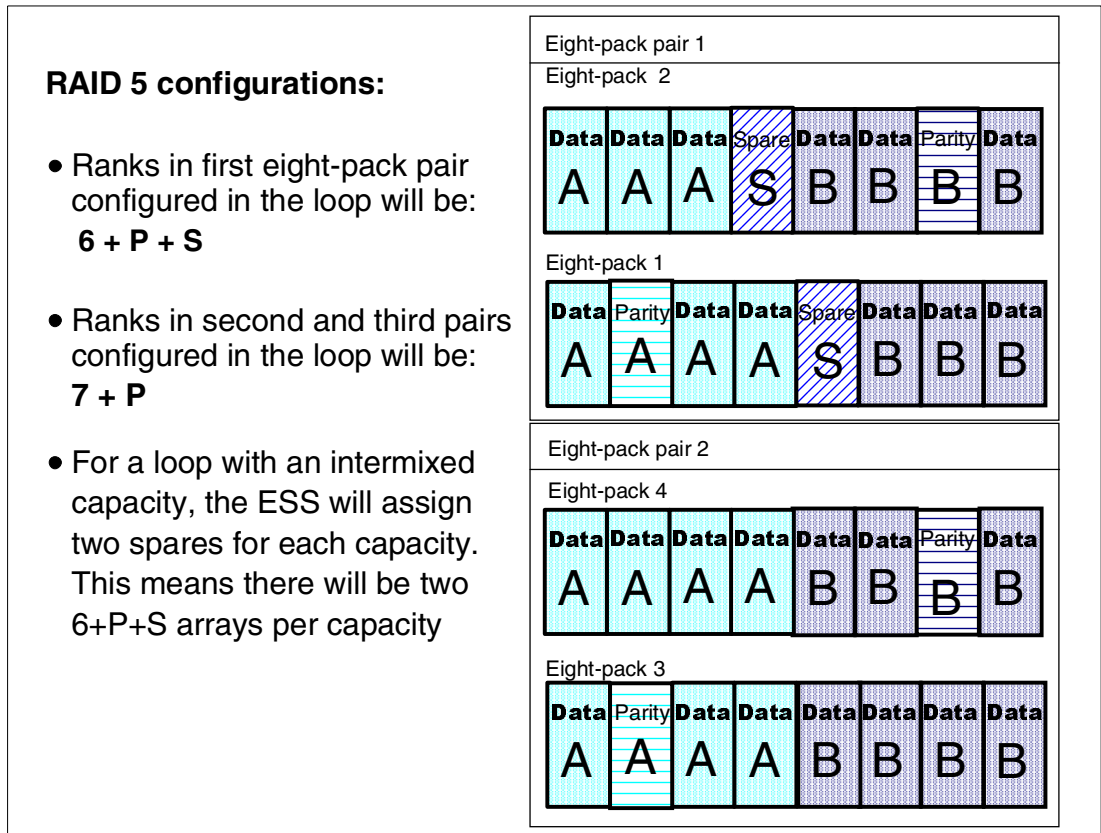


Figure 3-12 RAID 5 rank implementation

In the RAID 5 implementation, the disk access arms move independently for each disk, thus enabling multiple concurrent accesses to the array disk. This results in multiple concurrent I/O requests being satisfied, thus providing a higher transaction throughput.

RAID 5 is well suited for random access to data in small blocks. Most data transfers, reads or writes, involve only one disk and hence operations can be performed in parallel and provide a higher throughput. In addition to this efficiency for transaction oriented operations, the RAID 5 implementation of the ESS is able to work in a RAID 3 style when processing sequential operations (refer to “Sequential operations - write” on page 91) for maximum sequential throughput.

3.10.3 RAID 10 rank

The second possible RAID implementation in an ESS Model 800 rank is RAID 10 (also known as RAID 0+1). The RAID 10 rank consists of a set of disks for user data plus their mirrored disks counterparts. There is no parity disk to rebuild a failed disk. In case one disk becomes unusable, then its mirror will be used to access the data and also to build the spare.

Because the ESS architecture for maximum availability is based on two spare drives per SSA loop (and per capacity) if the first rank that is configured in a loop is defined as a RAID 10 rank, then it will be defined by the ESS with three data disks, plus their three mirrored counterpart disks, plus two spares (3 + 3 + 2S). This will happen for the first rank of each capacity installed in the loop if configured as RAID 10.

Once the two spare per capacity rule is fulfilled, then further RAID 10 ranks in the loop will be configured by the ESS as four data disks plus four mirror disks (4+4). Figure 3-13 illustrates the two arrangements of disks that are done by the ESS when configuring RAID 10 ranks.

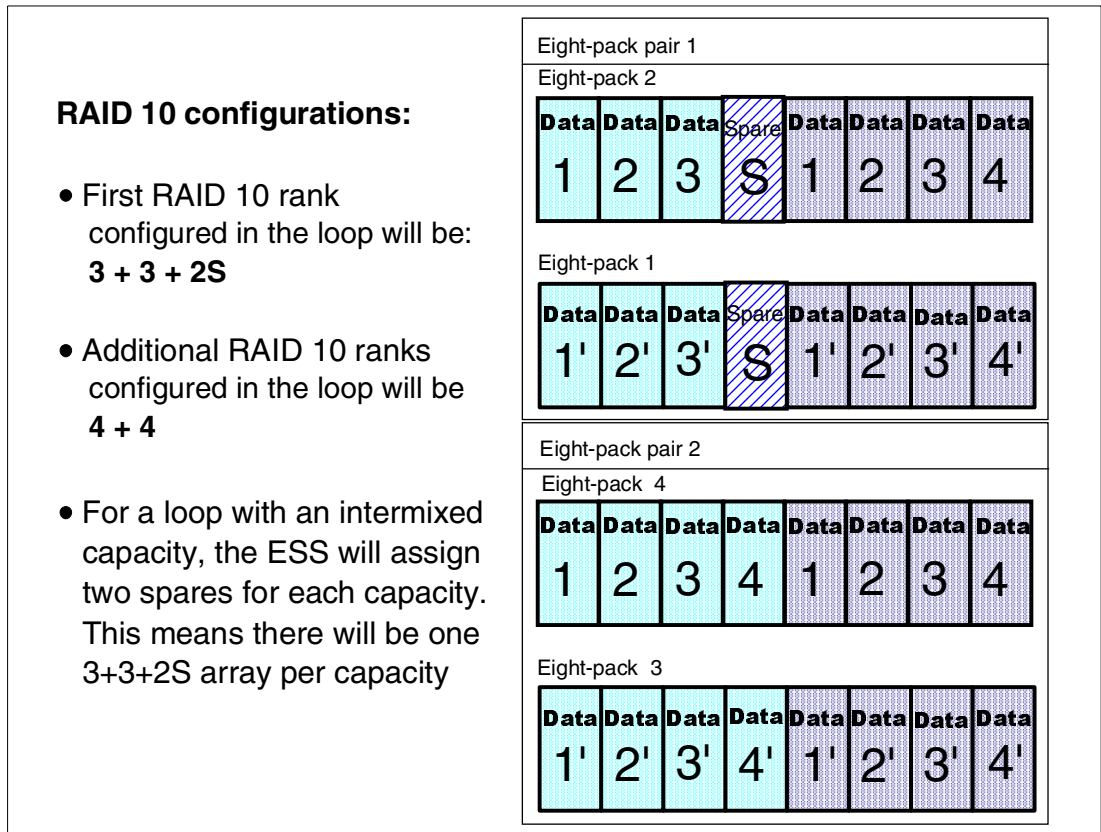


Figure 3-13 RAID 10 rank implementation

RAID 10 is also known as RAID 0+1, because it is a combination of RAID 0 (striping) and RAID 1 (mirroring). The striping optimizes the performance by striping volumes across several disk drives (in the ESS Model 800 implementation, three or four DDM's). RAID 1 is the protection against a disk failure by having a mirrored copy of each disk. By combining the two, RAID 10 provides data protection with I/O performance.

Storage balancing with RAID 10

For performance reasons you should try to allocate storage on the ESS equally balanced across both clusters and among the SSA loops. One way to accomplish this is to assign two arrays (one from loop A and one from loop B) to each Logical Subsystem (LSS, are explained in 3.12.2, "Ranks mapping" on page 69). To achieve this you can follow this procedure when configuring RAID 10 ranks:

1. Configure first array for LSS 0/loop A. This will be a 3 + 3 + 2S array.
2. Configure first array for LSS 1/loop B. This will be again a 3 + 3 + 2S array.
3. Configure second array for LSS1/loop A. This will be now a 4 + 4.
4. Configure second array for LSS 0/loop B. This will be also a 4 + 4.

Figure 3-14 on page 67 illustrates the results of this configuration procedure.

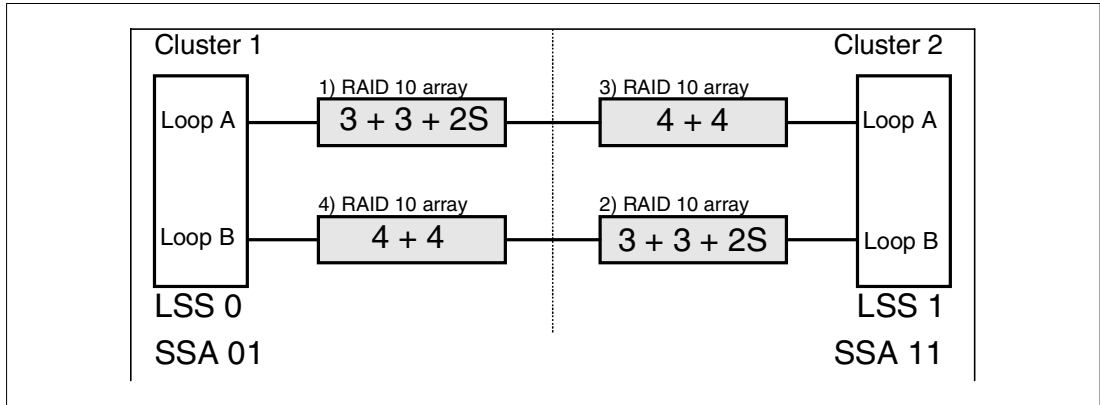


Figure 3-14 RAID 10 balanced configuration

Refer to Chapter 4, “Configuration” on page 93 for configuration recommendations.

3.10.4 Combination of RAID 5 and RAID 10 ranks

It is possible to have RAID 10 and RAID 5 ranks configured within the same loop, as illustrated in Figure 3-15. There are several ways in how a loop can be configured when mixing RAID 5 and RAID 10 ranks in it. If you configure RAID 5 and RAID 10 ranks on the same loop it is important to follow some guidelines in order to balance the capacity between the clusters. Chapter 4, “Configuration” on page 93 contains recommendations for mixed rank types configurations.

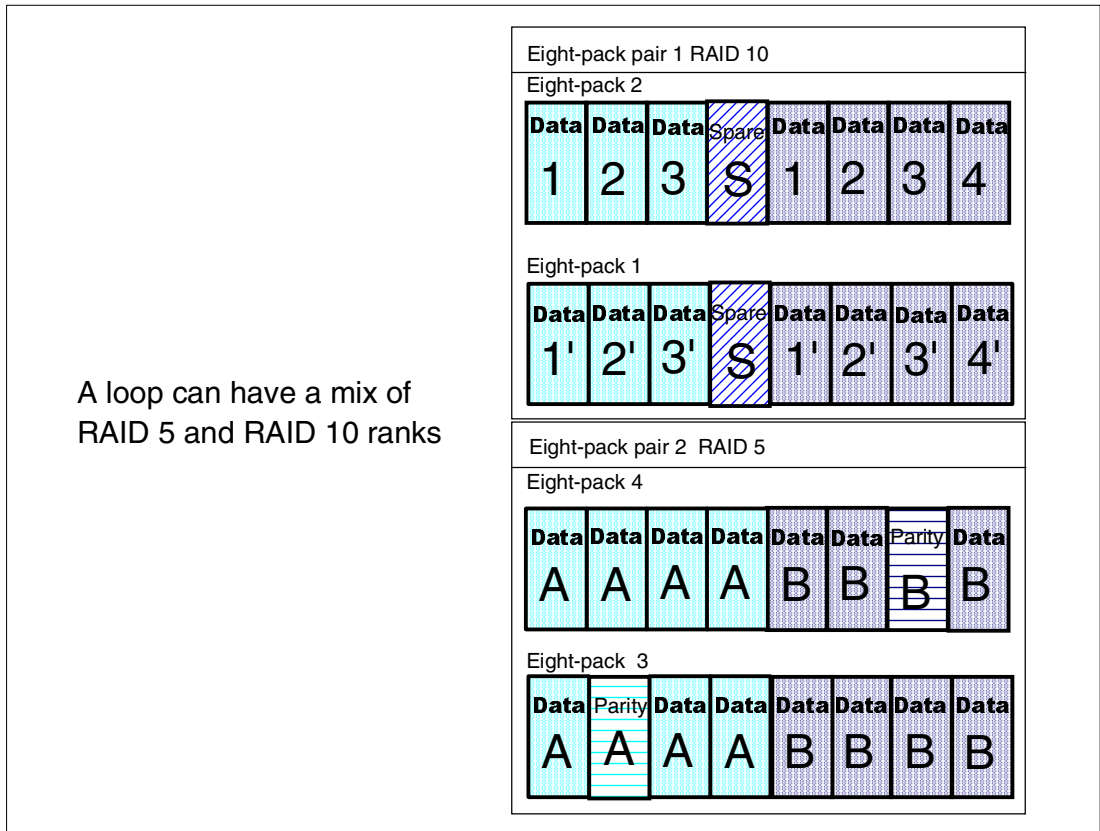


Figure 3-15 RAID 5 and RAID 10 in the same loop

3.11 SSA device adapters

The ESS Model 800 includes a new more powerful SSA device adapter to further improve the back-end disk performance. The device adapters manage two SSA loops (A and B) and perform all the RAID operations for the loops, including parity generation and striping for RAID 5, and mirroring and striping for RAID 10. This activity is done by the device adapters together with the normal reads and writes of user data, as well as the disk sparing if needed.

The SSA device adapter has on-board cache memory to hold the data and effectively off load the RAID functions from the clusters. No parity processing is done by the cluster processors or cache.

Sparing — the recovery of a failed disk drive onto one of the spare disk drives (explained in 3.7.1, “Sparing in a RAID rank” on page 59) —is also handled automatically by the SSA device adapter. The sparing process takes place in the background over a period of time, thus minimizing its impact on normal I/O operations. In fact, the sparing process dynamically adjusts its rate, in order not to impact when normal I/O processing to the array increases and subsequently taking advantage when normal I/O operations are lower. Then the failed disk drive can immediately be replaced and automatically becomes the new spare (floating spare).

The disk drives in a RAID rank are organized for performance by grouping four DDMs from two different eight-packs into a *disk group*. This allows the SSA adapter to achieve maximum throughput for each rank by having a path to each half of the rank down on each leg of the loop.

3.12 Logical Subsystems

The *Logical Storage Subsystem* (LSS, or Logical Subsystem) is a logical structure that is internal to the ESS. It is a logical construct that groups up to 256 logical volumes (logical volumes are defined during the logical configuration procedure) of the same disk format (CKD or FB) and is identified by the ESS with a unique ID. Although the LSS relates directly to the logical control unit (LCU) concept of the ESCON and FICON architectures, it does not directly relate to SCSI and FCP addressing.

The CKD Logical Subsystems are configured with the ESS Specialist at the time of configuring the Logical Control Units (LCUs) at S/390 storage allocation time. The fixed block Logical Subsystems are configured by the ESS Specialist at the time of allocating the fixed block logical volumes. In PPRC environments, Logical Subsystems are entities used for managing and establishing PPRC relationships.

3.12.1 Device adapters mapping

The device adapter (DA) to LSS mapping is a fixed relationship. Each DA supports two loops, and each loop supports two CKD Logical Subsystems and two FB Logical Subsystems (one from each cluster). So a DA pair supports up to four CKD LSSs and four FB LSSs, as Figure 3-16 on page 69 illustrates.

When all the eight loops have capacity installed, then there are up to 16 CKD LSSs and up to 16 FB LSSs available to support the maximum of 48 RAID ranks. Each LSS supports up to 256 logical devices (each logical device is mapped to a logical volume in the RAID rank).

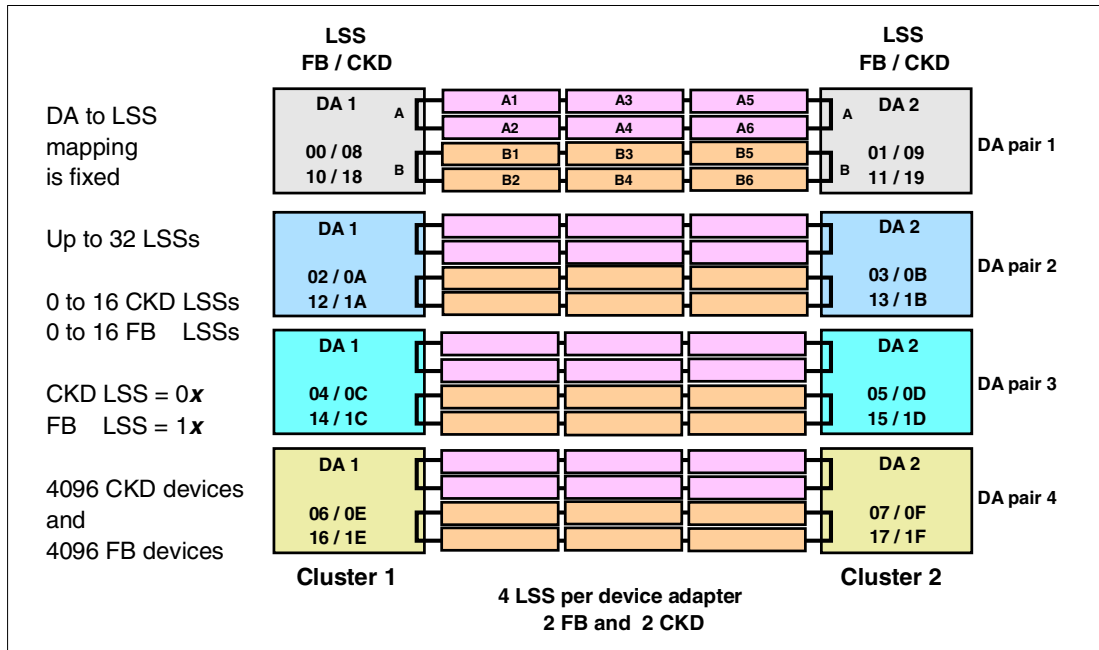


Figure 3-16 Logical Subsystems and device adapters mappings

The numbering of the Logical Subsystems indicates the type of LSS. CKD Logical Subsystems are numbered x'00' to x'0F' and the FB Logical Subsystems are numbered x'10' to x'1F'. As previously mentioned, for the CKD host view (i.e., zSeries server), a Logical Subsystem is also mapped one-to-one to a logical control unit.

As part of the configuration process, you can define the maximum number of Logical Subsystems of each type you plan to use. If you plan to use the ESS only for zSeries data, then you can set the number of FB LSSs to 0. This releases the definition space for use as cache storage. But you must also remember that going from 8 to 16 LSSs is disruptive, so you should decide in advance how many you will need.

3.12.2 Ranks mapping

This section discusses the relationship between the device adapters and the ranks (RAID 5 or RAID 10) defined in the loop.

Logical Subsystems are related to the device adapters, as presented in the previous section. All disk drives in the loop are configured as RAID 5 or RAID 10 ranks or as a combination of both.

As part of the configuration process, each rank is assigned to one LSS. This LSS is either CKD or FB. Two ranks, from two different loops of the same DA pair can be associated to the same LSS.

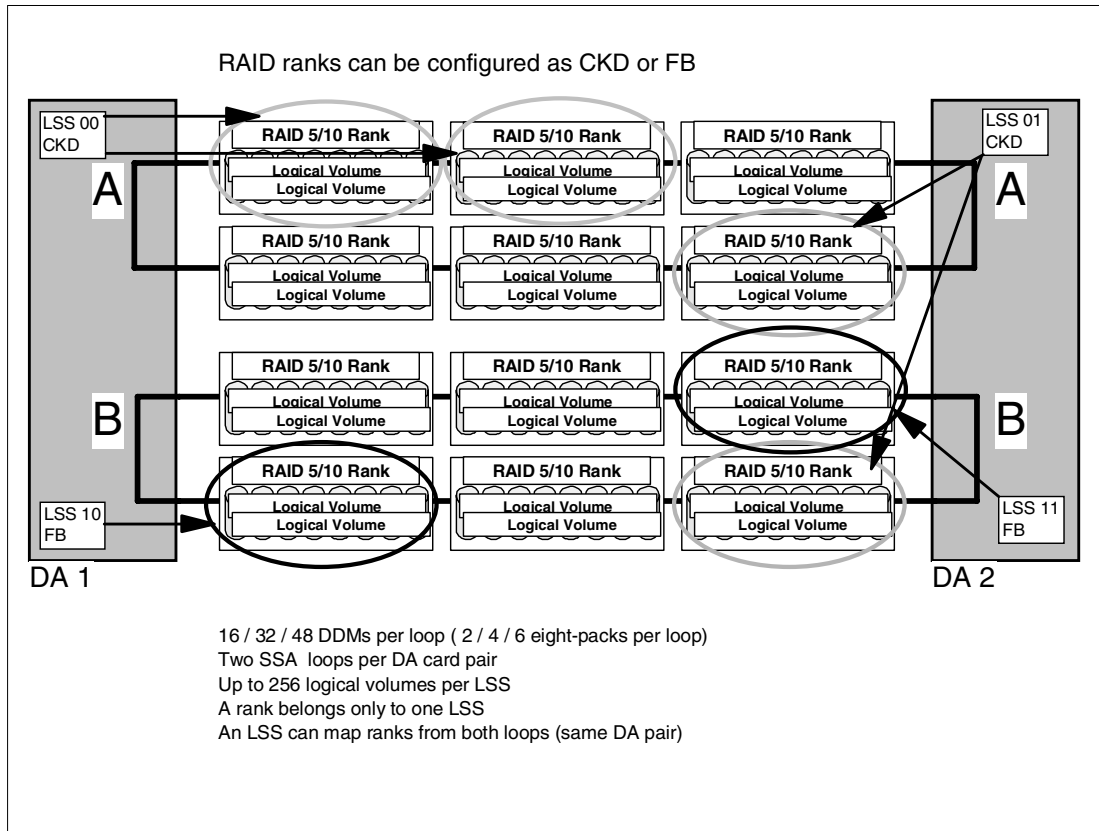


Figure 3-17 Logical Subsystem and rank relationship

Example

In the example shown in Figure 3-17, the maximum of 48 disk drives are installed on each of both loops. In the example, six disk groups are mapped onto the four LSSs of the first device adapter pair of the ESS. So six RAID ranks (RAID 5 or RAID 10) are defined, and this will result in:

1. DA1 Loop A LSS(00)—CKD: Two RAID 5 or RAID 10 ranks
2. DA1 Loop B LSS(10)—FB: One RAID 5 or RAID 10 rank.
3. DA 2 Loop B LSS(11)—FB: One RAID 5 or RAID 10 rank.
4. DA 2 Loops A and B LSS(01)—CKD: Two RAID 5 or RAID 10 ranks from two different loops.

Note there are still six remaining disk groups that can be mapped onto either existing or new LSSs.

For an LSS, allocating ranks from different loops (from the same DA pair) could be useful, especially for FlashCopy. However you must remember that all the capacity that an LSS can allocate, must be mapped within the 256 logical volume limit. 256 is the maximum number of logical volumes that can be defined for an LSS. Moreover, for CKD servers the PAV addresses should be taken in consideration.

3.13 Host mapping to Logical Subsystem

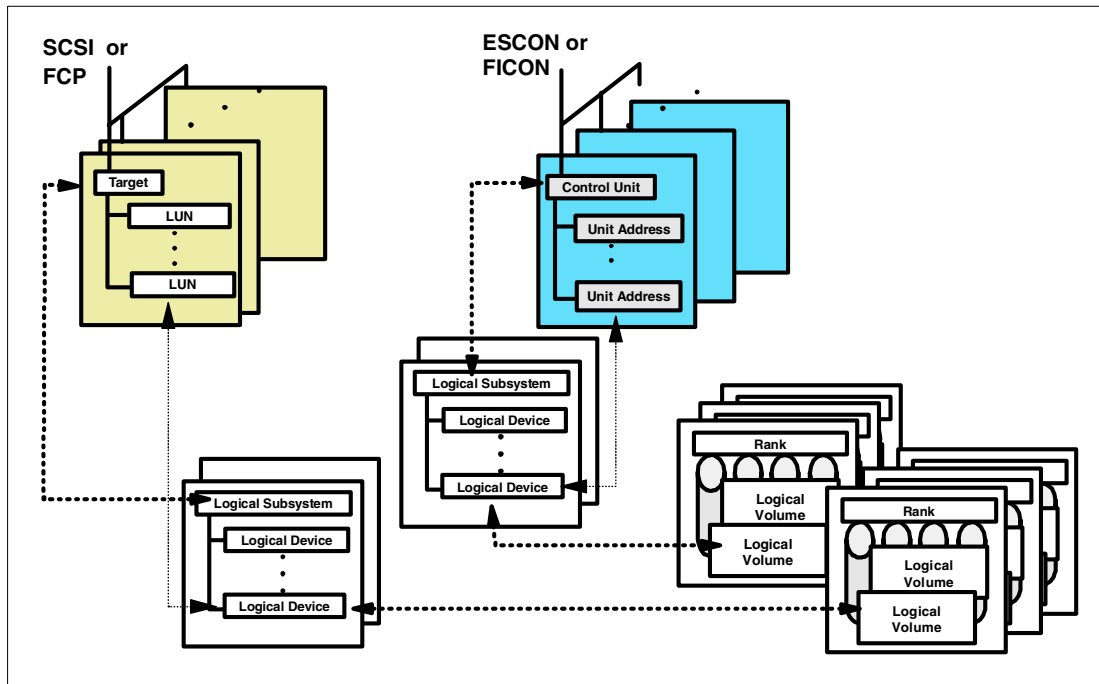


Figure 3-18 Host mapping

For the S/390 servers the data stored in the ESS is written in a count-key-data (CKD) format, in logical volumes (emulation of 3390 or 3380 volumes), and this data is read/updated with the host I/O operations. The device addressing and I/O operations to the logical volumes, for the S/390 servers, are done according to the ESCON or FICON architecture characteristics.

For the open system servers the data stored in the ESS is written in a fixed block (FB) format, in logical volumes, and this data is read/updated with the host I/O operations. The device addressing and I/O operations to the logical volumes, for the open system servers, are done according to the SCSI or Fibre Channel architecture characteristics.

Additionally, the ESS maps up to 256 of these logical volumes in internal logical constructs that are the Logical Storage Subsystems (LSS).

3.13.1 SCSI and Fibre Channel mapping

Each SCSI bus/target/LUN combination or each Fibre Channel device adapter/LUN combination is associated with one logical device (LD), each of which can be in only one Logical Subsystem. Another LUN can also be associated with the same logical device, providing the ability to share devices within systems or between systems.

3.13.2 CKD server mapping

For S/390 servers every Logical Subsystem relates directly to a logical control unit (LCU), and each logical device (LD) relates to a unit address.

Every ESCON or FICON port can address all 16 logical control units in the ESS Model 800.

3.14 Architecture characteristics

ESCON and SCSI architectures have respectively evolved to FICON and FCP architectures. Figure 3-19 lists the characteristics of these architectures.

- SCSI
 - Each parallel bus supports up to 16 targets (devices) or initiators (hosts)
 - Each target supports up to 64 logical unit numbers (LUNs)
- FCP
 - Each host N-port has only one target
 - Each target can address 16K LUNs (up to 2_{56} with Hierarchical Addressing)
- ESCON
 - Supports up to 16 logical control units (LCUs) per control unit port
 - Supports up to 1M logical volumes per channel
 - Maximum of 4K logical volumes per control unit
 - 64 logical paths per control unit port
- FICON
 - Supports up to 256 logical control units (LCU) per control unit port
 - Supports up to 16M logical volumes per channel
 - Maximum of 64K logical volumes per control unit
 - 256 logical paths per control unit port
- Common to ESCON and FICON
 - LCU has maximum of 256 logical volume addresses
 - All CKD logical devices accessed on every CU port

Figure 3-19 Architecture addressing characteristics

An architecture is a formal definition that, among other things, allows related components to interface to each other. The hardware and software components that make the systems implement architectures by following the definitions described in the official documentation. Some extensions to an architecture may not be publicly available because they are patented, and companies wanting to use them must be licensed and may have to pay a fee to access them.

Note: When looking at Figure 3-19, keep in mind that it is showing architectural characteristics. Remember that the product implementation of an architecture often limits its application to only a subset of that architecture. This is the reason we distinguish between architecture characteristics and implementation characteristics.

For detailed description of the characteristics of the FICON and the Fibre Channel architectures, and their ESS implementation you may refer to the following documents at the Web site:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essficonwp.pdf>

and

<http://www.storage.ibm.com/hardsoft/products/ess/support/essfcwp.pdf>

3.15 ESS Implementation - Fixed block

- Up to 16 FB LSSs per ESS
 - 4096 LUNs in the ESS
- SCSI
 - Host device addressing is target/LUN on each bus
 - Maximum of 15 targets per bus and 64 LUNs per target
 - Target on bus is associated with a single LSS
 - LUNs are associated to a specific logical volume on LSS
 - A specific logical volume can be shared using a different target/LUN association
 - LUN masking
- FCP
 - One target per host N-port
 - Must first define host N-port worldwide port name (WWPN)
 - Configure
 - Either 256 LUNs per host N-port
 - Or 4096 LUNs for the whole ESS and use LUNs enable mask for each host N-port
 - Access control by host by port

Figure 3-20 FB implementation for SCSI and FCP on the ESS

Servers attaching to the ESS by means of SCSI or Fibre Channel connections use the fixed block implementation of the ESS (refer to Figure 3-20).

3.15.1 SCSI mapping

In SCSI attachment, each SCSI bus can attach a combined total of 16 initiators and targets. Since at least one of these attachments must be a host initiator, that leaves a maximum of 15 that can be targets. The ESS presents all 15 targets to its SCSI ports. Also, in SCSI attachment, each target can support up to 64 LUNs. The software in many hosts is only capable of supporting 8 or 32 LUNs per target, even though the architecture allows for 64. Since the ESS supports 64 LUNs per target, it can support $15 \times 64 = 960$ LUNs per SCSI port. Each bus/target/ LUN combination is associated with one logical volume, each of which will be mapped in only one LSS. Another target/LUN can also be associated with the same logical volume, providing the ability to share devices within systems or between systems.

SCSI LUNs

For SCSI, a target ID and all of its LUNs are assigned to one LSS in one cluster. Other target IDs from the same host can be assigned to the same or different FB LSSs. The host adapter directs the I/O to the cluster with the LSS that has the SCSI target defined during the configuration process.

LUN affinity

With SCSI attachment, ESS LUNs have an affinity to the ESS SCSI ports, independent of which hosts may be attached to the ports. Therefore, if multiple hosts are attached to a single

SCSI port (The ESS supports up to four hosts per port), then each host will have exactly the same access to all the LUNs available on that port. When the intent is to configure some LUNs to some hosts and other LUNs to other hosts, so that each host is able to access only the LUNs that have been configured to it, then the hosts must be attached to separate SCSI ports. Those LUNs configured to a particular SCSI port are seen by all the hosts attached to that port. The remaining LUNs are “masked” from that port. This is referred to as *LUN masking*.

3.15.2 FCP mapping

In Fibre Channel attachment, each Fibre Channel host adapter can architecturally attach up to 256 LUNs in hierarchical mode. If the software in the Fibre Channel host supports the SCSI command **Report LUNs**, then it will support 4096 LUNs per adapter; otherwise it only supports 256 LUNs per adapter. The hosts that support Report LUNs are the AIX, OS/400, HP, and NUMA-Q DYNIX/ptx-based servers; hence they will support up to 4096 LUNs per adapter. All other host types, including NUMA-Q NT-based, will only support 256 LUNs per adapter.

The maximum number of logical volumes you can associate with any LSS is still 256, and the ESS has 16 FB LSSs. Therefore, the maximum LUNs supported by ESS, across all of its Fibre Channel and SCSI ports, is $16 \times 256 = 4096$ LUNs.

LUN affinity

In Fibre Channel attachment, LUNs have an affinity to the host's Fibre Channel adapter via the adapter's worldwide unique identifier (the worldwide port name, WWPN), independent of which ESS Fibre Channel port the host is attached to. Therefore, in a switched fabric configuration where a single Fibre Channel host can have access to multiple Fibre Channel ports on the ESS, the set of LUNs that may be accessed by the Fibre Channel host are the same on each of the ESS ports.

Access control by host by port is a function that enables the user to restrict a host's access to one or more ports rather than always allowing access to all ports. This capability significantly increases the configurability of the ESS.

LUN access modes

In Fibre Channel attachment, ESS provides an additional level of access security via either the `Access_Any` mode or the `Access_Restricted` mode. This is set by the IBM System Support Representative and applies to all Fibre Channel host attachments on the ESS.

In `Access_Any` mode, any host's Fibre Channel adapter for which there has been no access-profile defined can access all LUNs in the ESS (or the first 256 LUNs in the ESS if the host does not have the Report LUNs capability). In `Access_Restricted` mode, any host's Fibre Channel adapter for which there has been no access-profile defined can access none of the LUNs in the ESS. In either access mode, a host's Fibre Channel adapter with an access-profile can see exactly those LUNs defined in the profile, and no others.

You can find a detailed description of LUN affinity and LUN Access modes in the document at:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essfcwp.pdf>

3.16 ESS Implementation - CKD

- ESS implementations common to ESCON and FICON
 - LSS to logical control unit (LCU) 1:1 relationship
 - Up to 16 LCUs per ESS
 - Maximum 256 unit addresses per LCU
 - Maximum of 4096 unit addresses per ESS
- Specific to ESCON
 - Up to 1,024 logical volumes per channel
 - 64 CU logical paths per ESS ESCON port
- Specific to FICON
 - Up to 16,384 logical volumes per channel
 - 256 CU logical paths per ESS FICON port
 - 4096 CU logical paths per ESS (256 per LSS)

Figure 3-21 CKD implementation for ESCON and FICON on the ESS

For Count-Key-Data (CKD)-based servers such as the S/390 servers, every LSS relates directly to one logical control unit (LCU), and each logical volume (LV) to one host unit address (UA).

Every host channel adapter port effectively addresses all 16 logical control units in the ESS, similarly for both ESCON and FICON.

The maximum number of LVs you can associate with any LSS is 256, and the ESS supports up to 16 CKD type LSSs. Therefore, the maximum number of unit addresses supported by the ESS, both for ESCON or FICON attachment, is $16 \times 256 = 4096$ addresses.

FICON specifics

FICON architecture allows 256 LCUs per control unit port, but the ESS implementation is 16 LCUs per control unit (similar to ESCON). FICON implementation in the ESS also allows 16,384 unit address per channel, and 256 control unit logical paths per port. The ESS FICON implementation supports a total maximum of 4096 logical paths per ESS (256 per LSS).

The expanded addressing and performance characteristics of FICON allow for simpler configurations where logical daisy chaining of LSSs can be further exploited.

For more detailed description of the FICON architectural characteristics and its ESS implementation, refer to the document in the Web at:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essficonwp.pdf>

3.17 CKD server view of ESS

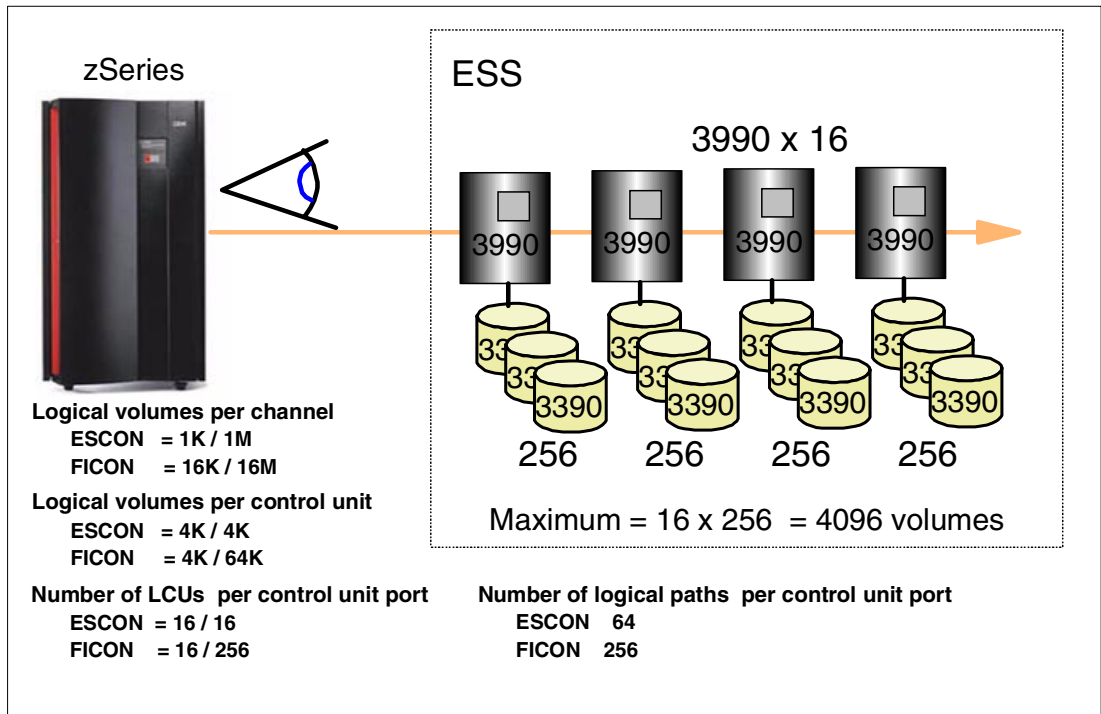


Figure 3-22 CKD server view of the ESS

For CKD servers, both ESCON and FICON attached, an ESS looks like multiple 3990-6 Storage Controls, each with up to 256 volumes. Up to 16 of the 3990s may be defined through HCD using the CUADD parameter to address each LCU. Each LCU is mapped directly to the CKD LSS number. So LSS 0 is mapped to LCU 0 and CUADD 0, and so on for all 16 CKD LSSs.

ESCON-attached CKD server view of ESS

CKD servers support 256 devices per LCU. Every LSS (and therefore LCU) can be addressed by every ESCON link. This means that, in theory, an ESCON channel could see all 16 LCUs, each with 256 devices (a maximum of 4096 devices). However, the ESCON channel hardware implementation limits the number of devices that can be addressed per channel to 1024. This is unlikely to be a restriction for most customers.

FICON-attached CKD server view of ESS

For FICON with ESS, each FICON channel also sees all 16 LCUs of the ESS, each with 256 devices (a maximum of 4096 devices). But the FICON hardware implementation extends the number of devices that can be addressed over a channel to 16,384. This expands the capabilities of logical daisy chaining when using FICON channels, allowing more flexibility at the time of laying out your configuration.

For more detailed description of the FICON architectural characteristics and its ESS implementation, refer to the document in the Web at:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essficonwp.pdf>

3.18 CKD Logical Subsystem

0 to 16 logical control unit images per ESS

- Up to 256 devices per CU image
- 4096 logical volumes maximum
- 1:1 Mapping between LCU and LSS

Emulation of 9390/3990-6, 3990-3, 3990-3+TPF

- 3390 2,3, and 9 emulation
- 3380 track format with 3390 capacity volumes
- Variable size 3390 & 3380 volumes (custom volumes) and 32k cylinder large volumes

Figure 3-23 CKD LSS

When configuring the ESS, you can specify whether you want 0, 8 or, 16 logical control units (LCUs) defined. If, for example, you plan to use an ESS for FB type data only, setting the CKD LSS number to zero frees up storage for use as cache.

The ESS emulates the 9390/3990-6, the 3990-3 and the 3990-3 with TPF LIC. The operating system will recognize the ESS as a 2105 device type when you have the appropriate maintenance applied to your system.

Devices emulated include standard 3390 Model 2, 3, 9 volumes. You can also define custom volumes, volumes whose size varies from a few cylinders to as large as a 3390 Model 9 (or as large as 32K cylinders with the DFSMS Large Volume Support). The selection of the model to be emulated is part of the ESS Specialist configuration process.

The ESS also supports 3380 track format, in a similar way to 3390 Track Compatibility Mode. A 3380 is mapped onto a 3390 volume capacity. So the 3380 track mode devices will have 2226 cylinders on a volume defined with the capacity of a 3390-2, or 3339 cylinders on a volume of 3390-3 size. If you wanted to have volumes that were exactly the same, for example, as a 3380-K, then you could use the custom volume function and define your logical volumes with exactly the same number of cylinders as a 3380-K.

3.19 SCSI server view of ESS

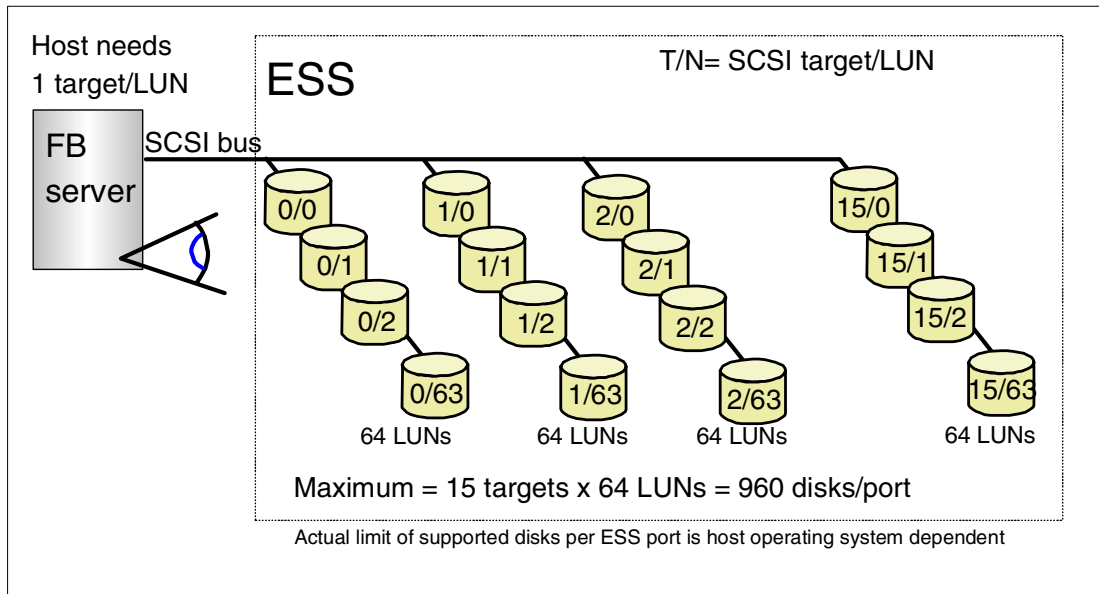


Figure 3-24 SCSI server view of the ESS

If you imagine a SCSI host's view of the ESS, it looks like a bunch of SCSI disks attached to a SCSI bus. The actual number that any fixed block SCSI-attached server system can support is considerably less than the maximum shown in Figure 3-24.

One target/LUN is used for each host attached to the ESS SCSI bus, and is commonly designated as the initiator ID. Typically, you will only have one host per SCSI bus that is attached to one ESS port, leaving you with 15 target IDs and a number of LUNs per target that varies, depending on the host system's LUN support. Today, this operating system support can range from four to 64 LUNs per target ID.

3.20 FB Logical Subsystem - SCSI attachment

0 to 16 logical subsystems per ESS

- Up to 256 FB logical devices per LSS
- Up to 4096 FB logical devices per ESS

0-32 SCSI ports per ESS

- 1-15 targets per SCSI bus/ESS port
- 1-64 LUNs per target (SCSI-3 architecture)
- Maximum 960 LUNs per SCSI bus/ESS port
- Up to 4 initiators per SCSI bus/ESS port

Figure 3-25 FB LSS - SCSI attachment

When configuring the ESS, you can specify the maximum number of FB type LSSs you plan to use. If you only have fixed block servers connecting to your ESS, then you can set the number of CKD type LSSs to zero.

Each FB type LSS supports up to 256 logical volumes. The size of the logical volumes vary from 100 MB to the maximum effective capacity of the rank (refer to Table 2-1 on page 27 for ranks' effective capacities). A single FB Logical Subsystem can contain logical volumes from multiple SCSI hosts.

0/8/16 LSSs

Either 8 or 16 LSSs can be specified when configuring the ESS. The choice of whether to use 8 or 16 LSSs will be influenced by whether or not you intend to use FlashCopy, and by how many volumes you wish to attach. If 16 LSSs are specified, then it will be possible to create up to 4096 logical volumes. But in most cases 2048 logical volumes in the ESS are more than sufficient for the open systems server you may plan to attach. This is because it is possible (and often desirable) to create a smaller number of large or very large volumes. For example, it is possible to create a single 509 GB volume for Windows NT (with 72.8 GB disks on a 7+P array) which can also help overcome the limited amount of drive letters available to Windows NT.

The more important consideration then is whether you intend to use FlashCopy. FlashCopy is only possible within an LSS and if you choose eight LSSs, it will be much easier to manage existing and/or future FlashCopy requirements than with 16 LSSs. For example, if your ESS had eight LSSs specified, with disk capacity in each LSS, then any additional arrays will be added to an existing LSS. If the existing LSS capacity was fully utilized, then the additional arrays could be easily used for FlashCopy. If your ESS had 16 LSSs, then additional arrays might cause a new LSS to be used. This is because although you specified 16 LSSs, the number of LSSs actually in use is dependent upon the installed capacity of the ESS and the number of logical volumes created within the ESS. If new LSSs are used because new arrays are installed, then it will be necessary to move some of your data to those LSSs in order to FlashCopy it. For this reason, it is generally best to configure eight LSSs in the ESS.

0-32 SCSI ports per ESS

You can install the SCSI host adapters into any of the ESS host adapter bays. Each SCSI card contains two SCSI ports. For a SCSI-only ESS, you can fill all the host adapter bays with SCSI cards, giving you a maximum of 16 cards and 32 SCSI ports.

Each SCSI port supports the SCSI-3 standard: 16 target SCSI IDs with 64 LUNs per target. This gives a total of $15 \times 64 = 960$ logical volumes on one SCSI port (only 15 because the host uses one SCSI ID).

You can attach up to four hosts to each ESS SCSI port (daisy chain). See 4.28, “Defining FB logical devices” on page 130 for details on the number of LUNs supported by different systems.

3.21 Fibre Channel server view of ESS

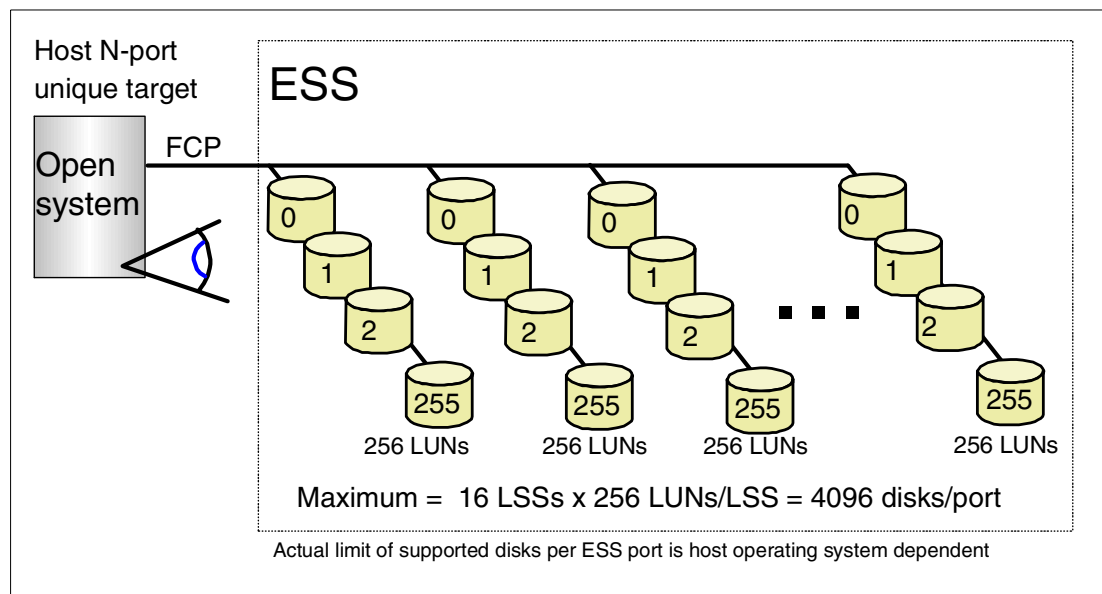


Figure 3-26 Fibre Channel server view of ESS

In Fibre Channel attachment, each Fibre Channel host adapter can architecturally attach up to 256 LUNs. If the software in the Fibre Channel host supports the SCSI command **Report LUNs**, then it will support 4096 LUNs per adapter.

In the ESS, each LSS supports a maximum of 256 LUNs, and there are 16 LSSs in an ESS. Therefore, the maximum LUNs supported by ESS, across all its SCSI and FCP ports, is $16 \times 256 = 4096$ LUNs.

For Fibre Channel architecture, where LUNs have affinity to the host's Fibre Channel adapter (via the world wide port name, WWPN), any fibre channel initiator can access any LUN in the ESS. This way each port in the ESS is capable of addressing all the 4096 LUNs in the ESS (should the host software not set a lower limit addressing for the adapter). This 4096 addressing may not be always desirable, so there are ways to limit it. Fibre Channel characteristics are further described in detail in the document at:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essfcwp.pdf>

3.22 FB Logical Subsystem - Fibre Channel attachment

- 0 to 16 logical subsystems per ESS
 - Up to 256 FB logical devices per LSS
 - Up to 4096 FB logical devices per ESS
- 0-16 FCP ports per ESS
 - One target per host N-port
 - Either 256 LUNs (server software sets limit)
 - Or 4096 LUNs (if server supports "Reports LUNs" command)
 - LUN affinity through switch fabric via worldwide unique identifier (WWPN)
 - LUN Access mode
 - Access any
 - Access restricted

Figure 3-27 FB LSS - Fibre Channel attachment

Fixed block LSSs are common for servers attaching either with SCSI connections or Fibre Channel connections. So the implementation characteristics are the same as already described for SCSI attachment in 3.20, "FB Logical Subsystem - SCSI attachment" on page 79.

0-16 Fibre Channel ports per ESS

Each Fibre Channel/FICON host adapter card has one port. For a Fibre Channel-only ESS, you can fill all the host adapter bays with Fibre Channel/FICON host adapter cards, giving you a maximum of 16 FCP ports.

Note: The Fibre Channel/FICON host adapters support FICON and FCP, but not both simultaneously. The protocol to be used is configurable on an adapter-by-adapter basis.

Each Fibre Channel port can address all the maximum 4096 LUNs that can be defined in the ESS. For hosts that support the **Report LUNs** command, this ESS maximum can be addressed by one Fibre Channel adapter (other hosts have a software-imposed limit of 256 per adapter). When you configure the LSS, you have the option to control the LUNs that a server can access through a specified port by means of the `Access_Any` or `Access_Restricted` modes of configuring it.

Fibre Channel characteristics are further described in detail in the document at:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essfcwp.pdf>

3.23 iSeries

Generally speaking, all the architectural considerations presented so far for the open systems servers, whether SCSI or Fibre Channel connected, relate to the iSeries family of IBM servers. With the Fibre Channel Disk adapter #2766, the iSeries Models 820, 830, 840 and 270 running OS/400 Version 5.1 can attach via Fibre Channel to the ESS. Also with the #6501 disk adapter, the other models of AS/400 running OS/400 Version 4.5 or earlier can connect SCSI to the ESS.

All the considerations about host mapping of the LSS, SCSI addressing, Fibre Channel addressing, fixed block implementation, SCSI server view of the ESS, and FCP server view of the ESS include the iSeries and AS/400 servers. However, there are some specific considerations to bear in mind when attaching iSeries or AS/400 server to the ESS, which we discuss in the present section and which are covered in greater detail in the publication *IBM @server iSeries in Storage Area Networks A Guide to Implementing FC Disk and Tape with iSeries*, SG24-6220.

When attaching an iSeries to an ESS, the iSeries will be like any other FB server, whether SCSI or Fibre Channel connected. Basically the ESS presents a set of LUNs to the iSeries, like it does for any other open systems server. The thing that distinguishes the iSeries is the LUN sizes it will be using: 4.19, 8.59, 17.54, 35.16, 36.00, and 70.56 GB. This way, with a SCSI adapter (#6501) attachment, the LUNs will report into the iSeries as the different models of device type 9337. And with a Fibre Channel disk adapter (#2766) connection, the LUNs will report into the iSeries as the different models of the 2105 device type. The models will depend on the size of LUNs that have been configured (see 4.26.3, “Assigning iSeries logical volumes” on page 128).

For Fibre Channel attachment, the ESS host adapter has to be set to Fibre Channel Arbitrated Loop and must be dedicated to Quickloop devices only. It cannot be shared with any other platforms unless they also use Quickloop. In addition, a maximum of 32 iSeries LUNs can be defined per ESS port.

For SCSI attachment, the iSeries #6501 adapter will only support eight LUNs per SCSI port, meaning that a maximum of 256 LUNs (32 maximum SCSI ports per ESS, times eight maximum LUNs per #6501 adapter) is possible on an ESS with all SCSI ports configured.

The other distinguishing characteristics of the iSeries come on an upper layer on top of the preceding architectural characteristics described so far for the FB servers. This is the Single Level Storage concept that the iSeries uses. This is a powerful characteristic that makes the iSeries and AS/400 a unique server.

To learn more about the iSeries storage architecture characteristics and how they complement to the ESS own characteristics, you may refer to the redbook *IBM @server iSeries in Storage Area Networks: A Guide to Implementing FC Disk and Tape with iSeries*, SG24-6220.

3.23.1 Single level storage

Both the main memory of the iSeries and the physical disk units are treated as a very large virtual address space, known as *single level storage*. This is probably the most significant differentiation of the iSeries when compared to other open systems. As far as applications on the iSeries are concerned, there is really no such thing as a disk unit.

3.23.2 iSeries storage management

The iSeries is able to add physical disk storage, and the new disk storage is automatically treated as an extension of the virtual address space. Thus the data is automatically spread across the entire virtual address space. This means that data striping and load balancing is already automated by the iSeries.

The iSeries keeps the objects in a single address space. The operating system maps portions of this address space as need arises, either to disk units for permanent storage, or to main memory for manipulation by the applications.

Storage management on the iSeries and AS/400 systems is automated. The iSeries system selects the physical disk (DASD - Direct Access Storage Device) to store data, spreads the data across the DASDs, and continues to add records to files until specified threshold levels are reached.

3.24 Data flow - host adapters

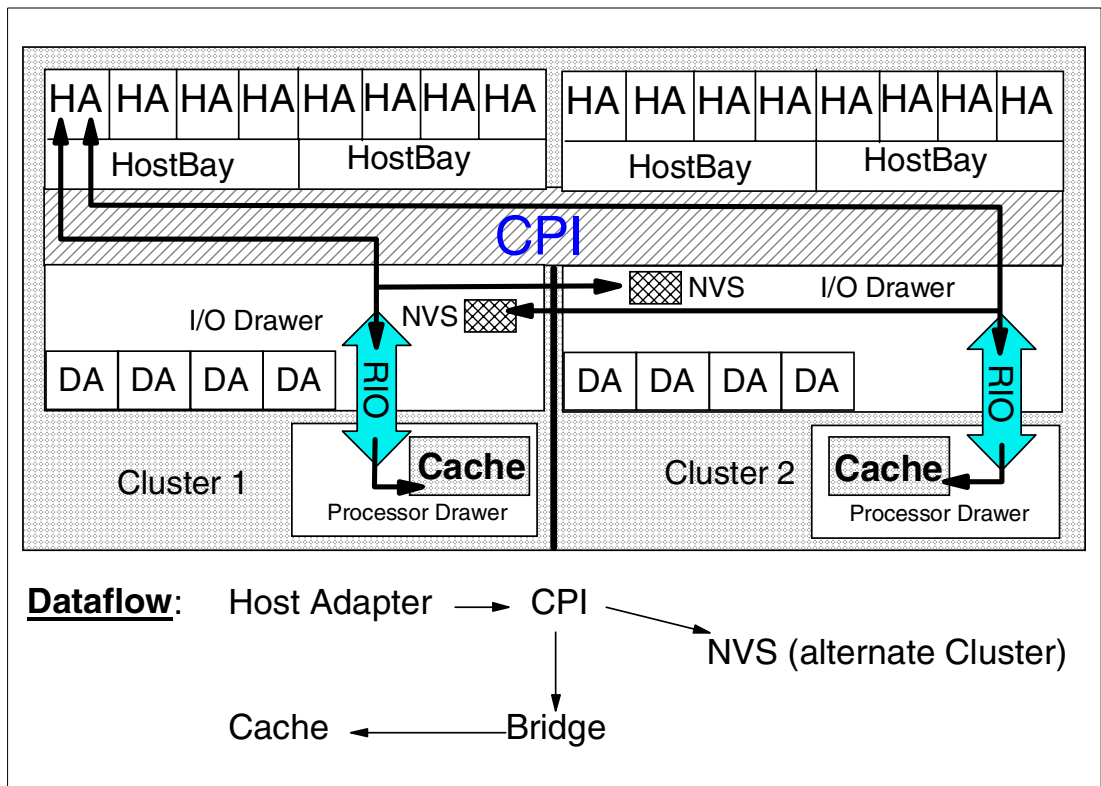


Figure 3-28 Data flow - host adapters

The host adapters (HA) are the external interfaces of the IBM TotalStorage Enterprise Storage Server Model 800 for server attachment and SAN integration. Each provides two ESCON or SCSI ports, or one FICON or Fibre Channel port. Each HA plugs into a bus in a host bay, and the bus is connected via the CPI to both clusters (Figure 3-28).

The host adapters direct the I/O to the correct cluster, based upon the defined configuration for that adapter port. Data received by the host adapter is transferred via the CPI to NVS and cache.

For an ESCON or FICON port, the connection to both clusters is an active one, allowing I/O operations to logical devices in the CKD LCUs in either cluster. The LCUs map directly to the ESS Logical Subsystems, each LSS being related to a specific SSA loop and cluster.

For SCSI, a target ID and its LUNs are assigned to one LSS in one cluster. Other target IDs from the same host can be assigned to the same or different FB LSSs. The host adapter in the ESS directs the I/O to the cluster with the LSS that has the SCSI target defined during the configuration process.

For FCP, any Fibre Channel initiator can access any LUN in the ESS. The LUNs can be associated in a LUN class that is related to the initiator WWPN (for Access_Restricted mode). Then the HA in the ESS directs the I/O to the cluster that owns the LSS that maps the LUN class that has been accessed. There are alternatives in the relationship between the ESS host adapter and the LUNs it will finally access. You can find a detailed description of this characteristics in the document at:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essfcwp.pdf>

Failover

The advantage of having both clusters actively connected to each host adapter is that, in the case of a failure in one cluster, all I/Os will automatically be directed to the remaining cluster. See 3.8, "Cluster operation: failover/failback" on page 60.

3.25 Data flow - read

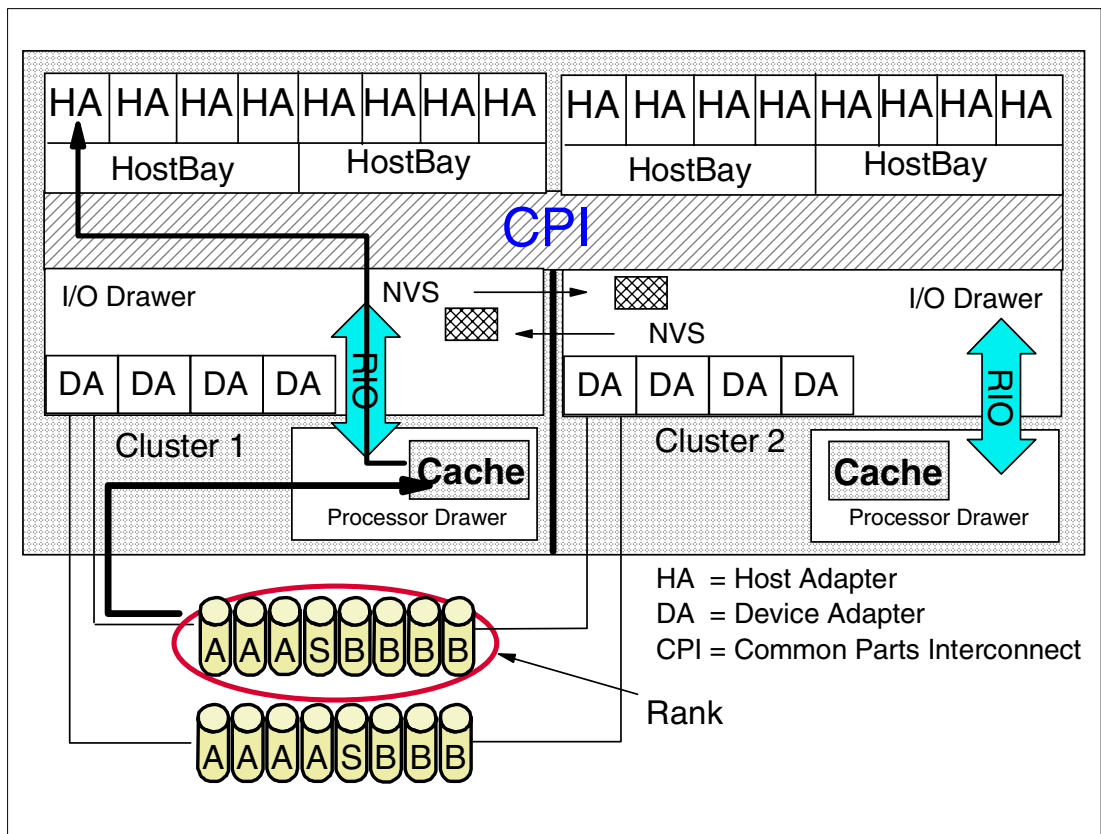


Figure 3-29 Data flow - read

Figure 3-29 on page 84 shows the data flow for a read operation. When a host requests data from the ESS, this data will be sent to the host via a host adapter from the cache, either after reading it from the cache or after reading it from the disk array into the cache (stage). Reads will be satisfied from cache if the requested data is already present from a previous operation (read hit).

3.25.1 Host adapter

The host adapter (HA) accepts the commands from the host and directs them to the appropriate cluster. For ESCON and FICON, each LCU is mapped to one LSS in one cluster, so the command can be directed to the correct cluster. For SCSI, each target is mapped to an LSS in either cluster, so part of the configuration process is to provide the HA with the SCSI target to LSS mapping. For FCP, any Fibre Channel adapter initiator can access any open systems device, so part of the configuration process is to provide the HA with the LUN class association to the WWPN of the server adapter (Access_Restricted mode).

3.25.2 Processor

The processor in the cluster's processor drawer processes the commands. If the data is in cache, then the cache-to-host transfer takes place. If the data is not in the cache, then a staging request is sent to the device adapter (DA) in the I/O drawer to fetch the requested data.

3.25.3 Device adapter

The device adapter (DA) is the SSA adapter for the loop that requests the data blocks from the disk drives in the rank. SSA can multiplex read operations, thus allowing multiple requests for searching and reading data on the disk drives to be started at the same time. The SSA adapter has buffers that it uses for holding recently used data, as well as for the RAID 5 and RAID 10 related activities.

3.25.4 Disk drives

The disk drives in the rank will read the requested data into their buffers and continue to read the rest of that track and the following tracks into a 64 KB buffer contained in the adapter. Once in the buffer, data can be transferred to the DA and the cache. Subsequent reads of data from the same track will find it already in the disk drive buffer, and it will be transferred without seek or latency delays.

3.26 Data flow - write

Figure 3-30 on page 86 illustrates the data flow for a write operation. The information in this section is complemented with information found in 3.28, "NVS and write operations" on page 88.

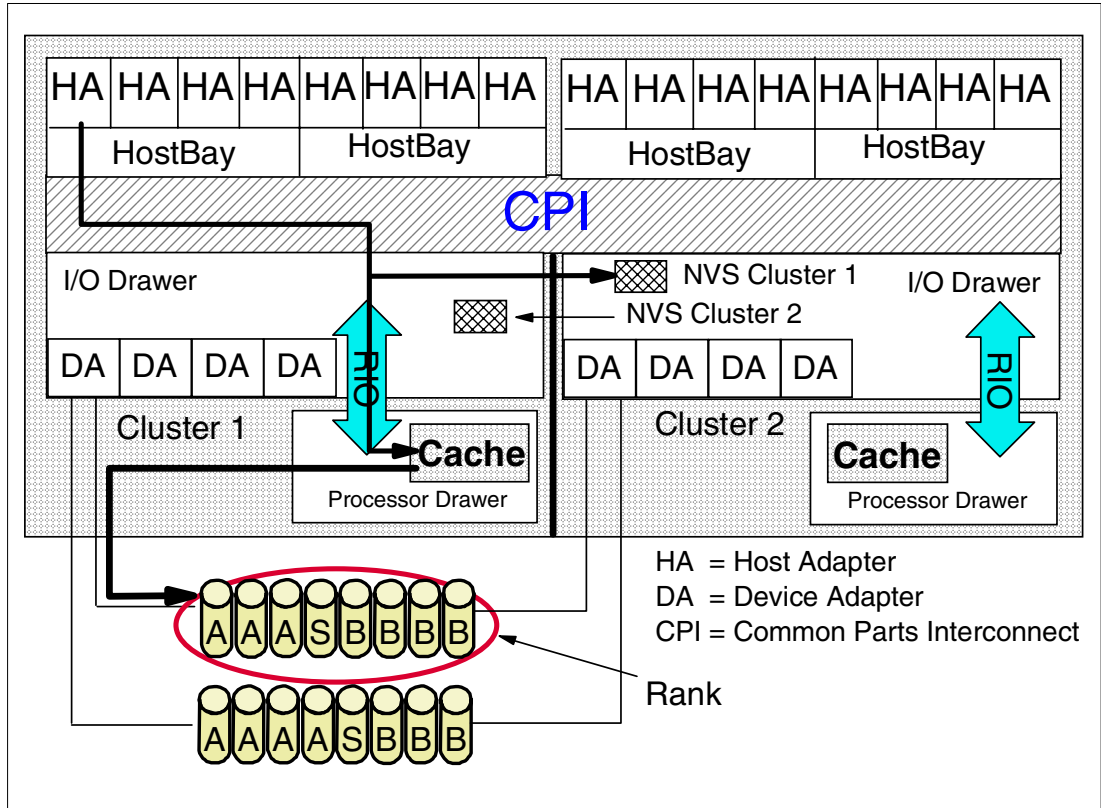


Figure 3-30 Data flow - write

3.26.1 Host adapters

The host adapter (HA) accepts the commands from the host and routes them to the correct cluster. For most write operations, data is already resident in cache from a previous operation, so the update is written to the NVS and cache (write hit). When data is in cache and NVS, the host gets channel end and device end and the write is complete.

3.26.2 Processor

The cache will hold a copy of the data until the Least Recently Used (LRU) algorithm of the cache (or NVS) determines that space is needed. At this moment the data is physically written to the array in a background operation (destage). All modified data for the same track is sent to the device adapter at the same time to maximize the destage efficiency.

3.26.3 Device adapter

The SSA device adapters (DA) manage the operation of the two loops. The SSA adapters also manage the RAID 5 and RAID 10 operations. For example for a RAID 5 rank when an update write to several blocks on a track is done, the data track and the parity must first be read into the SSA adapter RAM, then the updates are done, the parity re-calculated and then the data and new parity written back to the two disks. All this RAID related activity is done by the SSA adapter. Similarly with RAID 10, it is the SSA device adapter who manages its operation upon the disk array.

3.27 Cache and read operations

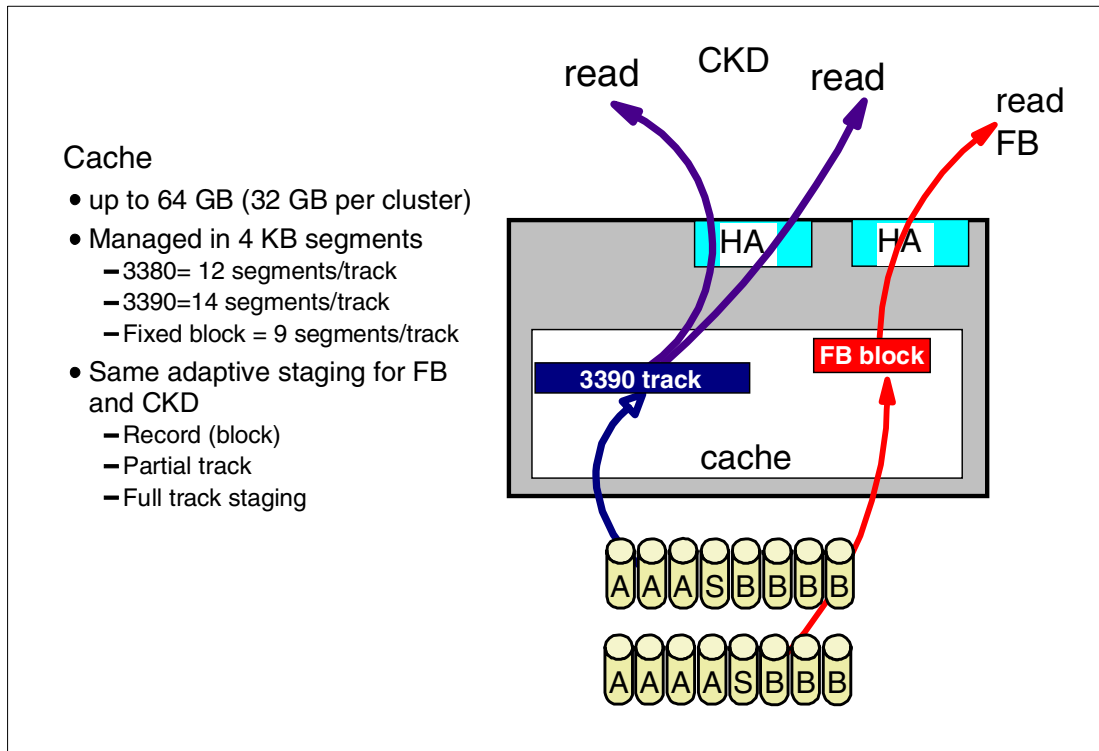


Figure 3-31 Cache - read

The cache in the ESS is split between the clusters and is not shared. As Figure 3-31 shows, each cluster has up to 32 GB of cache. The cache is managed in 4 KB segments (for fixed block a track is up to 9 segments, for CKD a full track of data in 3380 track format takes 12 segments and a full track in 3390 track format takes 14 segments). The small size allows efficient utilization of the cache, even with small records and blocks operating in record mode.

A read operation sent to the cluster results in:

- ▶ A cache hit if the requested data resides in the cache. In this case the I/O operation will not disconnect from the channel/bus until the read is complete. Highest performance is achieved from read hits.
- ▶ A cache miss occurs if the data is not in the cache. The I/O is logically disconnected from the host, allowing other I/Os to take place over the same interface, and a stage operation from the RAID rank takes place. The stage operation can be one of three types:
 - Record or block staging
 - Only the requested record or blocks are staged into the cache.
 - Partial track staging
 - All records or blocks on the same track until the end of the track are staged.
 - The entire track is staged into the cache.

The method selected by the ESS to stage data is determined by the data access patterns. Statistics are held in the ESS on each zone. A zone is a contiguous area of 128 cylinders or 1920 32-KB tracks. The statistics gathered on each zone determine which of the three cache operations is used for a specific track.

- ▶ Data accessed randomly will tend to use the record access or block mode of staging.

- ▶ Data that is accessed normally with some locality of reference will use partial track mode staging. This is the default mode.
- ▶ Data that is not a regular format, or where the history of access indicates that a full stage is required, will set the full track mode.
- ▶ The adaptive caching mode data is stored on disk and is reloaded at IML

Sequential reads

Cache space is released according to Least Recently Used (LRU) algorithms. Space in the cache used for sequential data is freed up quicker than other cache or record data. The ESS will continue to pre-stage sequential tracks when the last few tracks in a sequential staging group are accessed.

Stage requests for sequential operations can be performed in parallel on the RAID array, giving the ESS its high sequential throughput characteristic. Parallel operations can take place because the logical data tracks are striped across the physical data disks in the RAID array.

3.28 NVS and write operations

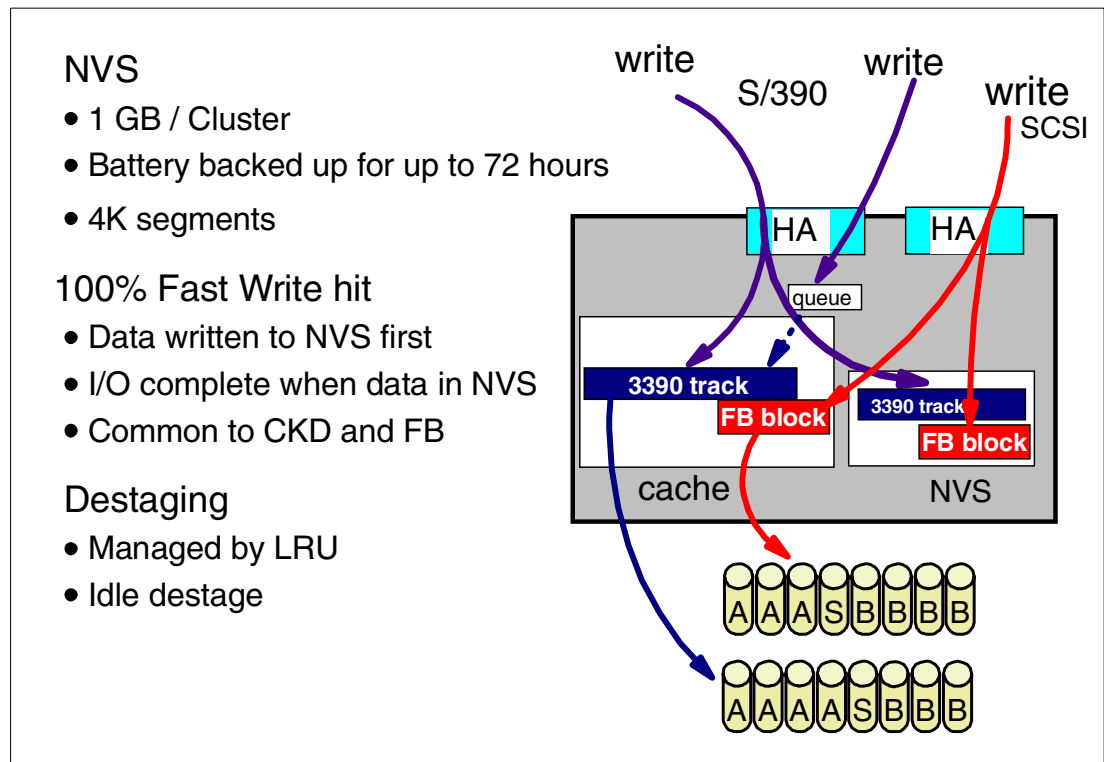


Figure 3-32 NVS - write

As Figure 3-32 illustrates, at any moment there are always two secured copies of any update into the ESS.

3.28.1 Write operations

Data written to an ESS is almost 100% fast write hits. A fast write hit occurs when the write I/O operation completes as soon as the data is in the ESS cache and non-volatile storage (NVS). The benefit of this is very fast write operations.

Fast write

Data received by the host adapter is transferred first to the NVS and a copy held in the host adapter buffer. The host is notified that the I/O operation is complete as soon as the data is in NVS. The host adapter, once the NVS transfer is complete, then transfers the data to the cache.

The data remains in the cache and NVS until it is destaged. Destage is triggered by cache and NVS usage thresholds.

3.28.2 NVS

The NVS size is 2 GB (1 GB per cluster). The NVS is protected by a battery. The battery will power the NVS for up to 72 hours following a total power failure.

NVS LRU

NVS is managed by a Least Recently Used (LRU) algorithm. The ESS attempts to keep free space in the NVS by anticipatory destaging of tracks when the space used in NVS exceeds a threshold. In addition, if the ESS is idle for a period of time, an idle destage function will destage tracks until, after about 5 minutes, all tracks will be destaged.

Both cache and NVS operate on LRU lists. Typically space in the cache occupied by sequential data is released earlier than space occupied by data that is likely to be re-referenced. Sequential data in the NVS is destaged ahead of random data.

When destaging tracks, the ESS attempts to destage all the tracks that would make up a RAID stripe, minimizing the RAID-related activities in the SSA adapter.

NVS location

NVS for cluster 1 is located physically in I/O drawer of cluster 2, and vice versa. This ensures that we always have one good copy of data, should we have a failure in one cluster.

See 3.8, “Cluster operation: failover/failback” on page 60 for more information.

3.29 Sequential operations - read

Sequential reads

- Sequential predict for both CKD and FB I/Os
 - Detects sequential by looking at previous accesses
 - More than 6 I/Os in sequence will trigger sequential staging
- Specific to OS/390 and z/OS
 - Access Methods specify sequential processing intent in CCW
- Stage tracks ahead
 - Up to two cylinders are staged (actual amount of tracks depends on rank configuration)

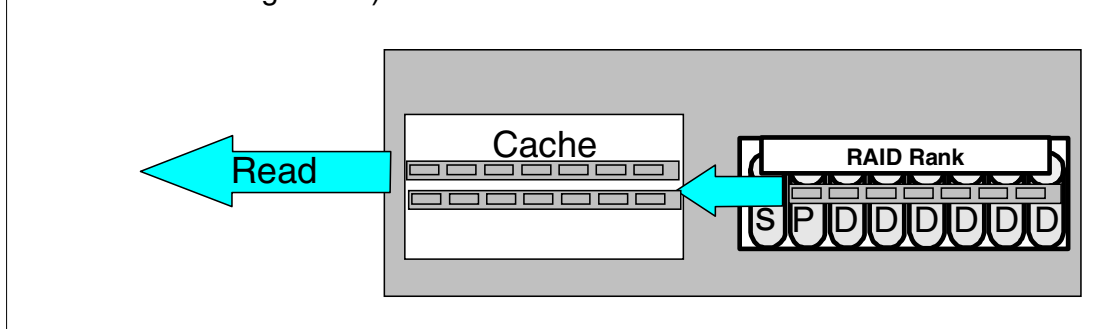


Figure 3-33 Sequential read

For sequential reading, either for RAID 10 or for RAID 5 ranks, the ESS has implemented unique algorithms. There are two ways to trigger the ESS sequential processing. One is automatically initiated by the ESS when it detects the sequential operations; the other is requested by the application when it is going to process sequential I/Os.

The sequential staging reads ahead up to 2 cylinders; the actual amount depends on the array configuration:

- ▶ On 18.2 GB disks arrays for 6+P, it is 30 tracks and for 7+P it is 28 tracks
- ▶ On 36.4 GB and 72.8 GB disks arrays for 6+P, it is 24 tracks and for 7+P it is 28 tracks
- ▶ For RAID 10 (both 3+3' and 4+4' rank configurations, all disk capacities) it is 24 tracks

As the tracks are read, when about the middle of a staging group is read then the next group starts to be staged. This delivers maximum sequential throughput with no delays waiting for data to be read from disk.

Tracks that have been read sequentially are eligible to be freed quickly to release the used cache space. This is because sequential data is rarely re-read within a short period.

Sequential detection

The ESS sequential detection algorithm analyzes sequences of I/Os to determine if data is being accessed sequentially. As soon as the algorithm detects that six or more tracks have been read in succession, the algorithm triggers a sequential staging process.

This algorithm applies equally when accessing CKD data or FB data. An example of environments that benefit from this ESS characteristic is the z/Architecture VSAM files. VSAM

does not set any sequential mode through software, and its sequential processing often skips areas of the data set because, for example, it has imbedded free space on each cylinder.

Software setting

The second method of triggering sequential staging, implemented by S/390 operating systems, is specifying the sequential access through the software in the channel program.

3.30 Sequential operations - write

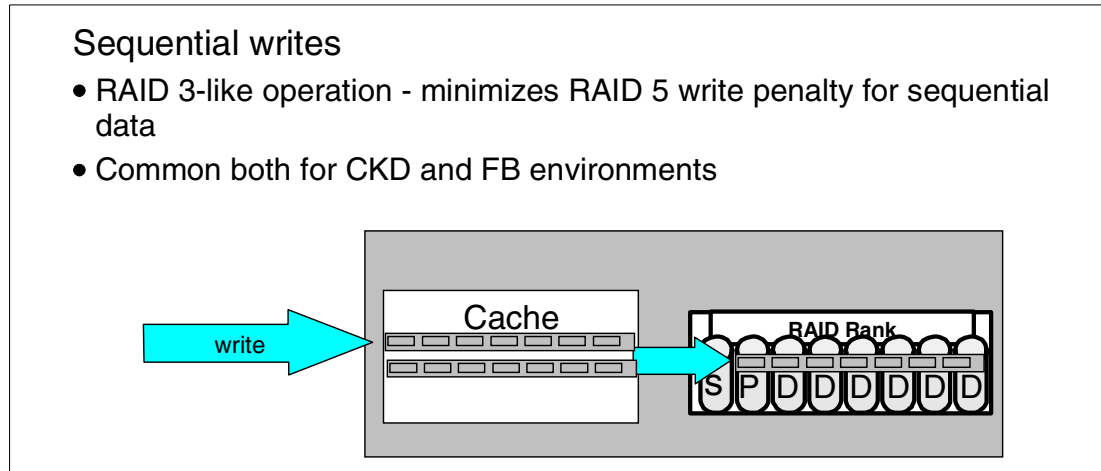


Figure 3-34 Sequential write

Sequential write operations on the RAID 5 ranks are done in a RAID 3 mode (parallel transfer of all stripes of the set). This is beneficial when operating upon the RAID 5 ranks because it avoids the read and recalculation overheads, thus neutralizing the RAID 5 write penalty. An entire stripe of data is written across all the disks in the RAID array, and the parity is generated once for all the data simultaneously and written to the parity disk (the rotating parity disk). This technique does not apply for the RAID 10 ranks, because there is no write penalty involved when writing upon RAID 10 ranks.

3.31 zSeries I/O accelerators

Parallel Access Volumes

- Multiple requests to the same logical volume within the same system image

Multiple Allegiance

- Multiple requests from different hosts to the same logical volume from multiple system images

Priority I/O Queuing

- Enhances the I/O queue management of the ESS

Figure 3-35 CKD accelerators

For the zSeries and S/390 servers, the IBM TotalStorage Enterprise Storage Server provides some specific performance features that are introduced in this section. These features are explained in detail in Chapter 6, “zSeries performance” on page 165.

3.31.1 Parallel Access Volumes (PAV)

Parallel Access Volumes (PAV) allows the host system to access the same logical volume using alternative device address UCBs (operating system unit control blocks). For the z/OS operating system, two types of device addresses can be defined: base device addresses and alias device addresses. The base represents the real device and the aliases represent an alternate access.

Multiple read requests for the same track in cache will be read hits and will provide excellent performance. Write operations will serialize on the write extents and prevent any other PAV address from accessing these extents until the write I/O completes. As almost all writes are cache hits, there will be only a short delay. Other read requests to different extents can carry on in parallel. All this parallelism is handled by the ESS, thus eliminating the high I/O enqueue and I/O re scheduling activity that otherwise is handled by the host operating system.

3.31.2 Multiple Allegiance

Multiple Allegiance (MA) allows multiple requests, each from multiple hosts to the same logical volume. Each read request can operate concurrently if data is in the cache, but may queue if access is required to the same physical disk in the array. If you try to access an extent that is part of a write operation, then the request will be queued until the write operation is complete.

3.31.3 I/O priority queuing

I/Os from different z/OS system images can be queued in a priority order. It is z/OS's Workload Manager with the ESS that can utilize this priority to favor I/Os from one system over the others.



Configuration

This chapter covers the configuration of the IBM TotalStorage Enterprise Storage Server Model 800. The initial part of the chapter describes the choices you have when deciding on the hardware configuration prior to ordering the ESS. The final part of this chapter details the logical configuration procedure.

4.1 Overview

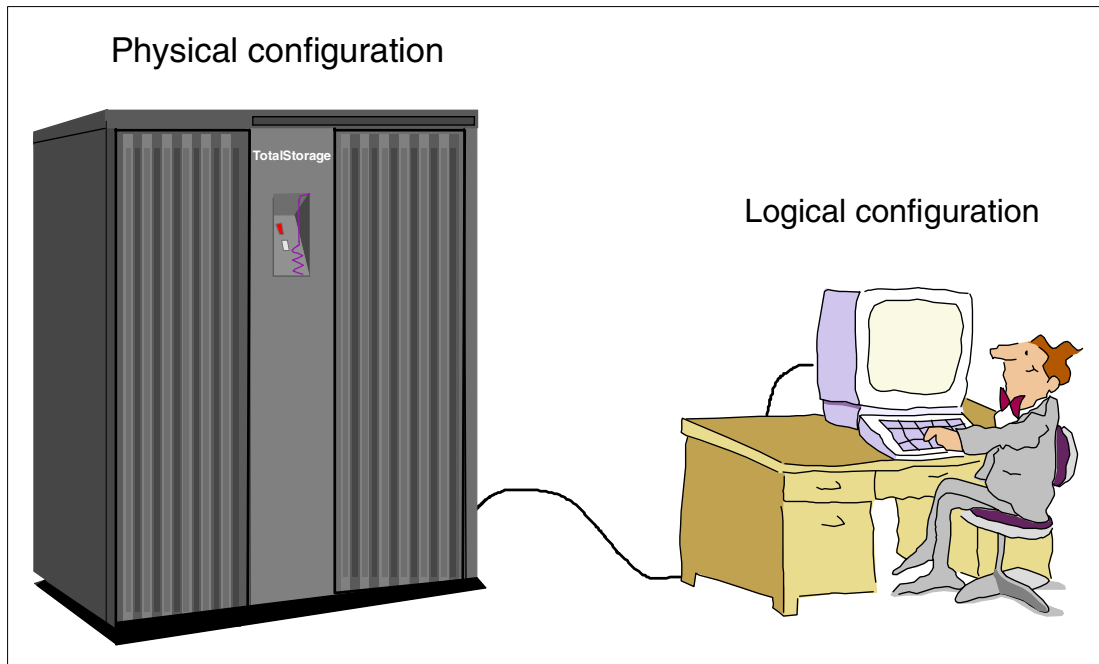


Figure 4-1 Configuration process

The configuration process of the ESS consists of two parts:

1. The physical configuration for ordering the hardware
2. The logical configuration when installing and, on subsequent modifications, for preparing the ESS to work with the attached servers and applications

For the physical configuration task, you may complement the information in this chapter with the information presented in Appendix A, “Feature codes” on page 263.

When doing the logical configuration of the ESS, refer to the publication *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448 for more information and how-to details. The publications *IBM TotalStorage Enterprise Storage Server Configuration Planner for Open-Systems Hosts*, SC26-7477 and *IBM TotalStorage Enterprise Storage Server Configuration Planner for S/390 and zSeries Hosts*, SC26-7476 provide the configuration worksheets to properly plan and document the desired logical configuration.

4.2 Physical configuration

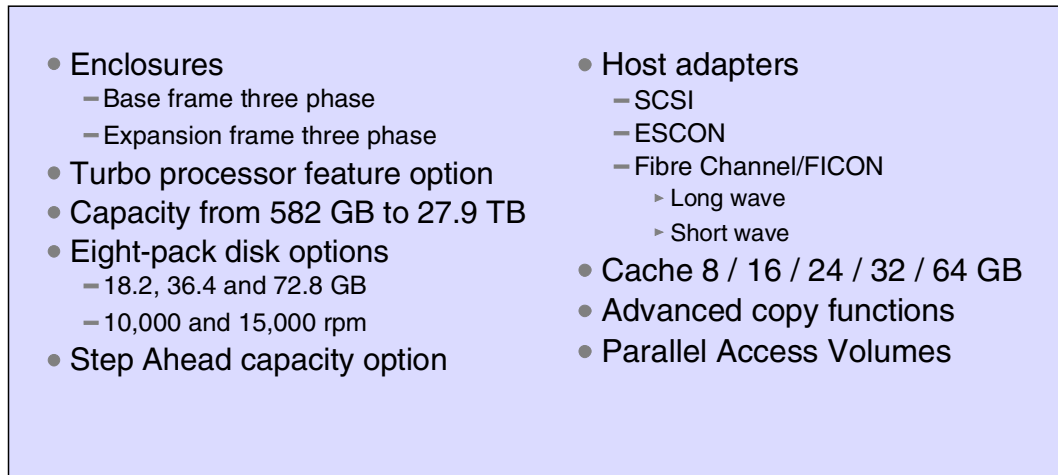


Figure 4-2 Physical configuration options

During the physical configuration, you determine the basic setup of the IBM TotalStorage Enterprise Storage Server hardware by determining:

- ▶ The number of disk eight-packs of the desired disk capacity and speed that will amount for the total storage capacity you need
- ▶ The rack configuration
- ▶ Cluster processor options
- ▶ Cache size
- ▶ The type and quantity of host attachments, which can be ESCON, SCSI and Fibre Channel/FICON
- ▶ As part of the physical configuration process, you determine the desired advanced copy functions
- ▶ If you are a z/OS user, whether you will be needing PAV or not

Figure 4-2 summarizes the major options that you should consider when configuring your ESS order.

4.3 Storage capacity

The IBM TotalStorage Enterprise Storage Server Model 800 provides a much richer capacity scalability, several options of disk drives, and a Step Ahead feature. These choices allow for a hardware configuration that more efficiently meets the requirements and demands of the particular operating environment where the ESS is installed.

4.3.1 Flexible capacity configurations

ESS storage capacity is measured in *raw capacity*, and ranges from 582 GB to 27.9 TB in increments of eight-pack pairs. You can scale easily from one capacity to another by adding more eight-pack pairs as required.

The ESS Model 800 disk arrays can be configured as either RAID 5 or RAID 10 ranks, resulting in different effective capacities for a given physical capacity (refer to Table 2-1 on

page 27). As such, all capacities quoted are *raw* capacities, which can be simply calculated by multiplying the DDM size (for example, 36.4 GB) by 8 (number of DDMs in an eight-pack) and then multiplying by the number of eight-packs. If eight-packs of different capacities are installed, then the calculation should be done for each size and the results added together.

When a disk group is formatted as a RAID array, some DDMs are used for parity or mirroring, and some may be reserved as spares. As such, the effective usable capacity will depend on the type of RAID selected and also the number of disk groups in an SSA loop. The details of this will be covered later.

In the physical configuration process, you select the capacity and speed of the disk drives that will come in the eight-packs. Currently there are three disk-drive sizes that can be configured with the ESS, with capacities of 18.2 GB, 36.4 GB, and 72.8 GB, and two speeds of 10,000 and 15,000 rpm (only 10,000 rpm for 72.8 GB). You also have the option to order an ESS Step Ahead configuration, where the plant will ship an extra eight-pack pair of any size pre-installed. This way you will be able to respond immediately to unexpected and urgent increases in demand for additional storage capacity.

4.3.2 Step Ahead configurations

The IBM TotalStorage Enterprise Storage Server provides a wide range of Step Ahead configurations. This capacity upgrade can be performed concurrently with normal I/O operations on the ESS. This option allows you to be prepared to immediately respond to unexpected demands for extra storage capacity.

Step Ahead works by allowing you to order an ESS with an additional eight-pack pair of any capacity already installed, and only be billed for a carrying fee. This enables *capacity on demand* by configuring the extra capacity when required, at which point you will be billed for the extra capacity. This feature must be renewed annually if not used. See Appendix A, "Feature codes" on page 263 for more details.

Note: The advanced functions PPRC and FlashCopy are billed by total ESS capacity, and this includes any Step Ahead capacity installed.

4.4 Logical configuration

<ul style="list-style-type: none">• For FB servers<ul style="list-style-type: none">– Generic LUN sizes from 100 MB to maximum rank capacity• For iSeries servers<ul style="list-style-type: none">– SCSI: 9337-48x 59x 5Ax 5Cx 5Bx– FCP: 2105 Ax1 Ax2 Ax3 Ax4 Ax5– protected / non-protected• For CKD servers<ul style="list-style-type: none">– 3390-2/-3/-9 in 3390 track format– 3390-2/-3 in 3380 track format– CKD custom volumes from 1 to 32,760 cylinders	<ul style="list-style-type: none">• CKD CU images for zSeries servers<ul style="list-style-type: none">– 3990-3– 3990-3 TPF– 3990-6• Rank definition<ul style="list-style-type: none">– RAID 5 (6+P+S / 7+P)– RAID 10 (3+3+2S / 4+4)
---	--

Figure 4-3 Logical configuration characteristics

Once the ESS is installed, you will be ready to do the logical configuration, where you define the relationship between the ESS and the attached hosts. You can configure the ESS for the open systems FB servers, either with SCSI or Fibre Channel attachment. You also configure the ESS for the zSeries servers, either with ESCON or FICON attachment. For FB architecture servers, the ESS is viewed as generic LUN devices, except for the iSeries servers for which the devices are 9337 or 2105 volumes. For the zSeries servers, the ESS is seen as up to sixteen 3990 Storage Controllers with 3390-2, 3390-3, 3390-9 standard volumes and custom volumes. For these servers, the ESS can also emulate 3380 track formats to be compatible with 3380 devices. You can also configure a combination of FB and CKD space to share the ESS between open systems and zSeries servers.

Figure 4-3 shows some initial options you will be considering when starting with the logical configuration of the ESS. The logical configuration process is done primarily using the ESS Specialist. During this procedure, you present to the ESS all the required definitions needed to logically set up the ESS and make it operational.

4.4.1 Logical standard configurations

To assist with the configuration process, there is the alternative of specifying some standard formatting options. Once the ESS has been installed, the IBM System Support Representative will format each loop according to the standard configurations you may have selected. This alternative eliminates some of the ESS Specialist steps needed for the logical configuration, shortening the installation procedure. These standard formatting options are available for both CKD and FB servers. See 4.35, “Standard logical configurations” on page 145 for descriptions of these configuration options.

4.5 Base enclosure

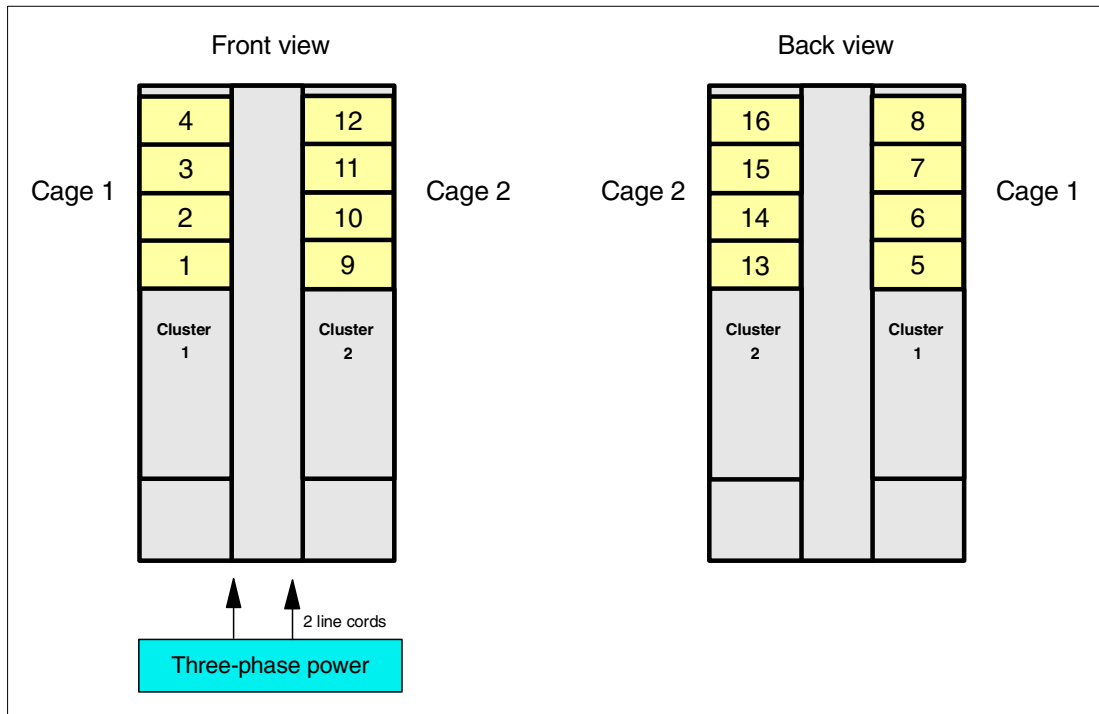


Figure 4-4 ESS base enclosure— eight-pack installation sequence

The base enclosure of the ESS always contains two cages. The eight-packs are installed in the cages that provide them with power. The minimum configuration is four eight-packs in cage 1. The numbers 1 to 16 in Figure 4-4 indicate the sequence in which the eight-packs are installed in the cages. The first two eight-packs are installed in the lowest position of the front of cage 1. All cages are filled from front to back, and from bottom to top.

When the ESS base enclosure is fully populated with eight-packs, it is then holding 128 disk drives. The base frame is powered by two three-phase power supplies, each with its own line cord. See Chapter 2, “Hardware” on page 19 for more details.

4.6 Expansion Enclosure

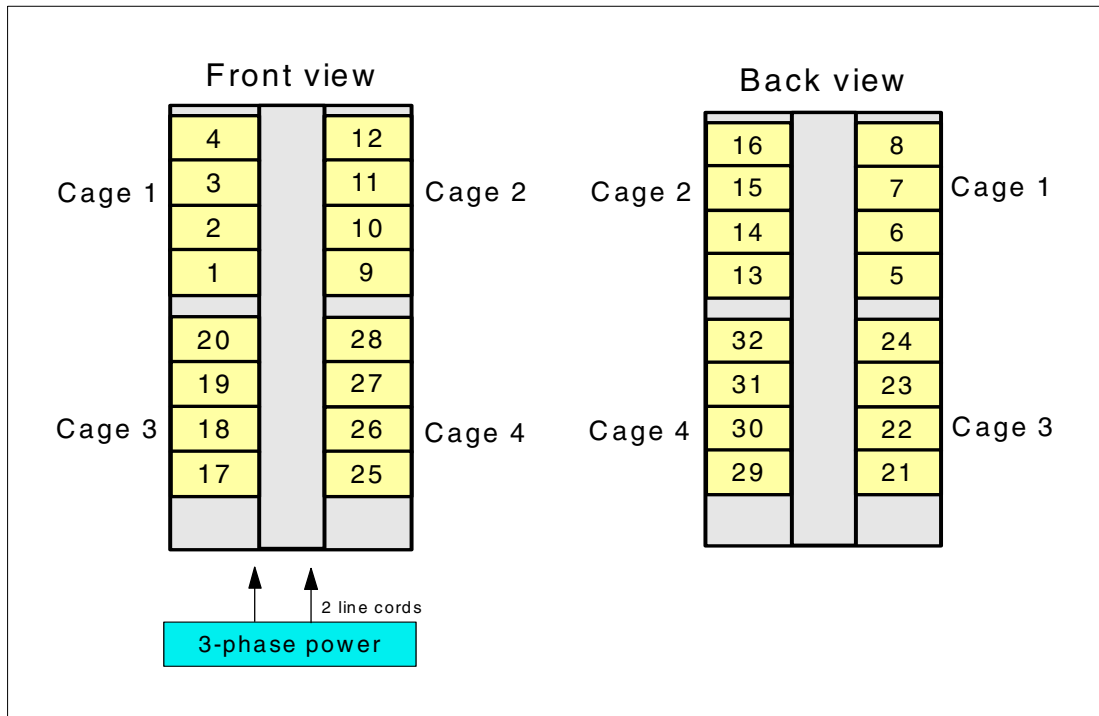


Figure 4-5 ESS Expansion Enclosure — eight-pack installation sequence

The ESS Expansion Enclosure (feature 2110 of the ESS Model 800) can accommodate four cages. The minimum configuration is zero eight-packs, which will give you an empty expansion frame (minimum with one cage pre-installed) ready for future upgrades. When adding disk drives to the Expansion Enclosure rack, it will be done by installing eight-pack pairs, starting with cage 1. The numbers 1 to 32 in Figure 4-5 indicate the sequence in which the eight-packs are installed in the cages of the Expansion Enclosure. The first two eight-packs are installed in the lowest position of the front of cage 1. All cages are filled from front to back, and from bottom to top.

When the Expansion Enclosure is fully populated with eight-packs, it is then holding 256 disk drives. The Expansion Enclosure is powered by two three-phase power supplies (two power line cords are used by the Expansion Enclosure rack itself, in addition to the ESS base frame power line cords).

4.7 Base and Expansion Enclosure loop configuration

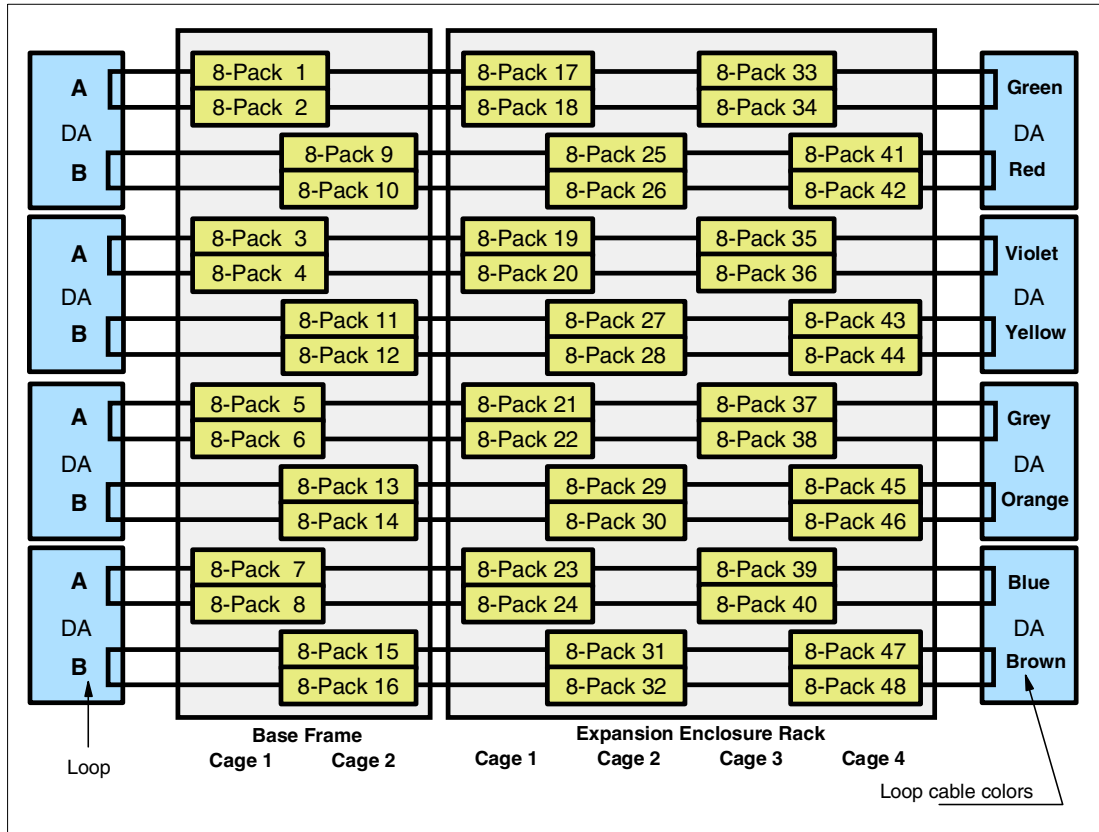


Figure 4-6 Eight-pack logical loop configuration — with Expansion Enclosure

Figure 4-6 shows a fully configured IBM TotalStorage Enterprise Storage Server with the Expansion Enclosure rack and the maximum number of eight-packs. The eight-packs are numbered in the sequence they are installed on the loops. As you can see, the eight-packs in the Expansion Enclosure are on the same loops as the eight-packs in the base frame.

4.8 Upgrade with eight-packs

If you plan to increase the capacity of the ESS with more disk drives, you do it by adding pairs of eight-packs. All machines are shipped with two Disk Eight-Pack Mounting Kit features (sheet metal cages, power supplies, fans, cables, etc., referred to as *cages*) installed in the base enclosure. If an Expansion Enclosure is also ordered, then you can choose to have up to four Disk Eight-Pack Mounting Kits (cages) pre-installed. If the cages are not already present, they will be shipped with the disk upgrade order. Sufficient cages must be available to hold the number of eight-packs required (eight per cage).

For the base frame, cage 1 is filled first, working bottom up at the front and then bottom up at the back. Cage 2 is then filled in the same sequence. Once the base frame is completed, then the Expansion Enclosure is filled using the same sequence in cages 1 through 4. This sequence can be seen in Figure 4-5 on page 99.

4.9 Physical configuration

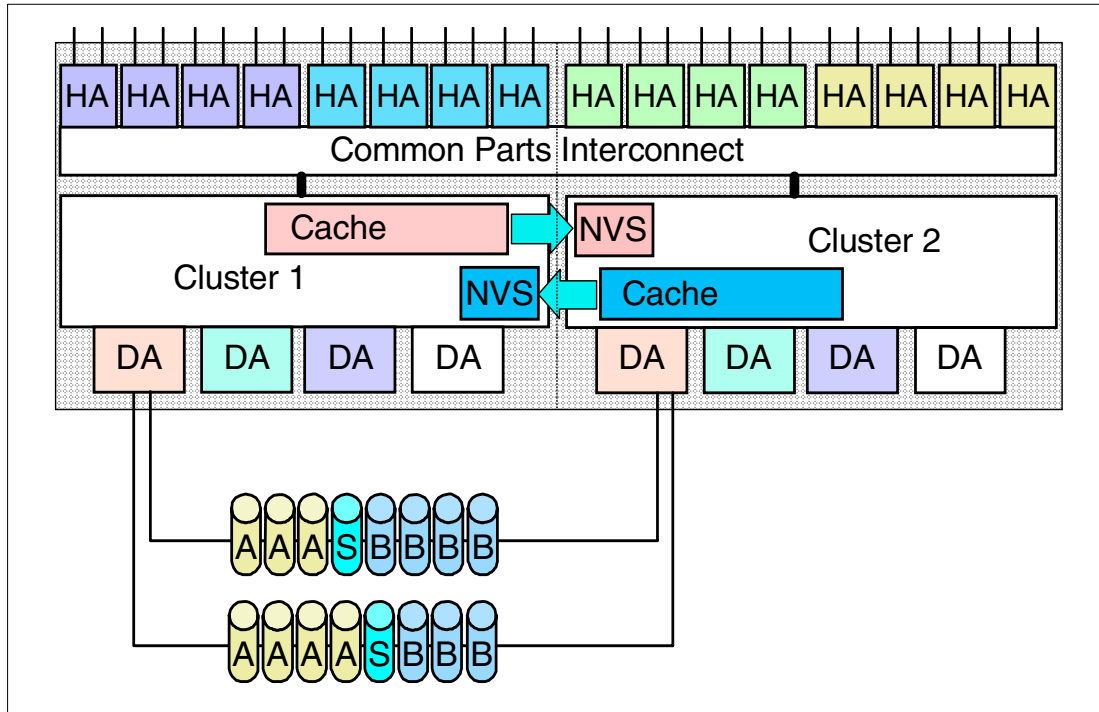


Figure 4-7 Block diagram of an ESS

Figure 4-7 illustrates the basic layout of the IBM TotalStorage Enterprise Storage Server architecture showing only one SSA loop with two eight-packs formatted as two RAID 5 ranks. You may find it helpful to refer to Appendix A, “Feature codes” on page 263 during the following section.

4.9.1 Cluster Processors

Each cluster of the ESS Model 800 has a powerful SMP (symmetrical multiprocessor) standard processor, with an option for a more powerful Turbo processor feature (this feature can be ordered initially or can be field installed at a later date). In either case, both clusters have the same processors. See 2.7.1, “Processors” on page 29 for more details.

4.9.2 Cache

You have the option of selecting any of five cache sizes to accommodate your performance needs: 8, 16, 24, 32 or 64 GB. This total cache capacity, which is split between both clusters, is selected when you order the IBM TotalStorage Enterprise Storage Server Model 800 from the plant by specifying one of the appropriate feature codes. It can be field upgraded at a later date if required. See 2.7.2, “Cache” on page 29 for more details.

4.9.3 Host adapters

The host adapters (HAs) are mounted in bays. Each of the four bays is able to hold up to four host adapters, making a maximum of 16 host adapters for the ESS. Each cluster has two host adapter bays installed. The host adapter cards can either be ESCON, SCSI or Fibre Channel/FICON (long wave or short wave). The Fibre Channel/FICON host adapters can be

configured as either Fibre Channel or FICON (one or the other but not simultaneously) on an adapter-by-adapter basis.

All of the host adapter cards are connected to the clusters by the *Common Parts Interconnect* (CPI), as shown in Figure 4-7 on page 101. This allows any of the cards to communicate with either cluster. You can mix ESCON, SCSI, and FICON/Fibre Channel host adapter cards in the ESS. Remember that the total number of host adapter cards, be it ESCON, SCSI or Fibre Channel/FICON, cannot exceed 16. The Fibre Channel/FICON card is a single port host adapter, whereas SCSI and ESCON have two ports for connection. The upgrade of host adapters is done by installing additional HA cards in the bays.

You must specify the type and number of host adapters for the machine you are ordering. The minimum order is two HAs of the same type. Later, once the ESS is installed, you can also add or replace HAs (See Appendix A, "Feature codes" on page 263). Note that the four bays are a standard part of the ESS and always come with the machine independently of the configuration ordered.

The order in which the ESS host adapter cards are installed in the machine during the manufacturing process is:

- ▶ Cluster 1 - Bay 1 - Adapter 1
- ▶ Cluster 2 - Bay 4 - Adapter 1
- ▶ Cluster 1 - Bay 2 - Adapter 1
- ▶ Cluster 2 - Bay 3 - Adapter 1
- ▶ Cluster 1 - Bay 1 - Adapter 2
- ▶ Cluster 2 - Bay 4 - Adapter 2
- ▶ Cluster 1 - Bay 2 - Adapter 2
- ▶ Cluster 2 - Bay 3 - Adapter 2
- ▶ Cluster 1 - Bay 1 - Adapter 3
- ▶ And so on, until filling the 16 host adapter card positions across the four bays

In addition, the ESCON host adapters are installed first, then the long-wave Fibre Channel/FICON adapters, then the short-wave Fibre Channel/FICON adapters, and finally the SCSI adapters.

4.9.4 Device adapters

The device adapter (DA) cards are installed into the I/O drawer below the processor drawer. There are no bays for the DA cards. The device adapter cards connect in pairs and support two SSA loops. This pairing of DAs not only adds performance, but also provides redundancy. There are four pairs of DAs in the IBM TotalStorage Enterprise Storage Server, each supporting two SSA loops. When a loop holds disk-drive capacity, the minimum it can have installed is 16 disk drives (two eight-packs). For capacity upgrades, additional disk drives in groups of 16 (two eight-packs) are installed in a predetermined sequence (as explained in 4.8, "Upgrade with eight-packs" on page 100). A maximum of 48 disk drives are supported in each SSA loop. A pair of device adapters will always have access to all the disk drives that belong to the SSA loop.

There is no feature specification needed for the device adapters, since they are a standard component of the ESS. The eight DAs will come installed four per cluster, whether you order a small-capacity configuration or a large-capacity configuration.

4.9.5 Performance accelerators

The only optional performance accelerator feature is the Parallel Access Volume (PAV) feature for z/OS, described in detail in 6.2, “Parallel Access Volume” on page 166. All other performance features come standard. The PAV feature can be field installed if required.

4.9.6 ESS copy functions

There are three optional copy functions:

- ▶ The FlashCopy point-in-time copy function
- ▶ Synchronous Peer-to-Peer Remote Copy (PPRC), which includes non-synchronous PPRC Extended Distance
- ▶ The asynchronous Extended Remote Copy (XRC) function

These features are described in detail in Chapter 7, “Copy functions” on page 193 and they all can be field installed if required.

4.10 Loop configuration

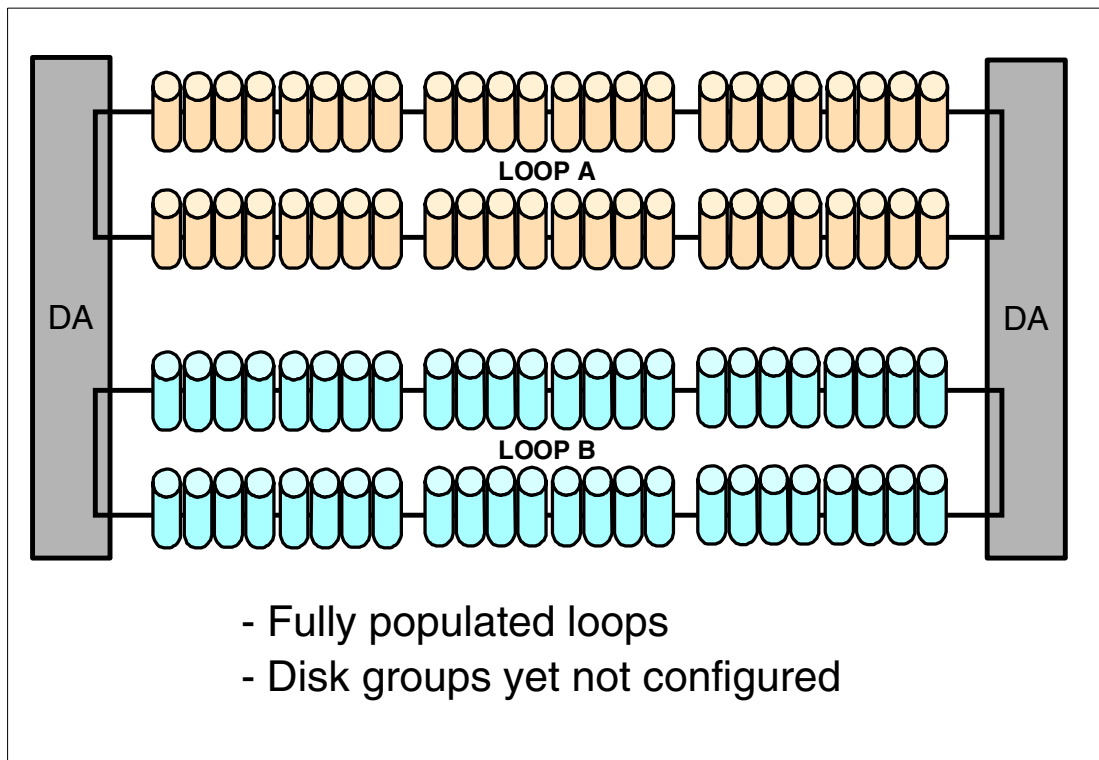


Figure 4-8 DA pair - maximum loop configuration

Figure 4-8 shows a device adapter pair with two loops (A and B), each containing the maximum number of six eight-packs, which results in 48 disk drives per loop. Initially the *disk groups* in the loops are yet not configured nor assigned. In the following pages we show how these disks groups are configured for the different RAID combinations.

4.11 Loop configuration – RAID 5 ranks

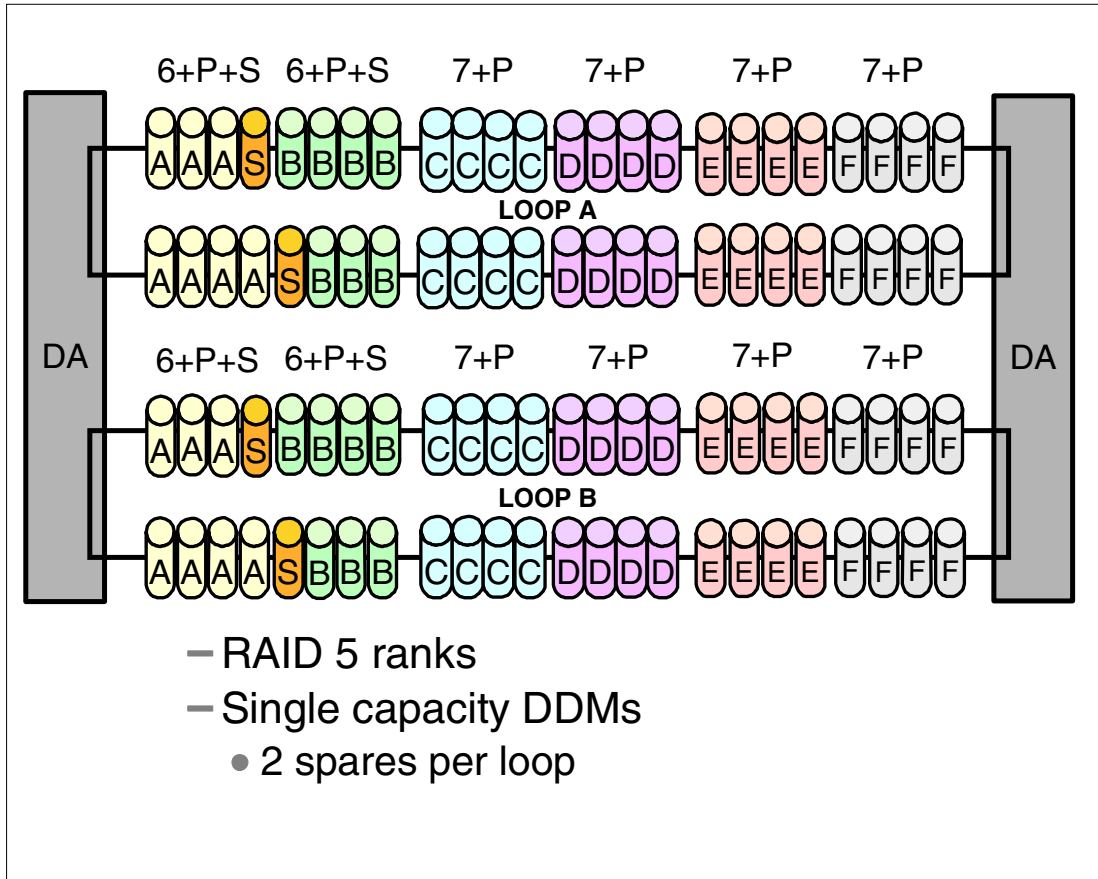


Figure 4-9 Single-capacity DA pair loop configuration - RAID 5 ranks

Figure 4-9 provides details about the SSA loop configuration in an ESS where all the disks are of the same capacity and all disk groups are formatted as RAID 5 ranks (disk drives of the same letter within a loop make up one RAID rank). The logical configuration procedure that you run will ensure two spare drives assigned to each loop. These drives are shown in the figure with an S, and are globally available for any array in that loop. Since every loop must have a minimum of two spare disks, each of the first two RAID 5 ranks configured in a loop will contain one spare disks. After this, subsequent arrays (of similar raw capacity) added to the loop will not be configured with spares.

The RAID 5 rank is managed by the SSA device adapter, and provides drive redundancy within the array to protect against disk failure. This option provides high availability with a low overhead in terms of disk capacity needed to store the redundant information. See 3.10.2, “RAID 5 rank” on page 64 for more details.

4.12 Loop configuration – RAID 10 ranks

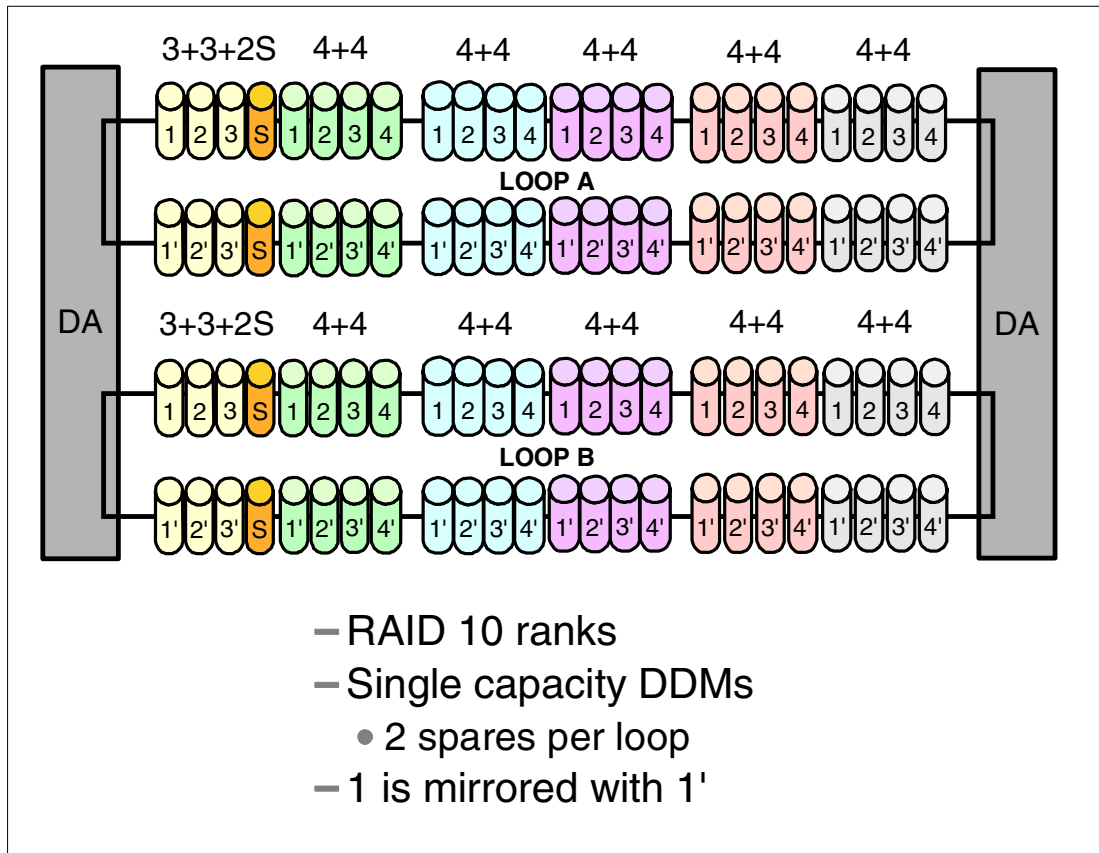


Figure 4-10 Single capacity DA pair loop configuration - RAID 10 ranks

Figure 4-10 provides details about the SSA loop configuration in an ESS where all the disks are of the same capacity and all the disk groups formatted as RAID 10 ranks (disk drives of the same color within a loop make up a RAID rank). The logical configuration procedure that you run will ensure two spare drives assigned to each loop. These drives are shown in the figure with an S, and are globally available for any array in that loop. Since every loop must have a minimum of two spare disks, the first RAID 10 rank configured in a loop will contain the two spare disks. After this, subsequent arrays (of similar raw capacity) added to the loop will not be configured with spares.

The RAID 10 rank is managed by the SSA device adapter, and provides drive redundancy within the array to protect against disk failure by mirroring (RAID 1) the three or four striped (RAID 0) data disks (in Figure 4-10, n is mirrored to n'). RAID 10 provides data protection by means of redundancy and better performance for selected applications, but with a higher cost overhead due to the capacity required for the mirrored data as compared to RAID 5. See 3.10.3, “RAID 10 rank” on page 65 for more details.

4.13 Loop configuration – Mixed RAID ranks

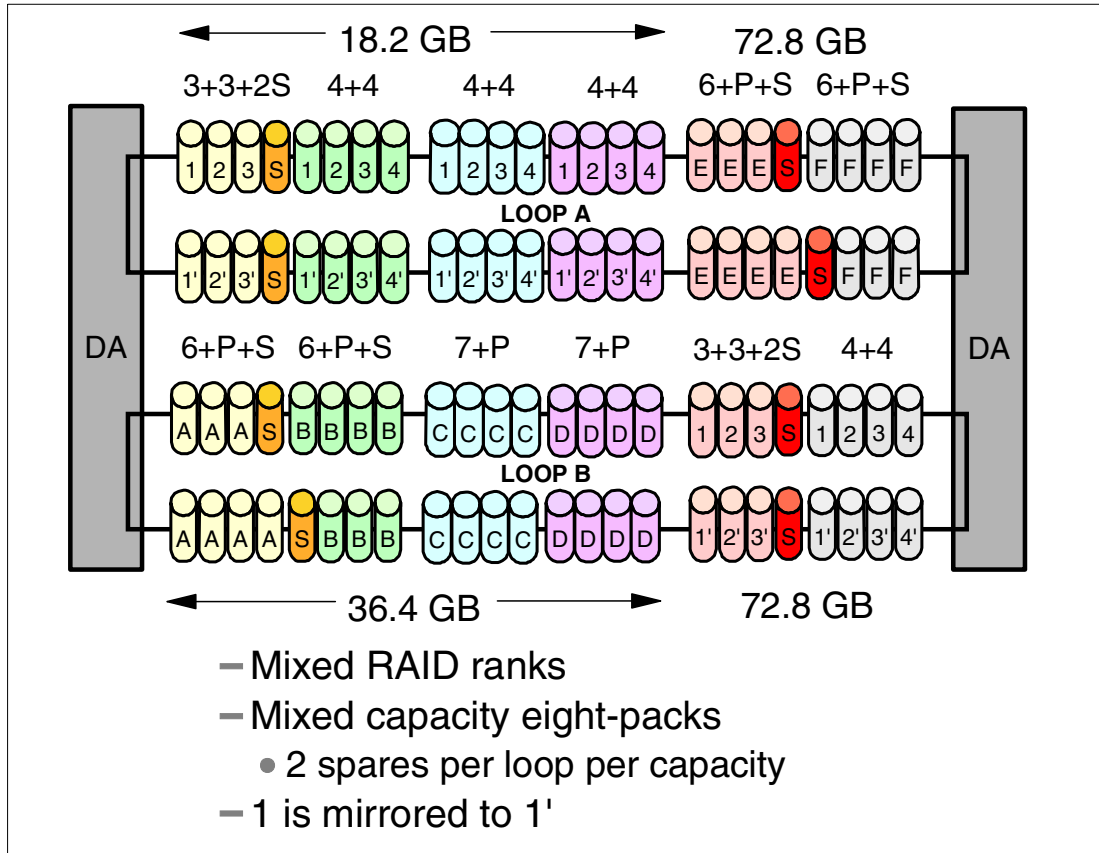


Figure 4-11 Mixed capacity DA pair loop configuration - RAID 10 and RAID 5 ranks

Figure 4-11 provides details about the loop configuration in an ESS where some disk groups are formatted as RAID 10 ranks and others are configured as RAID 5. Also (as illustrated in Figure 4-11) the eight-packs installed on these loops differ in their raw capacities because of the different disk drives models (for each eight-pack, the disks are all of the same capacity and speed). The logical configuration procedure that you run will leave at least two spare drives *per capacity per loop*. These drives are shown in the figure with an S, and are globally available for any array in that loop of the same capacity DDM.

Since every loop must have at least two spare disks of each capacity, either the first two ranks that are configured (if RAID 5) will hold these spares, or the very first rank that is configured (if RAID 10) will hold both spares. After this, subsequent arrays added to the loop will not be configured with spares, provided they use the same capacity DDMs. If the subsequent arrays use DDMs of a different capacity, then two spare disks will again be reserved.

In the example in Figure 4-11, all the 72.8 GB arrays contain spares, since the four arrays are split across two loops and hence cannot share spare DDMs. See 4.22, “RAID 5 and RAID 10 rank intermixing” on page 120 for more about mixing capacities or RAID formats within a loop.

4.13.1 RAID 5 versus RAID 10 considerations

RAID 5 and RAID 10 both provide protection against disk drive failure through redundancy.

RAID 5 does this by using an N+1 design, where there are logically N data disks (in the ESS either six or seven) and one parity disk. In fact, the data and parity are interleaved and then striped across the array to avoid hot spots and hence improve performance.

RAID 10 is a combination of RAID 0 and RAID 1. The data is striped across N disk drives (in the ESS either three or four) and then mirrored to a second set of N disk drives in the same array. The mirrored writes can occur in parallel, and read requests can be satisfied from either disk in the mirrored pair.

RAID 10 performance is normally better for selected applications than RAID 5 (see 5.6.3, “RAID 10 vs RAID 5” on page 153 for a further discussion of performance), but there is a much higher cost overhead in terms of extra disk drives needed, as shown in Table 4-1.

Table 4-1 RAID 5 versus RAID 10 overheads compared to non-RAID

Array type	Data DDMs per eight-pack	Overhead	Utilization
RAID 5: 7+P	7	12.5%	87.5%
RAID 5: 6+P+S	6	25.0%	75.0%
RAID 10: 4+4	4	50.0%	50.0%
RAID 10: 3+3+2S	3	62.5%	37.5%

4.13.2 Reconfiguration of RAID ranks

You can reconfigure a RAID array from RAID 5 to RAID 10, or RAID 10 to RAID 5, at any time, *but all data contents will be lost*, so you must offload the data first and reload it afterwards.

To maintain the minimum number of spare disks, the ESS will not allow you to reconfigure in isolation the initial array in the loop of a given capacity from RAID 10 to RAID 5. If the RAID 10 array was the first one created in the loop, then it will contain the two spares (3+3+2S). Conversion to RAID 5 would leave only one spare (6+P+S), and a loop must always have at least two spares for each size of DDM in the loop. To achieve this, another array must be re-configured to RAID 5 at the same time to provide the second spare.

4.14 ESSNet setup

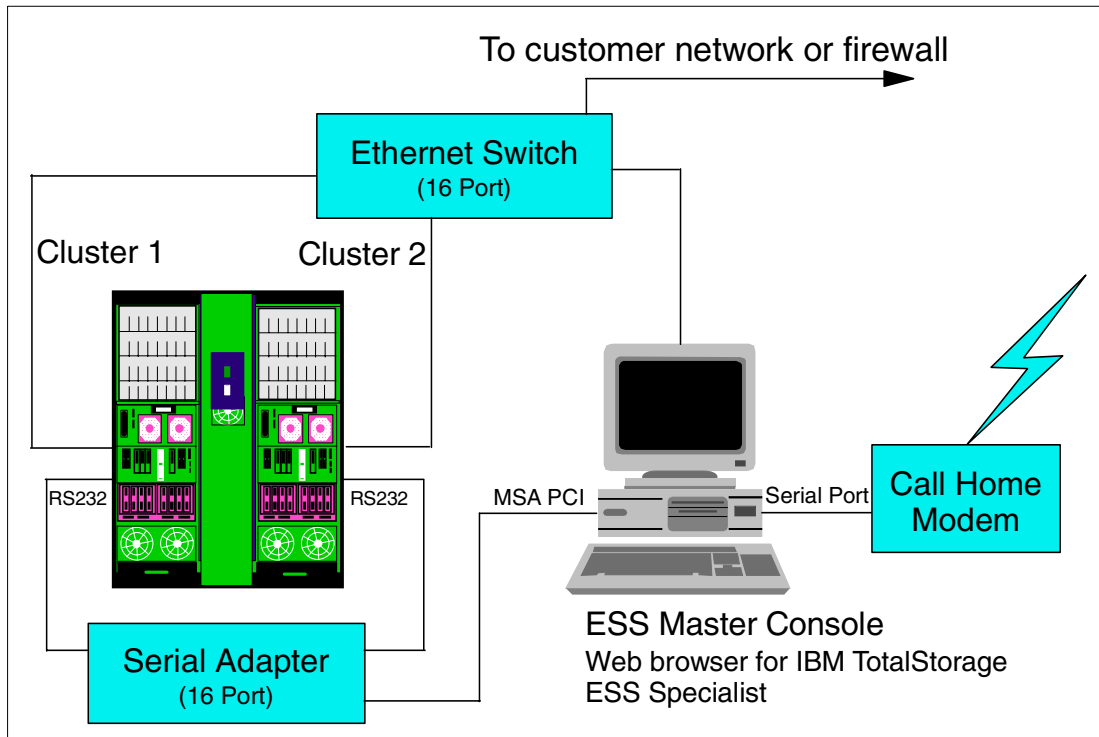


Figure 4-12 ESSNet setup

In order to start using the ESS Specialist for the logical configuration process, you first set up a local area network connecting to the IBM TotalStorage Enterprise Storage Server Model 800, and a Web browser interface to access the ESS Specialist windows.

4.14.1 ESSNet and ESS Master Console

The *Enterprise Storage Server Network* (ESSNet) is a dedicated local area network to support configuration, copy services communications between machines, Call Home, and remote support capabilities. At the center of the ESSNet is the IBM TotalStorage Enterprise Storage Server *Master Console* (ESS Master Console), which serves as a single point of control for up to seven ESSs. Multiple ESSNets can also be interconnected and even connected into an enterprise-wide area network to enable control of the ESSs from a central location, regardless of where the machines may be located.

The ESS Master Console consists of a dedicated PC running Linux, a modem, a 16-port Multiport Serial Adapter (MSA) with PCI adapter card, and an Ethernet switch with cables. The IBM System Support Representative (SSR) will attach two LAN cables (one per cluster) from the ESS to the Ethernet switch, and two RS-232 serial cables (again, one per cluster) to the MSA, as shown in Figure 4-12.

See 2.18, “ESS Master Console” on page 46 for more information on the use of the ESS Master Console and ESSNet.

4.14.2 User local area network

Attachment to the user's local area network permits access to the ESS Specialist outside the immediate vicinity of the ESS, but it also has the potential of increasing the number of people who can access and configure the ESS.

If you want to attach the ESSNet switch to your LAN, you will need to provide the appropriate TCP/IP information to the IBM System Support Representative, since each ESS cluster plus the ESS Master Console will need to be configured with static TCP/IP addresses that are recognized as part of your TCP/IP network. The IBM SSR will enter the required TCP/IP information and connect your LAN cable to the ESSNet switch.

An additional benefit of connecting the ESSNet to your LAN is the ability to enable e-mail problem notification and SNMP notification. For these, you should provide the IBM SSR with additional network configuration details, such as default gateway and name server addresses.

Tip: IBM recommends that the ESS network be configured such that connectivity is limited to those requiring access to the ESS. It is not recommended that it be installed on the enterprise intranet, nor the worldwide Internet, unless sufficient security is implemented. Installing the ESS behind a firewall is one method of improving ESS security.

Additional information can be found in the publication *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444.

4.14.3 Physical setup

See 2.18.1, "ESS local area network" on page 47 for details of the physical installation of the ESSNet and the ESS Master Console.

4.14.4 Web browser

The ESS Specialist provides an easy-to-use Web-based interface connection. It is implemented via a Web server running in each of the ESS clusters. Using a Web browser, such as Netscape Navigator or Microsoft Internet Explorer, you can access the ESS Specialist from a desktop or mobile computer as supported by your network, in addition to the ESS Master Console. The key is that the browser must support the correct level of Java. See 8.9, "ESS Specialist" on page 242 for where to find details of supported Web browsers.

Access to the ESS Specialist requires a valid user name and password. When started for the first time on a newly installed ESS, the ESS Specialist has one administrator user name predefined. Log on with this user name and immediately create one or more administrative users with secure passwords. This is required because the provided user name will automatically be deleted by the ESS Specialist after the first new administrator user name has been added to the list, and this user name cannot be redefined. As such, it is wise to immediately create more than one new account with administrator authority.

Tip: The user name and password fields are case sensitive.

4.15 The ESS Specialist

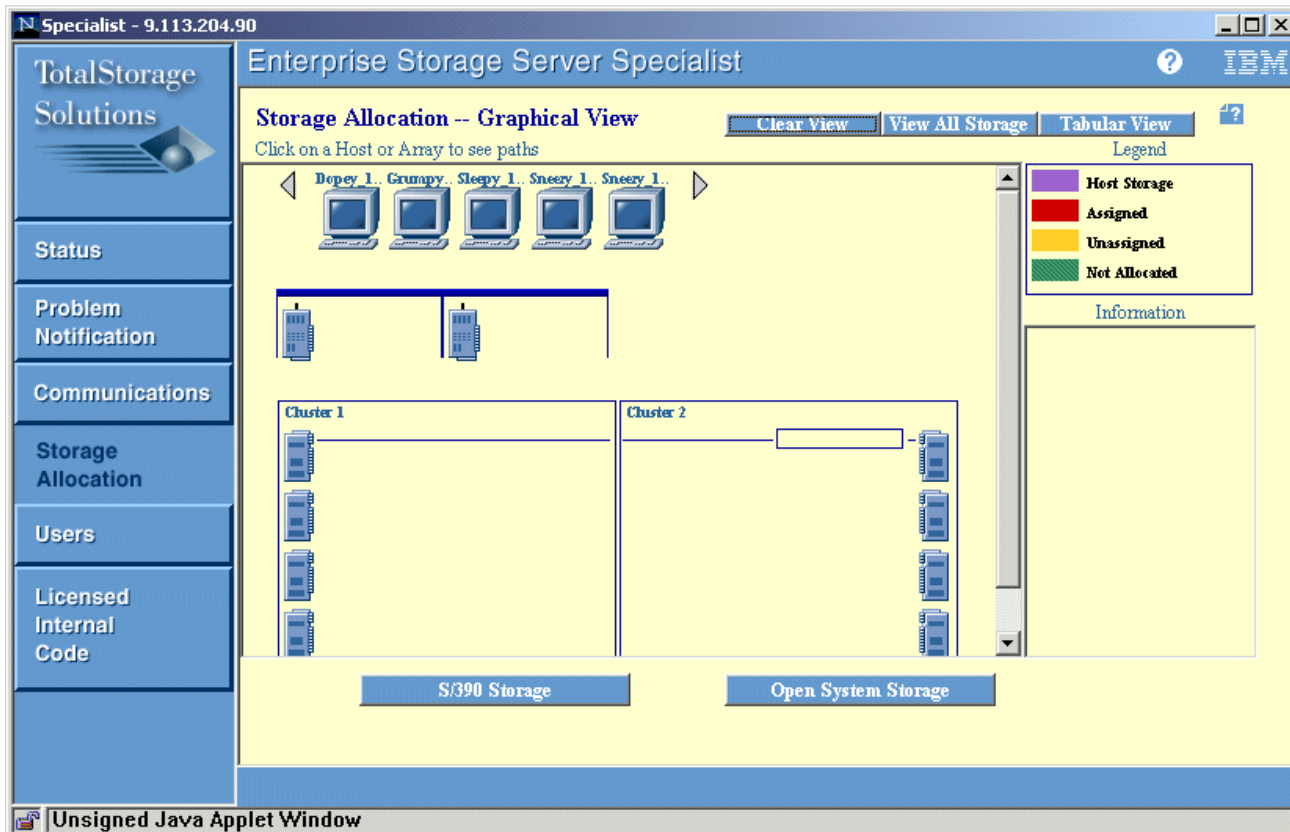


Figure 4-13 ESS Specialist — Storage Allocation window

Before explaining the logical configuration, let us see the interface you will be using to do it. The IBM TotalStorage Enterprise Storage Server Specialist (ESS Specialist) is the interface that will assist you in most of the logical configuration definitions.

In this section we do a very brief introduction of the ESS Specialist. To learn more about its functions and how to use it, refer to the publication *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448.

4.15.1 ESS Specialist configuration windows

Figure 4-13 shows the Storage Allocation window of the ESS Specialist. You arrive at this window by clicking the **Storage Allocation** button from the initial ESS Specialist Welcome window, that you get once you access the ESS Specialist from your Web browser.

The Storage Allocation window shows the ESS logical view for the host systems, host adapter ports, device adapters, arrays and volume assignments. You start the logical configuration process from the Storage Allocation window by selecting either the **Open System Storage** button or the **S/390 Storage** button, which will take you to their associated windows.

Use the Open Systems Storage window to configure new host system attachments and storage, or to modify existing configurations, for fixed block servers that attach using SCSI or Fibre Channel. Use the S/390 Storage window to configure the LCUs, their associated

volumes, aliases and FICON host adapters for the CKD servers that attach using ESCON or FICON (ESCON ports do not need to be configured).

4.15.2 Logical standard configurations

As an alternative to performing all the configuration steps yourself via the ESS Specialist, the IBM System Support Representative can use the Batch Configuration Tool to perform some of the steps for you. Even if you choose this option, you will still have to use the ESS Specialist to complete the configuration process. This is possible when your logical configuration is according to the standard logical configuration options. See 4.35, “Standard logical configurations” on page 145 for more on this option.

4.16 ESS logical configuration

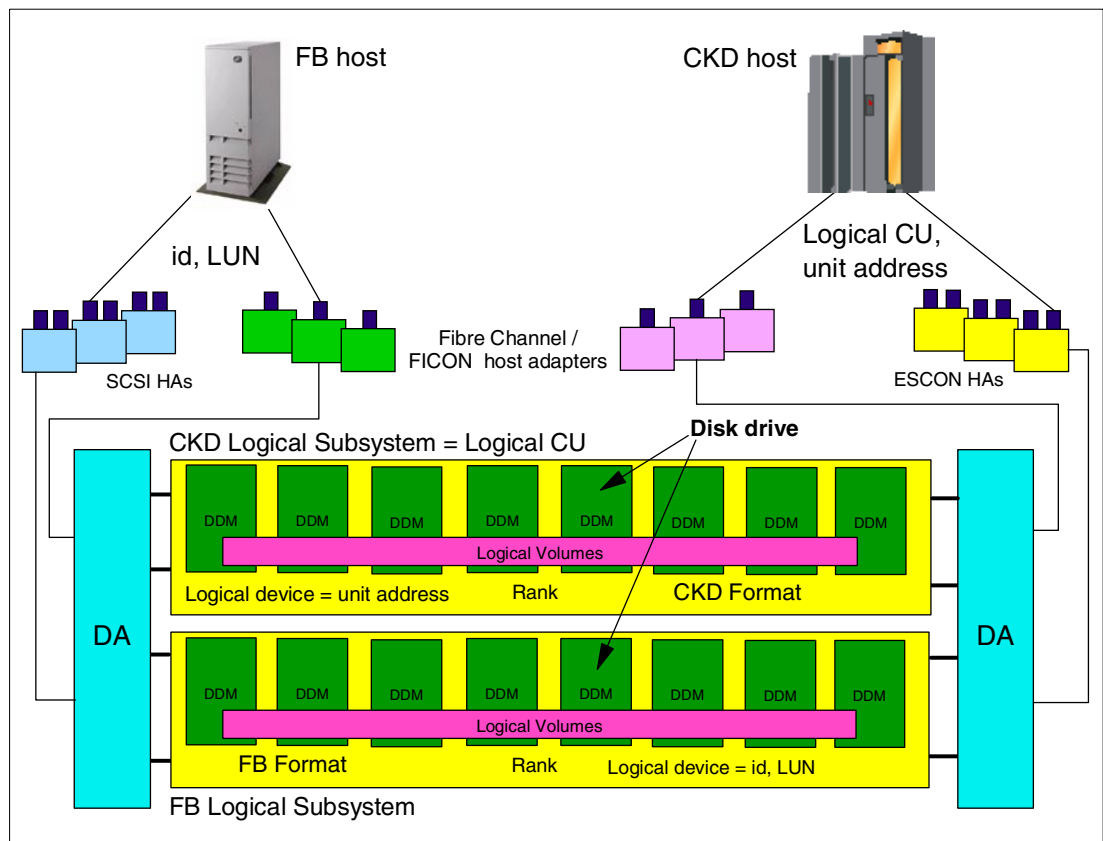


Figure 4-14 Logical configuration terminology

Figure 4-14 illustrates the most frequent terms we use in the following sections when describing the logical configuration. These terms have already been presented in previous chapters, and are illustrated in this figure to help you understand their relationships.

4.17 Logical configuration process

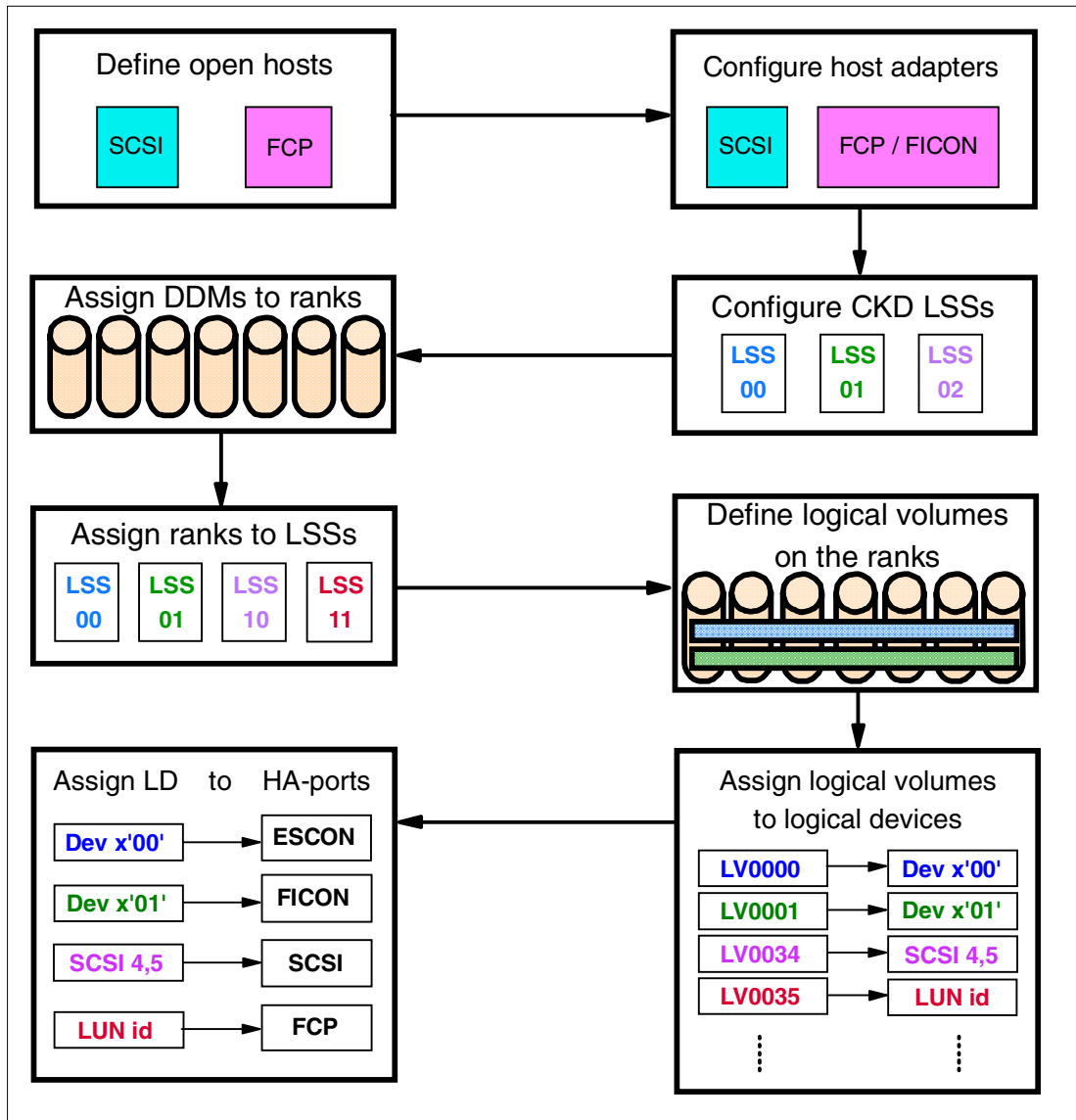


Figure 4-15 Logical configuration process

The diagram in Figure 4-15 provides a basic idea of the logical configuration process. The logical configuration requires that the physical installation of the IBM TotalStorage Enterprise Storage Server has been completed by the IBM System Support Representative. Then the logical configuration is done using either the ESS Specialist only, or also using the Batch Configuration Tool if choosing standard logical configuration options. The basic steps that are done during logical configuration are:

1. Identify the open systems hosts that are attaching to the ESS.
2. Configure the non-ESCON host adapters.
3. Configure LCUs (for CKD LSSs only).
4. Select groups of disk drives (DDMs) to form ranks.
5. Define the ranks as fixed block (FB) or count-key-data (CKD) and assign to the corresponding Logical Subsystems (LSSs).

6. Define logical volumes (LVs) on the ranks.
7. Assign LVs to host logical devices (LDs).
8. Relate LDs with HAs (for SCSI only). CKD LDs will have an exposure to all ESCON and FICON host adapters in the ESS. For Fibre Channel attachment, you assign LDs to HAs when setting the adapter in Access_Restricted mode (using the ESS Specialist).

Some of these steps will not require any action from you. For example, when assigning FB ranks to an LSS, the ESS Specialist will do this. The process of logical configuration is described in the following pages.

Important: For detailed information when doing the logical configuration procedure, refer to the publication *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448. Also the publications *IBM TotalStorage Enterprise Storage Server Configuration Planner for Open-Systems Hosts*, SC26-7477 and *IBM TotalStorage Enterprise Storage Server Configuration Planner for S/390 and zSeries Hosts*, SC26-7476 provide the configuration worksheets to properly plan and document the desired configuration.

4.18 SCSI and Fibre Channel hosts and host adapters

With the Modify Host Systems window of the ESS Specialist, you define to the ESS the attached open system hosts, identifying them by name, operating system type, and type of attachment. Once you have finished with the Modify Host Systems window definitions, then you proceed to configure the host adapters. You do so from the Open Systems Storage window, by clicking the **Configure Host Adapter Ports** button that will take you to the Configure Host Adapter Ports window.

For each SCSI port you define the type of server it will handle. The ESS Specialist will provide a list of hosts that are compatible with the selected bus configuration. This is required to run the correct protocols.

Unlike SCSI, where you link the server host icon to the SCSI host adapter port attached to it, Fibre Channel requires a host icon for every Fibre Channel *host bus adapter* (HBA) installed in the hosts (even if the HBAs are installed in the same host). This is because each Fibre Channel HBA has a unique *worldwide port name* (WWPN), and hosts are defined based on this adapter identification. Figure 4-16 on page 114 shows the ESS Specialist Storage Allocation window, where one host server with two Fibre Channel HBAs appears as two host icons in the first row of the window.

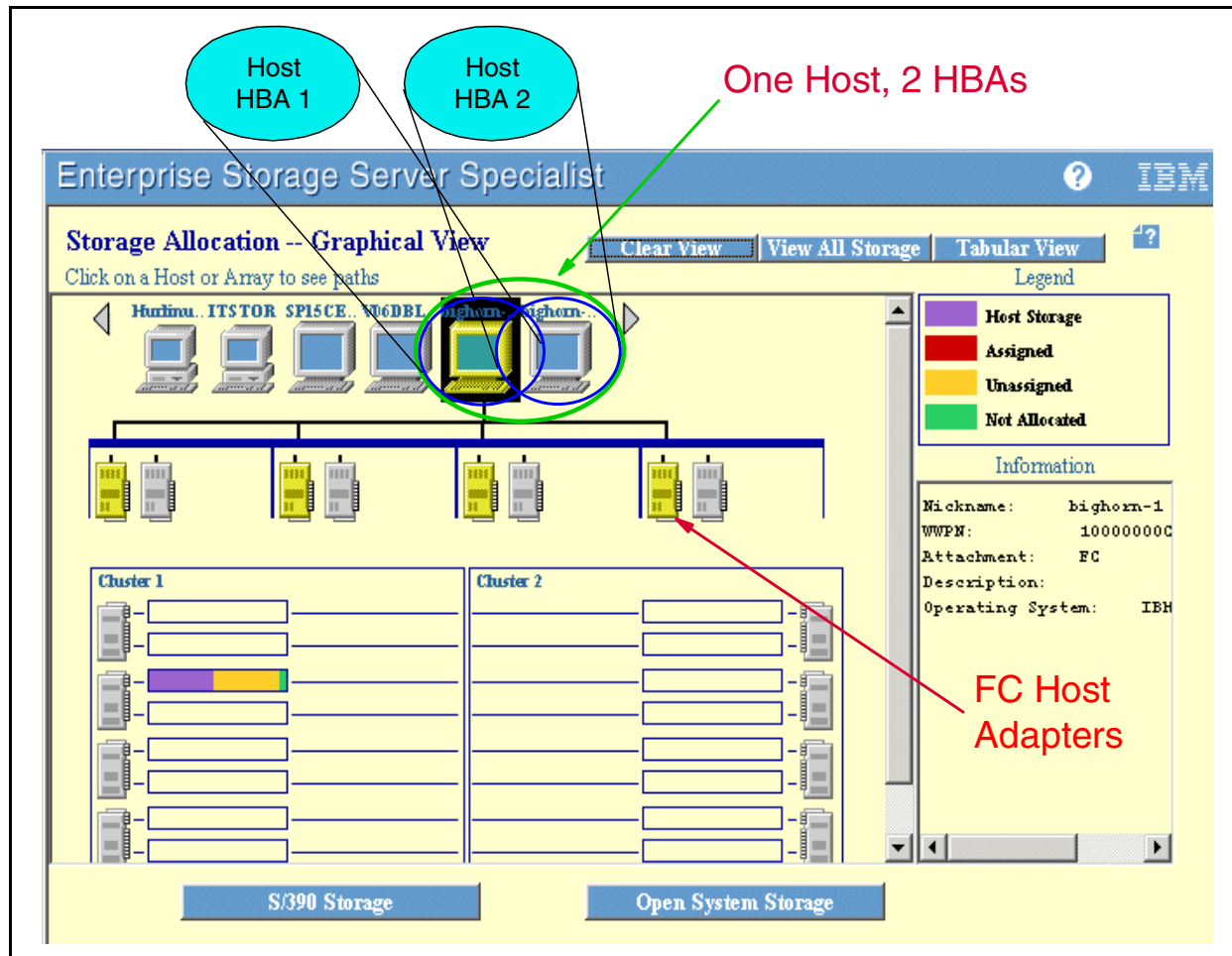


Figure 4-16 Fibre Channel adapters

For SCSI attachments, another important consideration when configuring the ports is the SCSI host initiator IDs used by the host adapters. The default values that ESS Specialist uses are in accordance with the SCSI protocol, which defines SCSI ID 7 as having the highest priority. The SCSI ID priority order is 7–0 then 15–8. The first host system that you add is assigned to SCSI ID 7, the second is assigned to SCSI ID 6. You must verify that these assignments match the SCSI ID setting in each host system SCSI adapter card, and make adjustments to the map of SCSI IDs if necessary.

For Fibre Channel attachments, you will have some different considerations when configuring the host adapter ports. Because Fibre Channel allows any Fibre Channel initiator to access any logical device, without access restrictions, you will have the option to limit this characteristic, or not. The ESS Specialist Configure Host Adapter Ports window, when a Fibre Channel host adapter (HA) has been selected, will allow you to specify the access mode as `Access_Any` mode or `Access_Restricted` mode. The `Access_Restricted` mode allows access to only the host system for which you have defined a profile. The profile limits the access of the host system to only those volumes assigned to the profile. The ESS adds anonymous hosts whenever you configure one or more Fibre Channel host adapter ports in the ESS and set the access mode to `Access_Any`. The `Access_Restricted` mode is recommended because it provides greater security.

For Fibre Channel you also specify the fabric topology to which the port connects. If the topology is Undefined, you can use a drop-down menu to select either **Point-to-Point** or

Arbitrated-Loop. If the topology has already been defined, you will need to reset it to **Undefined** before you can select a new topology.

4.19 ESCON and FICON host adapters

The ESCON and FICON protocols support up to 16 *logical control unit* (LCU) images from x'00' to x'0F' per channel (FICON has the architectural capability of supporting 256 control unit images, but its current implementation is 16 as in ESCON). In other words, any ESCON or FICON channel link arriving at an ESS host adapter port will be able to access any of the 16 logical control units that the ESS has available for CKD hosts: x'00' to x'0F'. These settings are mapped directly to the LSSs IDs, which means that LSS 00 will be logical CU 0, LSS 01 will be logical CU 1, and so on. Access to these LCUs is controlled by means of the host IOCP or HCD definitions.

You don't need to identify the ESCON host adapter ports. The ESCON ports are identified to the ESS when the physical connection between the hosts and the ESS is made.

Figure 4-17 gives an example of the fields that are enabled on the Configure Host Adapter Ports window when you select a Fibre Channel host adapter.

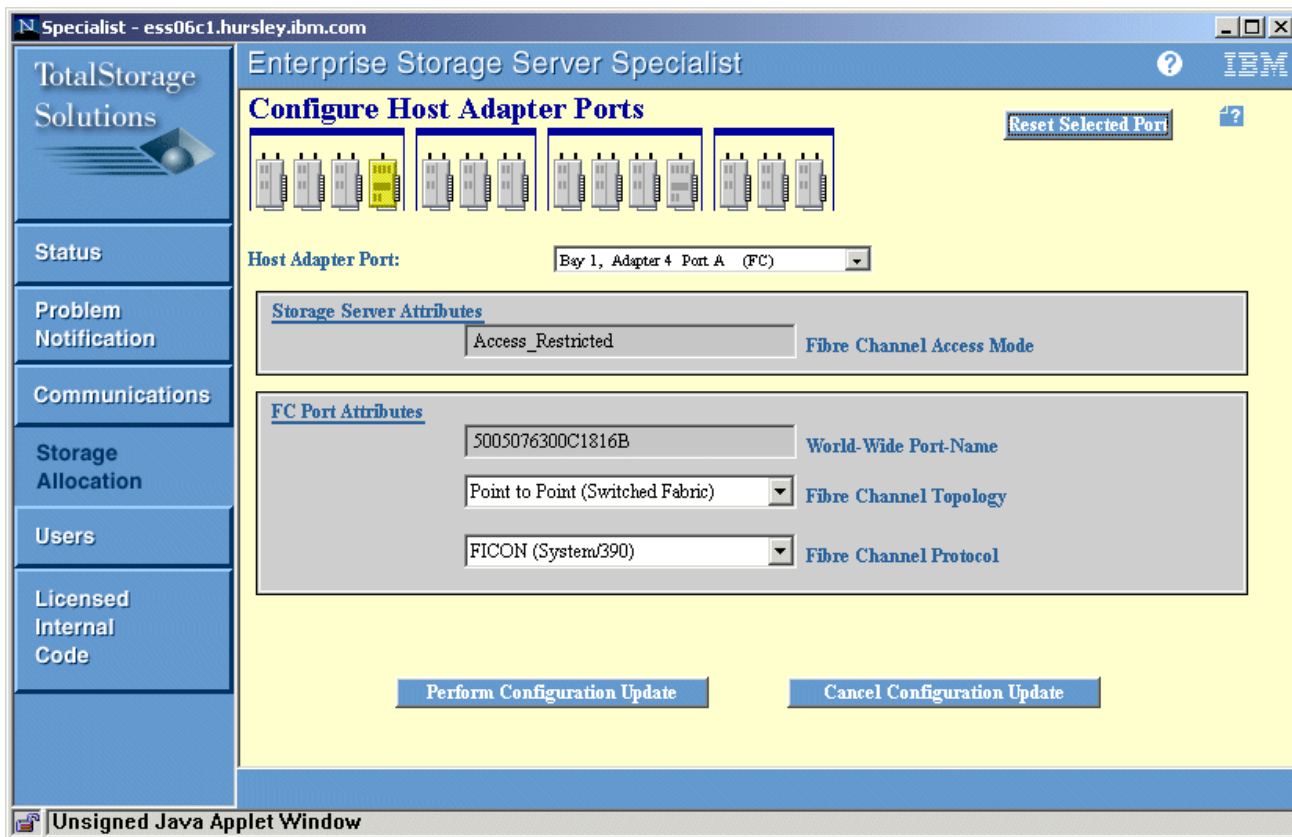


Figure 4-17 Configure Host Adapter Ports window — FICON connection

When you have FICON attachments in your configuration, you will have to configure the ports using the Configure Host Adapter Ports window. This window is similar to the window you use to configure the open systems Fibre Channel ports, but for FICON you arrive here by clicking the **Configure Host Adapter Ports** button from the S/390 Storage window (not from the Open System Storage window).

For unconfigured ports, Point-to-point (Switched Fabric) will be the only choice in the Fibre Channel topology field, when accessing this window from the S/390 Storage window.

The Fibre Channel Protocol field shows the current Fibre Channel protocol for the port you selected. If the topology is Undefined the protocol can be FCP (open systems) or FICON (zSeries). FICON will be your only choice for unconfigured ports, when you access this window from the S/390 Storage window.

If the topology is defined, you must first change the setting to Undefined before the ESS can make an alternate setting available for configuration.

To action your selection, click the **Perform Configuration Update** button shown in Figure 4-17 on page 115.

4.20 Defining Logical Subsystems

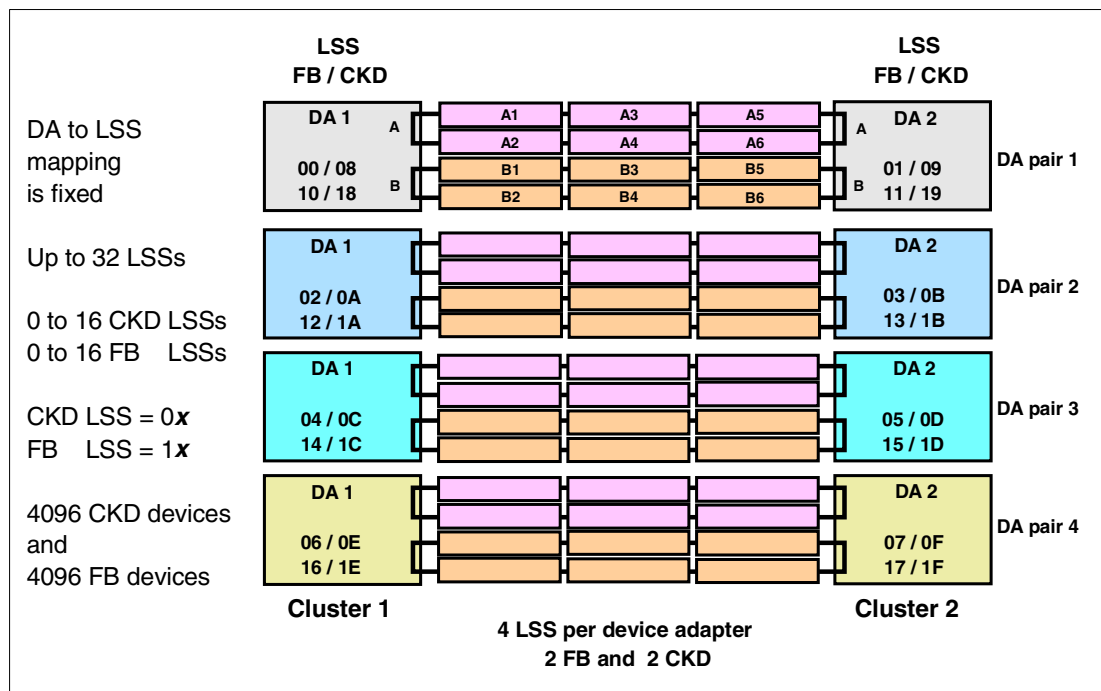


Figure 4-18 Logical Subsystem mapping

Up to 32 LSSs can be configured in an ESS, 16 for CKD servers and 16 for FB servers. Each LSS will get a hexadecimal identifier. LSSs 00x are CKD and LSSs 01x are FB, as Figure 4-18 shows. An LSS can have up to 256 logical devices defined to it. LSSs can map ranks from both loops of the DA pair.

4.20.1 CKD Logical Subsystems

The ESCON and FICON protocols support up to 16 LCU images from x'00' to x'0F'. The LSS concept is very straightforward for zSeries users because LSSs in the ESS map one-to-one the logical control units the zSeries server is viewing.

You will be using the ESS Specialist Configure LCU window to make the CKD LSS definitions (see Figure 4-19 on page 117).

For each control unit image, you must specify its emulation mode. You can choose between the following CU emulations:

- ▶ 3990-6
- ▶ 3990-3
- ▶ 3990-3 TPF

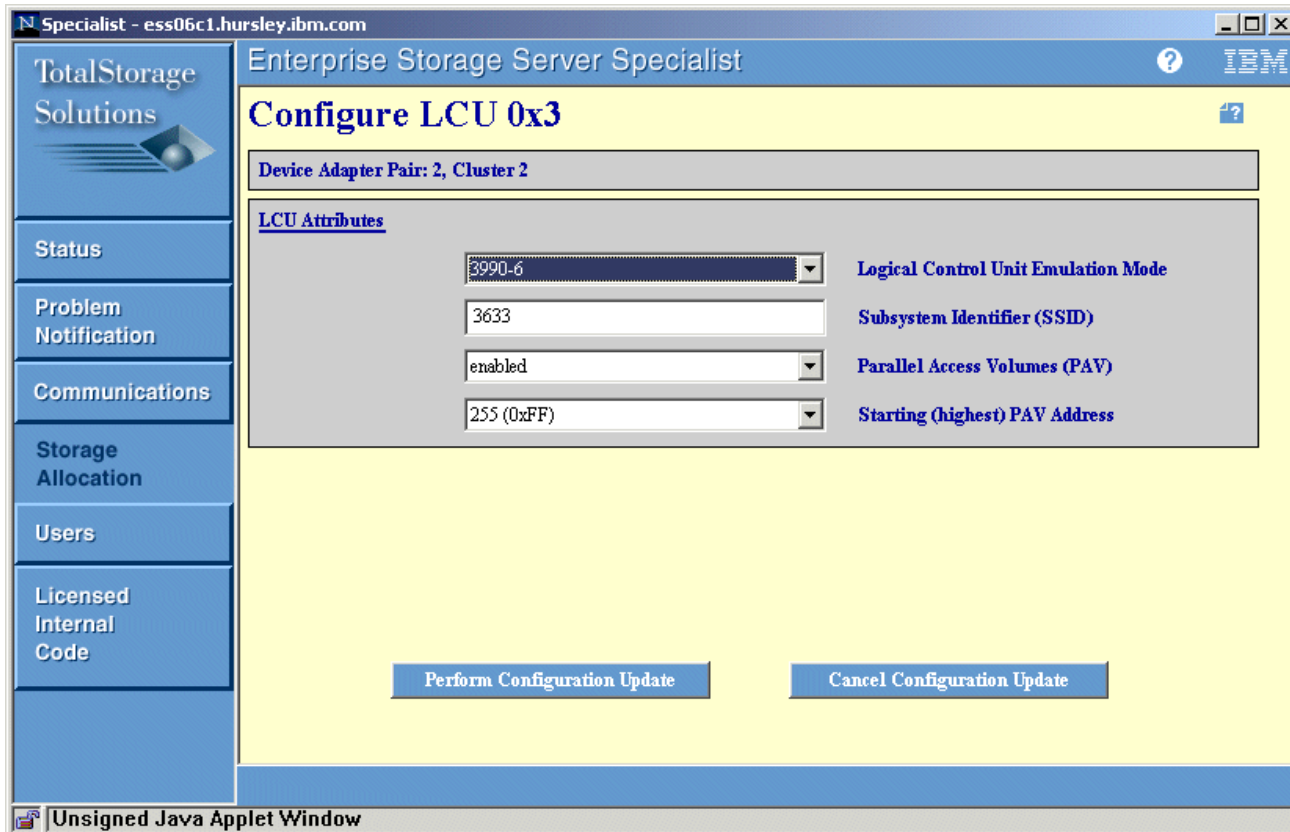


Figure 4-19 Configure LCU window

For each of the configured logical control units, you will need to specify a 4-digit *subsystem identifier* (SSID). This is the usual setting done for a real 3990, and it is required to identify the CU to the host for error reporting reasons and also for such functions as Peer-to-Peer Remote Copy (PPRC). If the Batch Configuration Tool has been used by the IBM SSR, then the SSIDs will already be assigned. If the tool has not been used, then the SSIDs will need to be configured.

Remember that SSIDs must be unique. The system does not allow bringing online a control unit with an already assigned SSID. Users must keep a record of the SSIDs in use, and must assign a unique SSID to each new LCU being configured. By default the Batch Configuration Tool assigns the SSIDs with an *xyxy* format, where *xx* is the two last digits of the ESS serial number, and *yy* can be from 01 to 16 in correspondence to LCUs 01 to 16 (this can be modified to adjust to the installation requirements).

Also in this window you may enable PAV (explained in “Configuring CKD base and alias addresses” on page 129) if your ESS has this optional feature, and you set the PAV starting address. Note that you can nondisruptively update PAV assignments at a later date. Next, you proceed to define the ranks as RAID 5 or RAID 10, as explained later in 4.24, “Configuring CKD ranks” on page 123.

4.20.2 FB Logical Subsystems

As with CKD, the ESS allows you to have either eight or 16 FB LSSs from x'10' to x'1F'. When you choose to have the IBM SSR configure eight LSSs, then each of the four DA pairs in the ESS will have two LSSs, one LSS per DA. For each DA pair, one LSS is assigned to cluster 1 and the other to cluster 2. When your choice is 16 LSSs, then each of the four DA pairs will have four LSSs, two LSSs per DA. For each DA pair, two LSSs are allocated to cluster 1 and two to cluster 2.

When you define the ranks as FB, the FB LSSs will be implicitly configured, and the ESS Specialist will assign the ranks to the LSSs. There is no specific ESS Specialist window for FB LSS configuration.

4.21 Disk groups – ranks

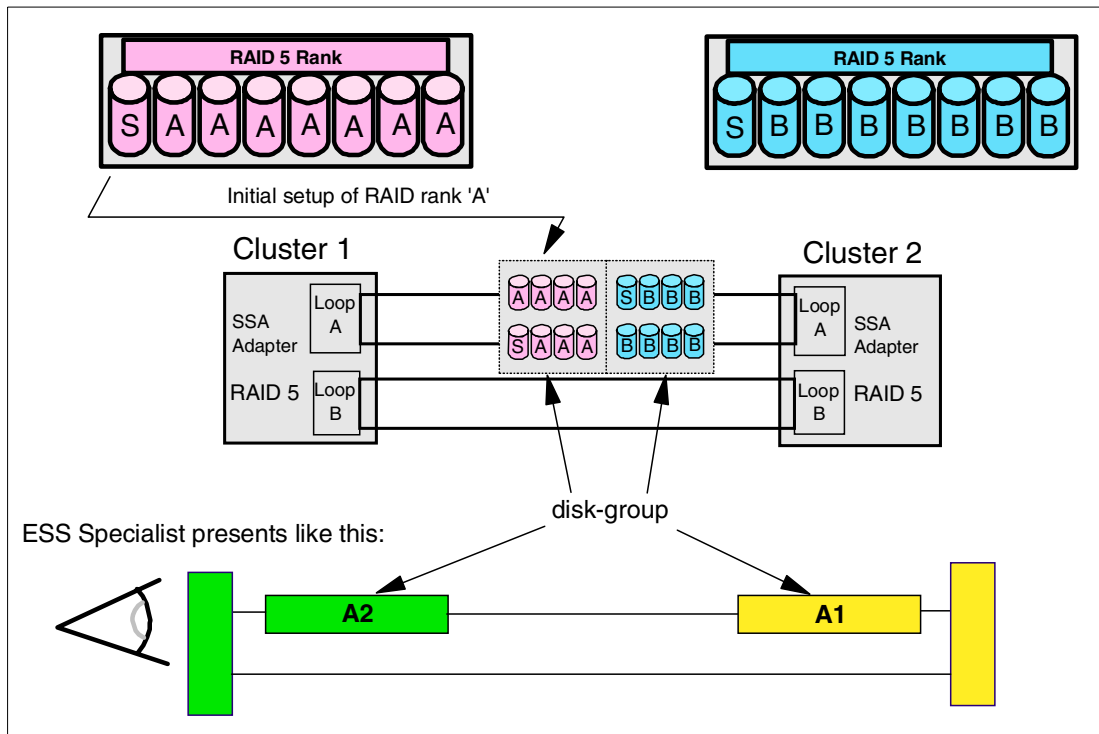


Figure 4-20 Disk groups and RAID ranks — initial setup

Before describing the configuration of CKD and FB ranks, let us review some definitions we have already presented and also present some additional basic concepts that are involved. Please refer to Figure 4-20 and Figure 4-21 on page 119 for the following explanations:

- ▶ The basic units of storage are the *disk drive modules* (DDMs) that hold the *hard disk drives* (HDDs), which we also refer to generically as *disk drives*.
- ▶ Eight DDMs with identical speed (rpm) and capacity are grouped into one *physical assembly* to form an *eight-pack*. The eight-packs must be installed in pairs (of the same capacity) on the same loop. These pairs of eight-packs are the way to add capacity to the ESS.
- ▶ There are two SSA loops (A and B) per device adapter (DA) pair.

- ▶ Each SSA loop can hold two, four, or six eight-packs, if it has some capacity installed on it. Otherwise, it is an unused loop without any eight-packs, and is available for future capacity upgrade.
- ▶ A collection of eight DDMs of the same capacity in the same SSA loop *logically* forms a *disk group*. Initially four DDMs from two different eight-packs in the loop are used. The ESS Specialist refers to disk groups and displays them graphically as whole rectangles. Later, when the spare disks have been used, the DDM set that forms the disk group may differ from the original set. The ESS Specialist will still show the disk group with the same whole rectangle representation as it did to show the initial disk group setup. So a disk group is basically eight DDMs on the same loop that are used to form a RAID array.
- ▶ There can be from zero to 12 disk groups per DA pair (A1 through A6 and B1 through B6).
- ▶ Disk groups A1, A3, A5, B1, B3 and B5 are associated with the device adapter in cluster 2; disk groups A2, A4, A6, B2, B4 and B6 are associated with the device adapter in cluster 1. This is the default association before any ESS Specialist configuration process.
- ▶ Using the ESS Specialist, *ranks* are created from these disk groups by formatting the set of eight DDMs. A rank results from the ESS Specialist formatting process of a disk group: it will become either a RAID 5 rank or a RAID 10 rank. RAID ranks are also known as RAID *arrays*.
- ▶ Each rank is mapped by only one of the DAs.

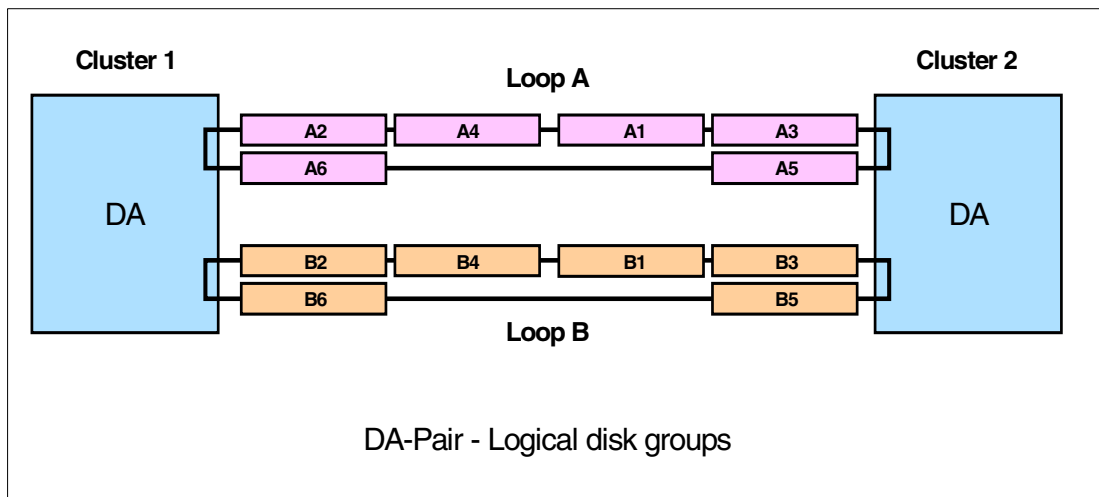


Figure 4-21 Disk group associations

When working with the ESS Specialist to do the logical configuration, you will recognize the following characteristics:

- ▶ The device adapter (DA) to Logical Subsystem (LSS) mapping is a fixed relationship (see Figure 4-18 on page 116).
- ▶ Each one of the DAs of a pair has up to four LSSs that map ranks: two are for CKD ranks and the other two are for FB ranks.
- ▶ For a pair of DAs, any of its LSSs can map ranks from either of its two loops (FB LSSs map FB ranks, and CKD LSSs map CKD ranks).
- ▶ Logical volumes (up to 256 per LSS) are created on these ranks.
- ▶ Each *logical volume* (LV) is part of a rank. The RAID rank is the whole of a disk group. This disk group is mapped in an LSS, which is associated with a single DA in a DA pair.

The assignment of disk groups to an LSS is made in the context of the corresponding addressing architectures, either fixed block architecture (FB) or count-key-data (CKD) architecture. Also this disk group assignment to the LSSs is made in the context of the attachment that will be used (SCSI, FCP, ESCON, or FICON).

As already discussed, ranks can either operate in RAID 5 mode or RAID 10 mode. The RAID 5 arrays will have the following setup:

- ▶ **6+P+S:** This setup is to reserve spares in the loop. Because 16 drives are the minimum configuration for a loop, whenever configuring for RAID 5, the first two arrays in the loop will be 6+P, leaving two drives in the loop as spares.
- ▶ **7+P:** This is the setup for all the other ranks in the loop you are configuring as RAID 5.

While the RAID 10 arrays will have the following setup:

- ▶ **3+3+2S:** This setup is to reserve spares in the loop. Whenever configuring for RAID 10, the first array in the loop will be three data + three mirror data, leaving two drives in the array as spares.
- ▶ **4+4:** This is the setup for all the other ranks in the loop you are configuring as RAID 10.

This assumes that all eight-packs installed in the loop are of the same capacity. The key point is that a loop must contain at least two spare DDMs for each DDM capacity present in the loop (see Figure 4-11 on page 106 for an example of mixed capacities). Hence if a loop contains four eight-packs of 36.4 GB DDMs and two eight-packs of 72.8 GB DDMs, there will be at least two 36.4 GB DDM spares and at least two 72.8 GB DDM spares. See 4.22, “RAID 5 and RAID 10 rank intermixing” on page 120 for an explanation of drive intermixing considerations.

4.22 RAID 5 and RAID 10 rank intermixing

As explained in 2.6.5, “Disk intermixing” on page 27, there are restrictions on the mixing of drive capacities and speeds (rpm) within an ESS or a loop. The following list summarizes these restrictions:

- ▶ All DDMs of a given capacity in an ESS must be of the same speed — for example, all 36.4 GB DDMs must be either 10,000 rpm or 15,000 rpm.
- ▶ All DDMs within an eight-pack are of the same capacity and hence speed.
- ▶ A loop must have zero, two, four or six eight-packs installed on it, and every eight-pack pair must be of matching capacity.
- ▶ Every loop must contain at least two spare DDMs for each DDM capacity present in the loop.

Each disk group is formatted independently within a loop to form either a RAID 5 or RAID 10 rank. As such, it is possible to define within an eight-pack pair both a RAID 5 rank and a RAID 10 rank. In this case, the sequence in which the two ranks are formatted will determine the number of spare DDMs allocated (assuming these are the first two eight-packs of a given capacity formatted within the loop).

As Figure 4-22 on page 121 shows, formatting the RAID 10 array first will be done as 3+3+2S, which provides the required two spares for the loop. Hence the RAID 5 array is formatted as 7+P.

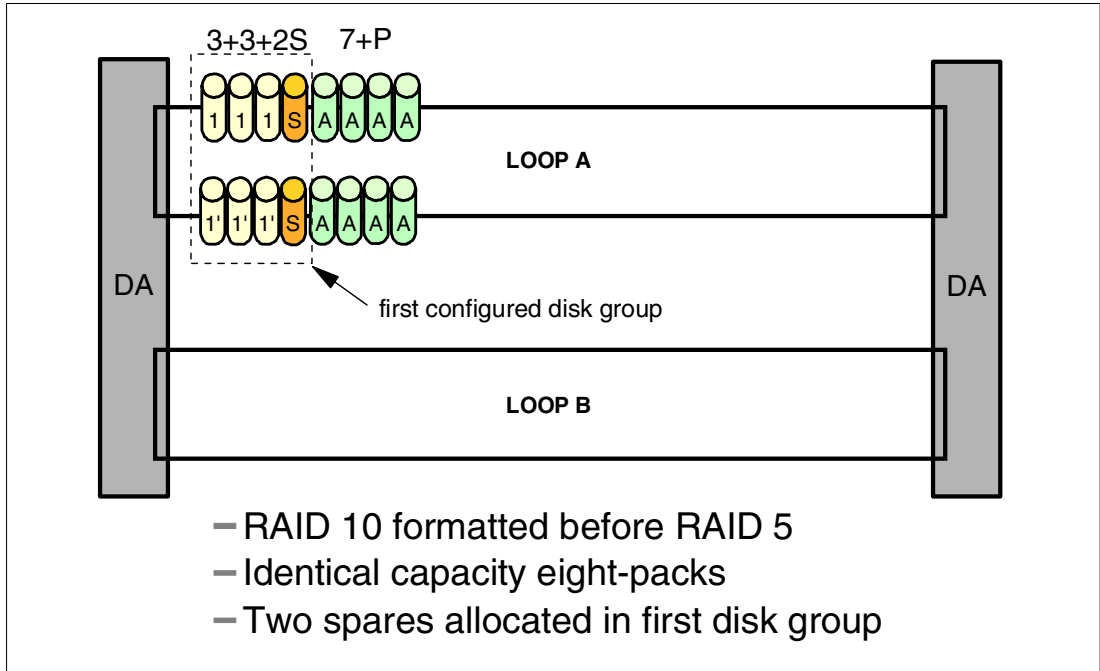


Figure 4-22 RAID 10 followed by RAID 5 formatting

Figure 4-23 shows that formatting the RAID 5 array first will be done as 6+P+S, which only provides one spare, and hence the RAID 10 array has to be formatted as 3+3+2S. This results in three spare DDMs in the loop.

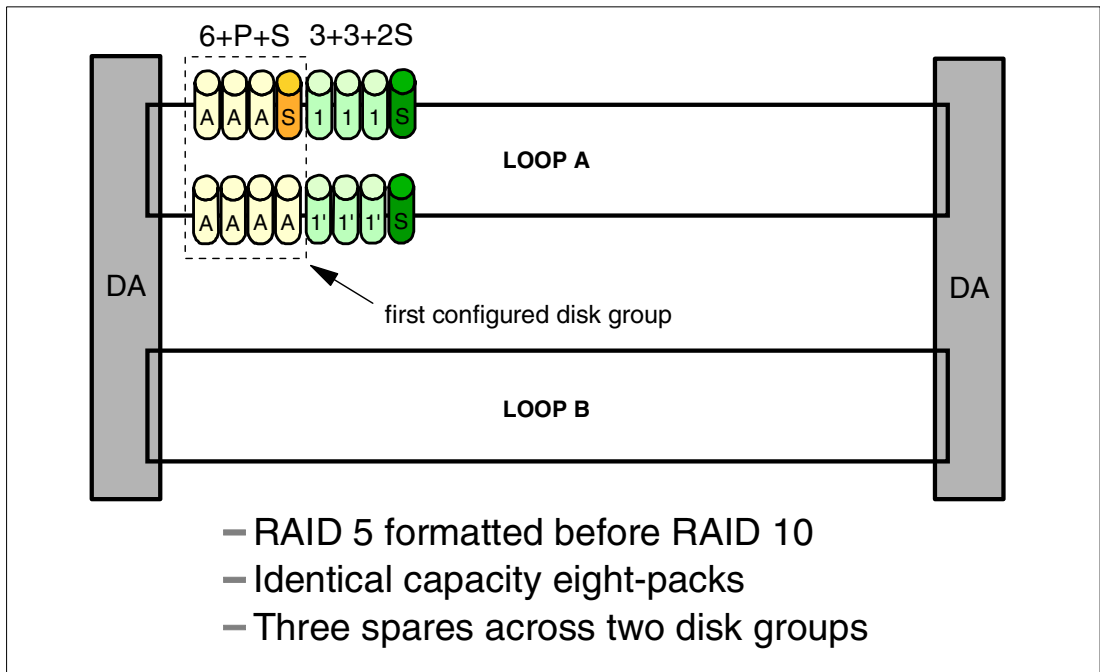


Figure 4-23 RAID 5 followed by RAID 10 formatting

Tip: To avoid “wasting” a DDM as an extra spare, consider carefully the sequence in which you format the arrays.

Remember that these considerations must be repeated for each set of eight-pack capacities within a loop.

4.23 Balancing RAID 10 ranks

Normally you will want to balance storage allocation within an ESS across clusters and SSA loops. Since the first RAID 10 array in a loop will provide the required two spare DDMs (3+3+2S), subsequent RAID 10 arrays in the same loop will be 4+4. This means the first array will have three DDMs for data and the others will have four DDMs for data. This can result in LSSs with unbalanced capacity for some loop configurations.

Unbalanced LSSs

Figure 4-24 shows the potential problem, where the disk groups are allocated in the numbered sequence. LSS 0 has ended up with two arrays of 3+3+2S, while LSS 1 has two arrays of 4+4, and hence has more usable capacity.

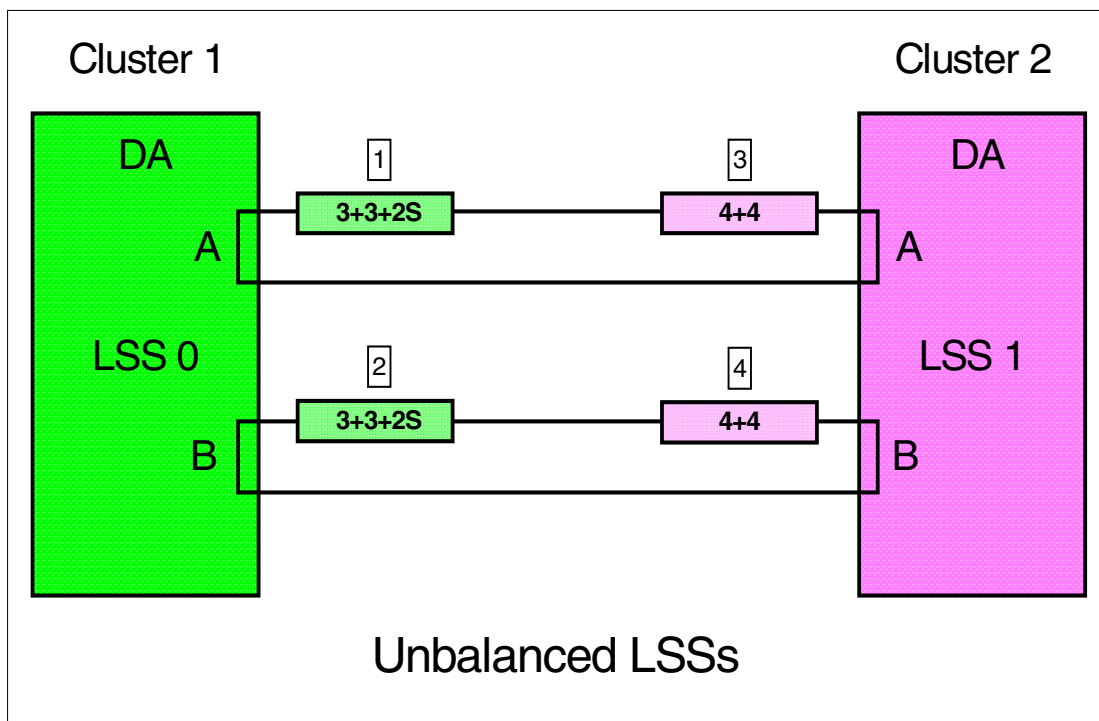


Figure 4-24 Unbalanced RAID 10 arrays

Balanced LSSs

To avoid this, allocate an initial RAID 10 array to each LSS within the DA-pair from different loops before allocating subsequent arrays to an LSS, using the following process:

1. Add the first array for LSS 0 to loop A. The ESS creates the array as 3+3+2 (three pairs of mirrored drives and two spares).
2. Add the first array for LSS 1 to loop B. The ESS creates the array as 3+3+2.
3. Add the second array for LSS 0 to loop B. The ESS creates the array as 4+4 (four pairs of mirrored drives).
4. Add the second array for LSS 1 to loop A. The ESS creates the array as 4+4.

This results in the configuration illustrated in Figure 4-25 on page 123.

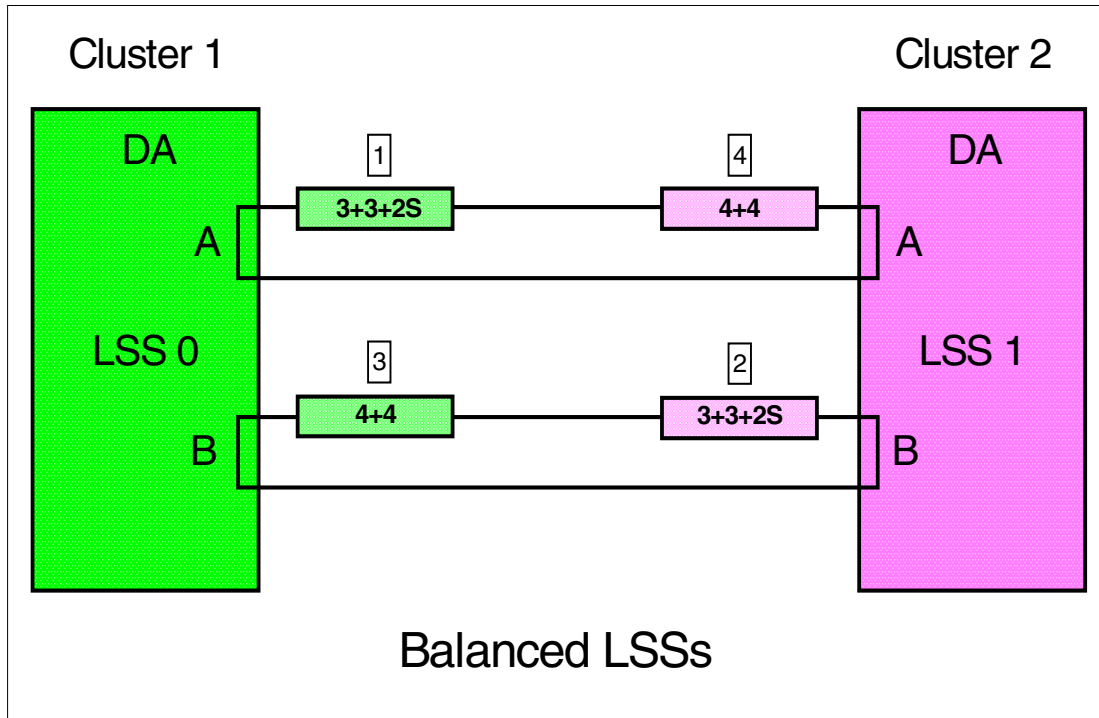


Figure 4-25 Balanced LSSs

This assumes that you only have one LSS per cluster in the DA-pair. If you have four LSSs per DA-pair, then you should still ensure that capacity is spread evenly between clusters and loops, but it will not be possible to have the same capacity in all four LSSs since only two of the RAID 10 disk groups will be 3+3.

4.24 Configuring CKD ranks

The CKD arrays can either be RAID 5 or RAID 10, and 3390 or 3380 track format. When working with the Configure Disk Groups window shown in Figure 4-26 on page 124, you select from the list of available disk groups displayed the ones you wish to associate with the LCU you are configuring. Then for the selected disk groups, you select the storage type as either **RAID 5 Array** or **RAID 10 Array**, and the Track Format as either **3390 (3380 track mode)** or **3390 (3390 track mode)**.

For RAID 10 ranks, or RAID 5 ranks formed from DDMs of at least 72.8 GB, the Standard volumes to auto-allocate pull-down is disabled, and you should proceed to defining logical volumes by clicking the **Perform Configuration Update** button.

For RAID 5 arrays formed from DDMs smaller than 72.8 GB, the Standard volumes to auto-allocate pull-down will be enabled, and you have the option at this stage to block allocate some standard-sized logical volumes. If you only need 3390-2 or 3390-3 or 3390-9 standard volumes, then this is the best way to allocate them as they will be allocated in *interleaved* mode. Interleaved mode causes the volumes to be striped in groups of four across the rank, but you are restricted to allocating all volumes of the same standard size. If you use this option, then after the last block of four volumes in the *interleaved partition*, there will be some leftover space of at least 5000 cylinders known as the *non-interleaved* partition. In here you will be able to allocate some extra volumes of whatever size will fit. If you do not use this option, then the array is created with the specified track format and zero logical volumes. You should now proceed by clicking the **Perform Configuration Update** button.

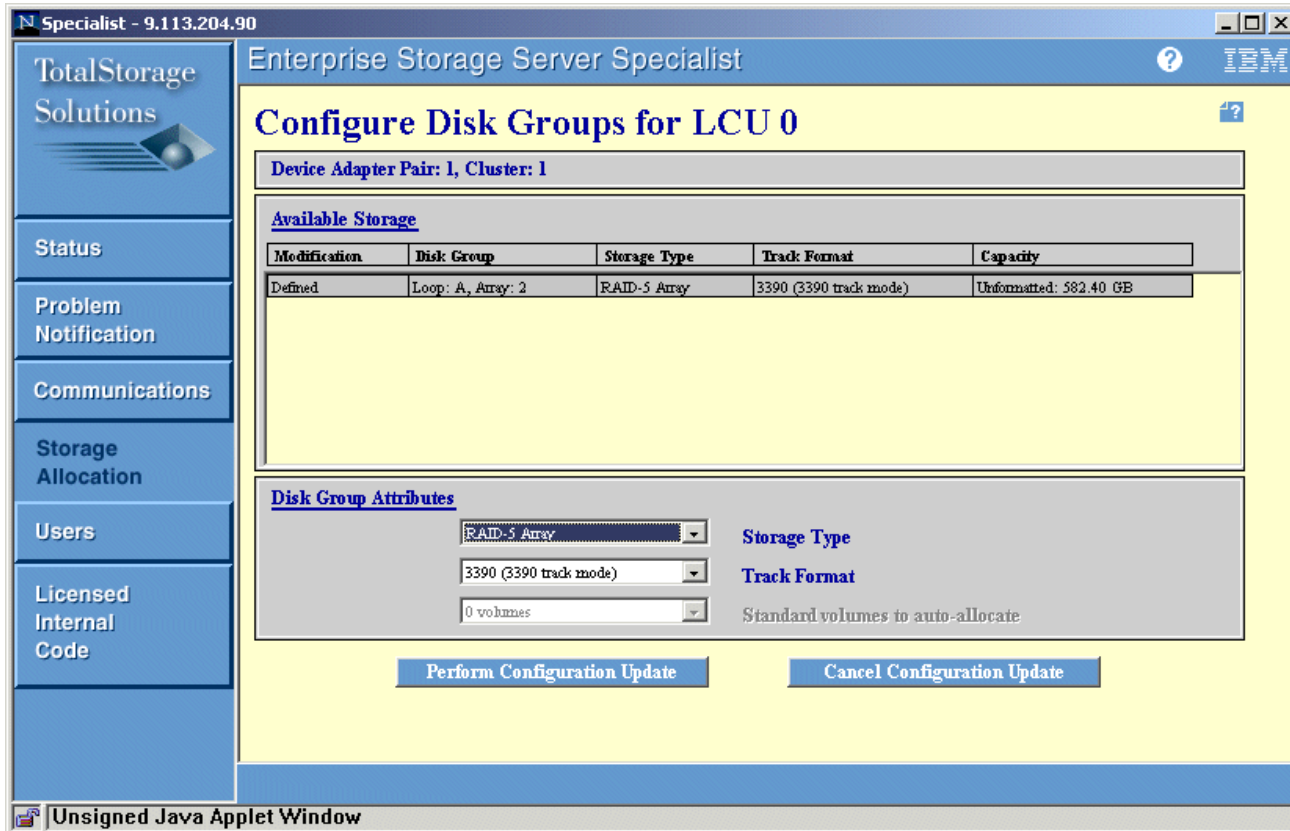


Figure 4-26 Configure Disk Group window

Important: Any modification of volumes (change of size or removal) in an array will require a redefinition of the disk group within the LSS, which is disruptive for the data. The data has to be moved or backed up before modification and restored afterwards.

You can add extra ranks to an LSS without affecting the existing data of that LSS.

4.25 Configuring FB ranks

In this step the fixed block ranks, from the disk groups available on the loop, are configured. The ESS Specialist will automatically assign the ranks to the LSSs in the DA pair. If you have defined eight LSSs in total, then there will be one LSS in each DA. If you have defined 16 LSSs in total then there will be two LSSs in each DA. This has implications for FlashCopy as discussed earlier in “0/8/16 LSSs” on page 79.

The ranks that are being formatted on even-numbered disk groups are assigned to an LSS belonging to cluster 1, while the ranks that are being formatted on odd-numbered disk groups are assigned to an LSS belonging to cluster 2. The allocation of ranks to LSSs is dependent upon the number of LSSs, the number of ranks, and the number of logical volumes that have been created. Although 16 LSSs may have been defined, the ESS Specialist will start assigning the ranks to the first of the two LSSs in the DA, and will start assigning ranks to the second LSS only when the ranks assigned to the first LSS contain a total of 192 or more logical volumes (remember that each LSS can contain a maximum of 256 logical volumes).

FB ranks can either be RAID 5 or RAID 10. You select the disk groups to configure and specify for them the storage type as either RAID 5 or RAID 10 and the track format as FB. You do this definition for all the disk groups you want to format, using the Fixed Block Storage window of the ESS Specialist shown in Figure 4-27. Then you finish the rank definitions by clicking the **Perform Configuration Update** button.

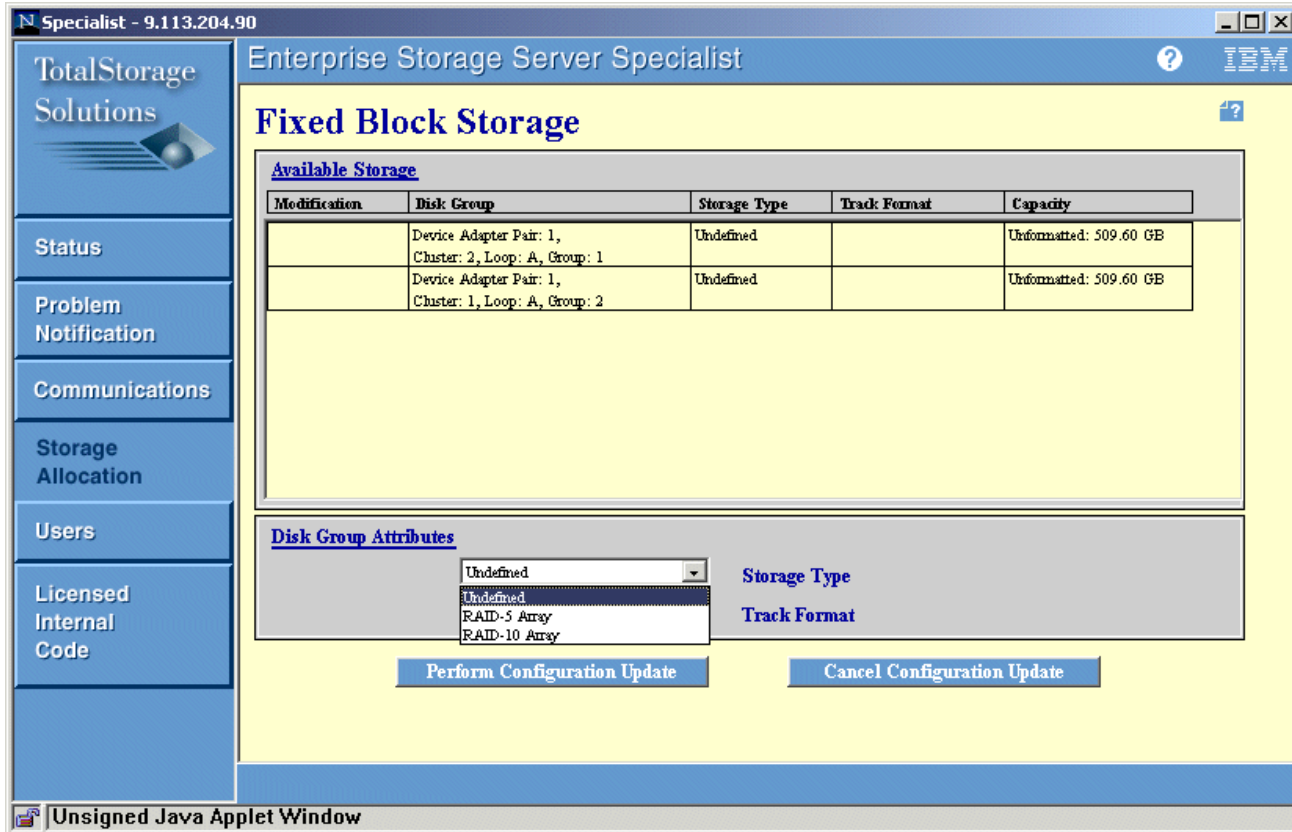


Figure 4-27 Fixed Block Storage window

Important: Changes to the logical volume structure on an array will require the whole array to be reformatted. For example, if you have two 16 GB logical volumes and you want to make them into one 32 GB logical volume, then the whole array has to be reformatted. This will require any data on that array to be backed up first and all volumes on that array to be deleted and redefined, including the host system assignments, before reloading the data.

You can add extra ranks to an LSS without affecting the existing data of that LSS.

4.26 Assigning logical volumes to a rank

Once the ranks have been set up, you can start defining logical volumes (LV) on them. For both FB and CKD ranks, you have the alternative of having the IBM System Support Representative use the Batch Configuration Tool that will define the logical volumes to the LSSs. This will happen if you choose the standard logical configuration options.

4.26.1 Rank capacities

See 2.6.4, “Disk eight-pack capacity” on page 26 for the effective usable capacity of the different rank formats and DDM sizes. This information can be used for this section to calculate how many volumes will fit in a RAID array.

4.26.2 Adding CKD logical volumes

Once the array type and format has been configured, or after auto-allocating standard volumes, you are ready to use the Add Volumes window shown in Figure 4-28 to define your logical volumes. All volumes are allocated on the rank in the order of creation, and may be of any size from 1 to 32,760 cylinders. Since the volumes may be of non standard size, they are known as *custom volumes*. When adding volumes you must plan how many volumes, and of what size, will fit the available space.

To build the list of logical volumes you wish to define, you first select a storage type and track format from the choices with available capacity. Then you select the number of cylinders for the volume and a multiple for the number of volumes, and click the **Add>>** button to add your selection to the New Volumes list. You can repeat these steps to build up a list of new volumes of varying sizes until you run out of free capacity. As you use capacity by adding volumes to the list, the available capacity figures at the top of the window will decrease. Figure 4-28 shows an example where all available capacity has been used to define custom 3390 Model 9 volumes of 32,760 cylinders each, but there was only sufficient space left for the last volume to be 13,837 cylinders.

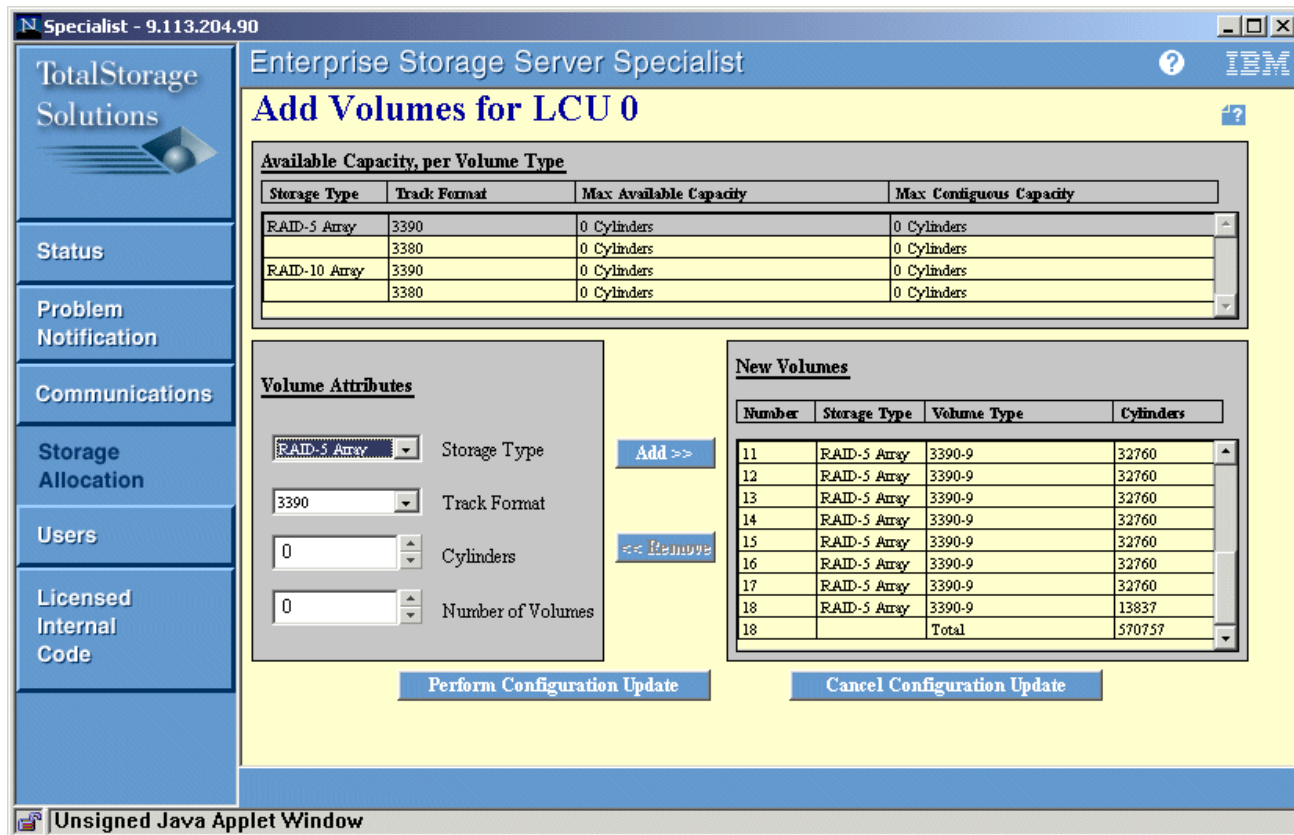


Figure 4-28 Add CKD Volumes window

A CKD custom volume will allow you to break down a data set or VM minidisk to a single logical volume. The advantage of this is that the data set will have a dedicated volume for it, which will result in less logical device contention.

Refer to Appendix A, “S/390 Standard Logical Configuration”, in the publication *IBM TotalStorage Enterprise Storage Server Configuration Planner for S/390 and zSeries Hosts*, SC26-7476, for the capacities available when you configure the CKD ranks. This appendix shows, for each type of CKD logical volume, how many volumes will fit in the rank’s interleaved and non-interleaved partitions. Normally you do not need to calculate these numbers, because when assigning CKD LVs to a rank, the configuration process will give you information about the available and remaining capacities in the rank you are configuring.

Table 4-2 shows the CKD logical device capacities, along with the physical capacity that is allocated in the rank, when defining the logical volumes. As you can see, the physical capacity allocated in the rank is slightly greater than the logical device capacity. This difference reflects the structure the ESS uses as it allocates space for logical volumes.

Table 4-2 CKD logical device capacities

Logical device type	Cylinders	Bytes per cylinder	Logical device capacity (GB)	Physical capacity used (GB)	524 byte sectors per cylinder
3390-2	2,226	849,960	1.892	1.962	1,680
3390-3	3,339		2.838	2.943	
3390-3 custom (1)	1 - 3,339		0.00085 - 2.838	0.00176 - 2.943	
3390-9	10,017 (2)		8.514	8.828	
3390-9 custom (1)	3,340 - 32,760 (2)		2.839 - 27.845	2.944 - 28.868	
3390-2 (3380)	2,226	712,140	1.585	1.821	1,560
3390-3 (3380)	3,339		2.377	2.731	
3390-3 (3380) custom (1)	1 - 3,339		0.00071 - 2.377	0.00163 - 2.731	
Notes:					
1. A CKD volume that has a capacity different from that of a standard 3390 device type is referred to as a custom volume.					
2. In an interleaved partition, the number of cylinders is 10,017. In a non-interleaved partition, the number of cylinders may be from 3,340 to 32,760.					

The following formulas allow you to approximate the physical capacity of a logical CKD device. The formulas do not completely reflect the algorithms of the ESS as it allocates space for a logical volume.

- ▶ The amount of physical capacity for 3390 devices can be approximated by:
Capacity = (((Nr. of cyls. + 1) * Bytes per cyl. * 524) / 512) * 1.013 * 10⁻⁹ GB
- ▶ The amount of physical capacity for 3380 devices can be approximated by:
Capacity = (((Nr. of cyls. + 1) * Bytes per cyl. * 524) / 512) * 1.122 * 10⁻⁹ GB

The equations compensate for any overhead in the logical device, such that the result is always greater than or equal to the physical capacity required to configure the logical device.

4.26.3 Assigning iSeries logical volumes

For the iSeries servers the logical volume sizes match the 9337 or 2105 devices with sizes of 4.19 GB, 8.59 GB, 17.54 GB, 35.17 GB, 36 GB, or 70.56 GB. With SCSI attachment, the LUNs will be reported to the iSeries as device type 9337. With Fibre Channel attachment, the LUNs will be reported to the server as device type 2105. The model will depend on the LUN sizes defined.

See the “Additional open-systems host information” section of the *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448 for complete details of the supported iSeries volumes.

Protected versus non-protected

With its RAID architecture, the ESS emulated 9337 and 2105 volumes are treated as protected logical volumes (9337 models that end in C), which prohibits software mirroring. To solve this, the ESS permits disk units to be defined as non-protected models (9337 models that end in A). Software mirroring is only allowed on non-protected 9337s and 2105s. From an ESS perspective, all iSeries volumes are defined on RAID ranks and are protected within the ESS. The ESS Specialist Add Volumes window allows you to define the volume as Unprotected.

For additional considerations when attaching external storage to the iSeries, refer to *IBM @server iSeries in Storage Area Networks: A Guide to Implementing FC Disk and Tape with iSeries*, SG24-6220.

4.26.4 Assigning fixed block logical volumes

For FB ranks, from the ESS Specialist Open Systems Storage window you click the **Add Volumes** button and that takes you to the first Add Volumes window where you select the server, the port, and the fixed block ranks to which the logical volumes will be available. The selections in this window take you to the second Add Volumes window, where you define the volume sizes and number of volumes. Once you click the **Perform Configuration Update** button, the ESS will define the logical volumes within the fixed block rank. The server will see these logical volumes as logical devices.

Logical volumes for open systems servers can have an LV size from 100 MB to the full effective rank capacity, in increments of 100 MB. This granularity of LUN sizes enables improved storage management efficiencies, especially for Windows NT systems that have a limited number of LUNs and therefore require full exploitation of the rank capacity.

Refer to Appendix A, “Open-systems standard configuration”, in *IBM TotalStorage Enterprise Storage Server Configuration Planner for Open-Systems Hosts*, SC26-7477 for the number of standard size LUNs that will fit in different rank capacities.

4.27 Defining CKD logical devices

The CKD logical volumes are mapped into a logical device map in a CKD LSS. Figure 4-29 on page 129 shows that the logical devices in such an LSS represent the device address of the logical volume, ranging from x'00' to x'FF'. Because each LSS is seen as a logical control unit, the zSeries systems will see it as a 3990-x with up to 256 devices. The logical devices (LD) need not be mapped to the ESCON or FICON host adapters, because the zSeries hosts have access to all the LDs through any of the ESCON or FICON connections available. The set of logical devices accessed by any S/390 image is defined with the *Hardware Configuration Definition* (HCD) in the IODF file that the operating system uses to recognize its hardware topology.

ESCON-attached and FICON-attached hosts are identified to the ESS when you make the physical connection between the hosts and the storage server. These hosts are grouped as single nets (EsconNet and FiconNet) in the ESS Specialist for improved graphical presentation.

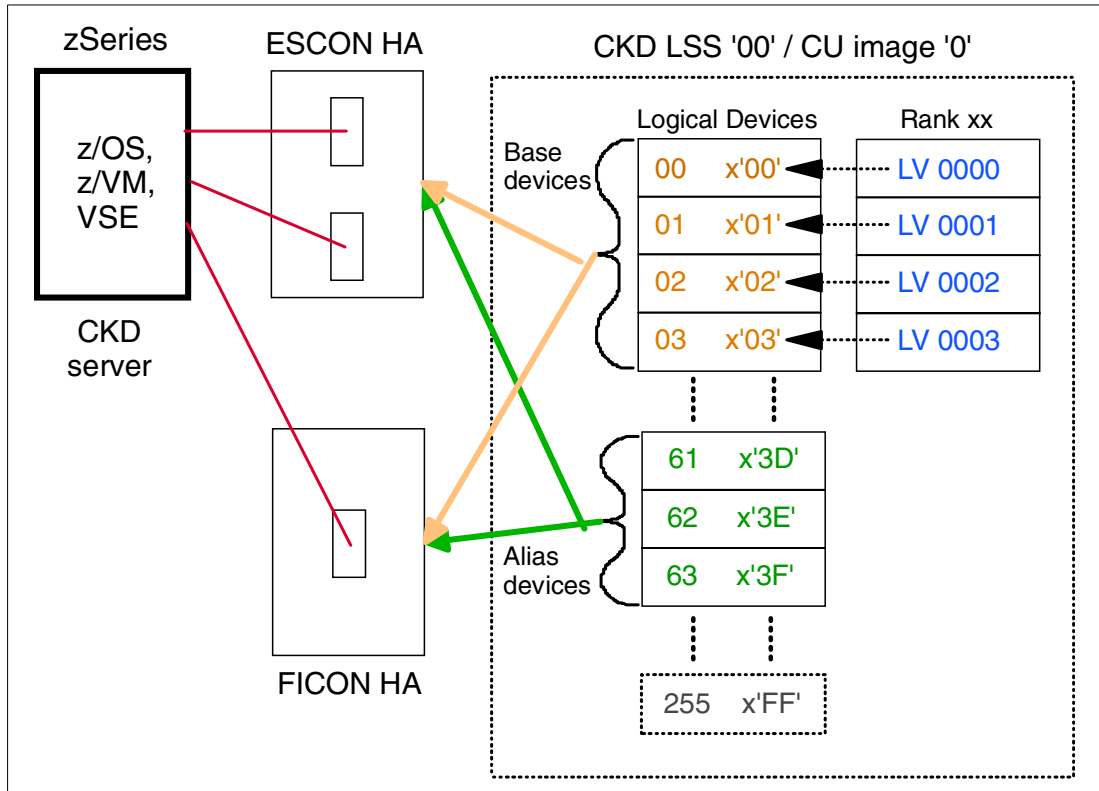


Figure 4-29 CKD logical device mapping

Configuring CKD base and alias addresses

Since the ESS also supports alias addresses for *Parallel Access Volumes (PAV)* with the z/OS and OS/390 environments (see 6.2, “Parallel Access Volume” on page 166 for details), for these environments you can specify two types of logical devices:

- ▶ Base devices, for primary addressing from the host
- ▶ Alias devices, as an alternate UCB to a base device

At least device x'00' must be a base device. The ESS is capable of having up to 4096 (16 LSSs x 256 devices) devices configured. The ESCON channel can handle 1024 devices, and the FICON channel can handle 16,384 devices. Considering that the ESS gives you the capability of having larger volumes sizes (up to 32,760 cylinders) for easier administration, still with good response times, you may not need to address all 256 devices on a 3990 CU image. Defining fewer base and alias addresses will also save space in memory, since fewer UCBs will be created. In any case, the volumes you define in the ESS must match your definition in the HCD for the operating system.

You set a starting address for PAVs using the ESS Specialist (see Figure 4-19 on page 117). The base devices are assigned upwards from the lowest address in the order of creation. Alias devices are assigned downwards from the PAV starting address. Base and alias addresses are defined both in the ESS Specialist and also in the zSeries IOCP IODEVICE macro specifying UNIT=3390B and UNIT=3390A respectively. The ESS and the ESS Specialist

allow you to define from zero to 255 aliases per base device address. The maximum devices (alias plus base) is 256 per LCU.

Figure 4-29 on page 129 shows a CKD storage map and an example of how base and alias devices are mapped into a 64 device address range (when `UNITADD=(00,64)`) in the `CNTLUNIT` macro of the IOCP definition). The figure shows the possible 256 logical devices available in the LSS, but which are not all defined.

Note: The intermixing of ESCON and FICON channels on one control unit, as shown in Figure 4-29 on page 129, is only supported for migration purposes. It is not a recommended configuration for a production environment as explained in 4.34, “ESCON and FICON connectivity intermix” on page 144.

4.28 Defining FB logical devices

Logical devices (LDs) are the way for the host to access the already defined logical volumes. Each logical volume will automatically receive an LD identification that will be used by the host to access that volume. Remember that each LSS supports up to 256 logical devices.

4.28.1 SCSI attached hosts

For FB logical volumes that are accessed by a SCSI host, you must set the SCSI targets of the host adapters. The SCSI target and the LUN ID of the devices are assigned by the ESS. The ESS can assign up to 64 LUNs per target, but not all host operating systems and host SCSI adapters are able to support 64 LUNs per target, so the number of LUNs the ESS will allow to be assigned to each target will depend on the host type. These SCSI-attached host types will be known to the ESS, since they must have already been defined when using the Modify Host Systems window.

When you first define the FB logical volumes, you are simultaneously making them accessible to a host and to a SCSI port. This way you are relating the logical volumes to the logical devices view (target, ID LUN) of the host. Once each of the logical volumes has a logical device and SCSI port assignment, you can map the logical devices to additional SCSI ports using the ESS Specialist Modify Volume Assignment window. Doing this will result in shared logical volumes. It is the host application’s responsibility to handle shared logical volumes. These definitions may be of interest if you wish to configure for high availability. The Subsystem Device Driver (SDD) program, which comes with the ESS, allows for the handling of shared logical volumes. SDD runs on the host system side (See 5.8, “Subsystem Device Driver” on page 157 for further information).

Figure 4-30 on page 131 shows the SCSI port assignment and logical device mapping that occurs when initially defining the logical volumes in the FB ranks, using the ESS Specialist Add Fixed Block Volumes window. The pSeries server is utilizing SDD to view LDs 00 – 03 in LSS 10 via two SCSI adapters, each connected to a different ESS port, while the iSeries server has one SCSI adapter connected to another port on the ESS with access to LDs 04 – 06 in LSS 10. The pSeries is addressing LDs 00 – 03 via SCSI target ID 14 with LUNs 0 – 3, while the iSeries is addressing LDs 04 – 06 via SCSI target ID 6 with LUNs 0 – 2.

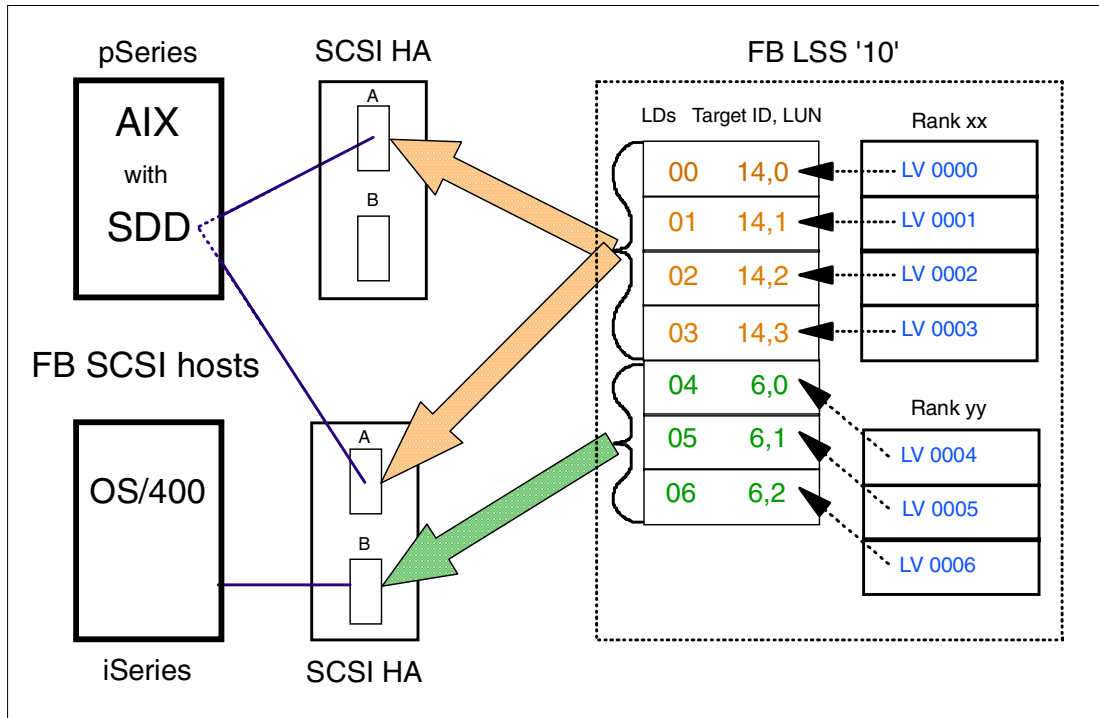


Figure 4-30 FB logical device mapping for SCSI attached hosts

The ESS supports a full SCSI-3 protocol set, and because of that, it allows the definition of up to 64 LUNs per target. However, not every host operating system can support 64 LUNs per target. For information on host LUN support, see the appropriate chapter in *IBM TotalStorage Enterprise Storage Server Host System Attachment Guide, SC26-7446*.

4.28.2 Fibre Channel-attached hosts

For FB logical volumes that are made accessible to Fibre Channel-attached hosts, things are different from SCSI. In SCSI, the LDs are assigned based on SCSI ports, independent of which hosts may be attached to those ports. So if you have multiple hosts attached to a single SCSI port (ESS supports up to four hosts per port), all of them will have access to the same LDs available on that port.

For Fibre Channel, the LD affinity (LUN affinity) is based on the WWPN of the HBA in the host, independent of which ESS Fibre Channel HA port the host is attached to (see “LUN affinity” on page 74 for further discussion on LUN affinity).

Figure 4-31 on page 132 shows the logical device assignment to the FB logical volumes for Fibre Channel-attached hosts when initially defining the logical volumes in the FB ranks, using the ESS Specialist Add Fixed Block Volumes window. The pSeries server is addressing LDs 00 – 03 in LSS 10 via one FC HBA connected to an ESS FC port, while the xSeries server also has one HBA connected to an FC port on the ESS with access to LDs 04 – 06 in LSS 10. Depending on how zoning is configured in the FC switch, each host may see its LDs only via one FC HA or via both FC HAs. In the latter case, host software such as SDD (see 5.8, “Subsystem Device Driver” on page 157) should be used to manage the multiple views.

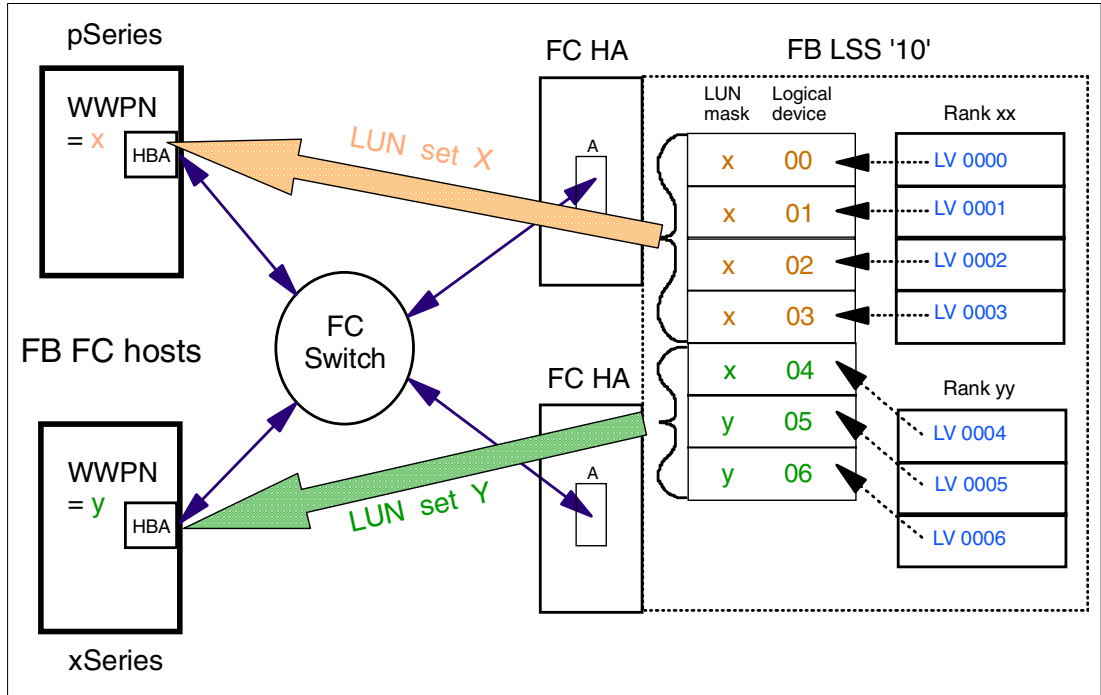


Figure 4-31 FB logical device mapping for Fibre Channel-attached hosts

These FB logical volumes mapping to the host-viewed logical devices (LDs) result from the definitions you do using the ESS Specialist, when initially working with the Add Fixed Block Volumes window. Previous to this step, you already identified the host, its Fibre Channel attachment, and the Fibre Channel host adapter WWPN when using the ESS Specialist Modify Host Systems window.

Remember that Fibre Channel architecture allows any Fibre Channel initiator to access any open system logical device without access restrictions. However, the default mode for the ESS is Access_Restricted, which means that no server can access a LUN in the ESS unless it has been defined to that LUN via its WWPN. If you want to change from Access_Restricted to Access_Any, then the IBM SSR has to change this parameter via the service window and the ESS has to be rebooted (see 4.18, "SCSI and Fibre Channel hosts and host adapters" on page 113).

4.29 LSS/ranks configuration example

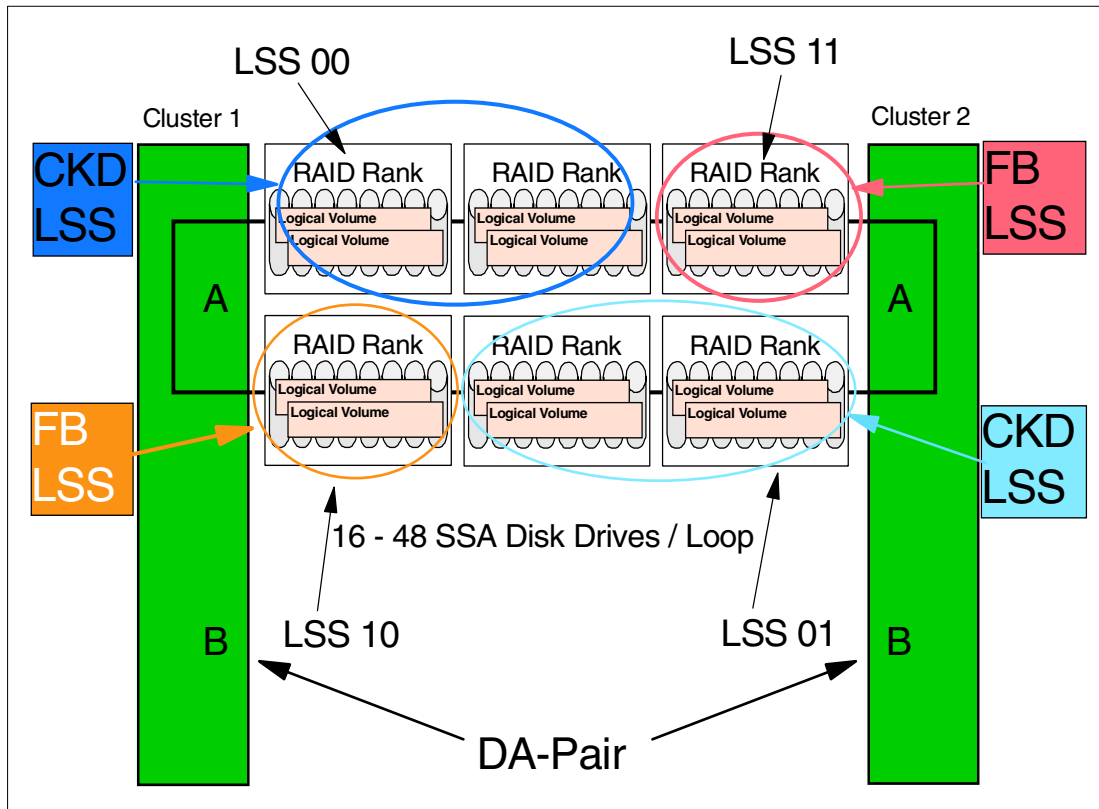


Figure 4-32 LSS ranks assignment

In the example in Figure 4-32, you can see how a final logical configuration of a loop may look. In this case, loop A has the maximum possible (48) drives installed. The loop has been configured with six RAID 5 ranks (RAID 10 ranks could also have been used). A single DA pair loop can have up to four LSSs assigned to it: two CKD LSSs and two FB LSSs. Assuming that this example shows the first DA pair, then the LSSs defined are:

- ▶ DA Cluster 1 Loop A: CKD LSS 00 (CU image 0)
 - Two RAID ranks
- ▶ DA Cluster 1 Loop A: FB LSS 10
 - One RAID rank
- ▶ DA Cluster 2 Loop A: CKD LSS 01 (CU image 1)
 - Two RAID ranks
- ▶ DA Cluster 2 Loop A: FB LSS 11
 - One RAID rank

4.30 SCSI host connectivity

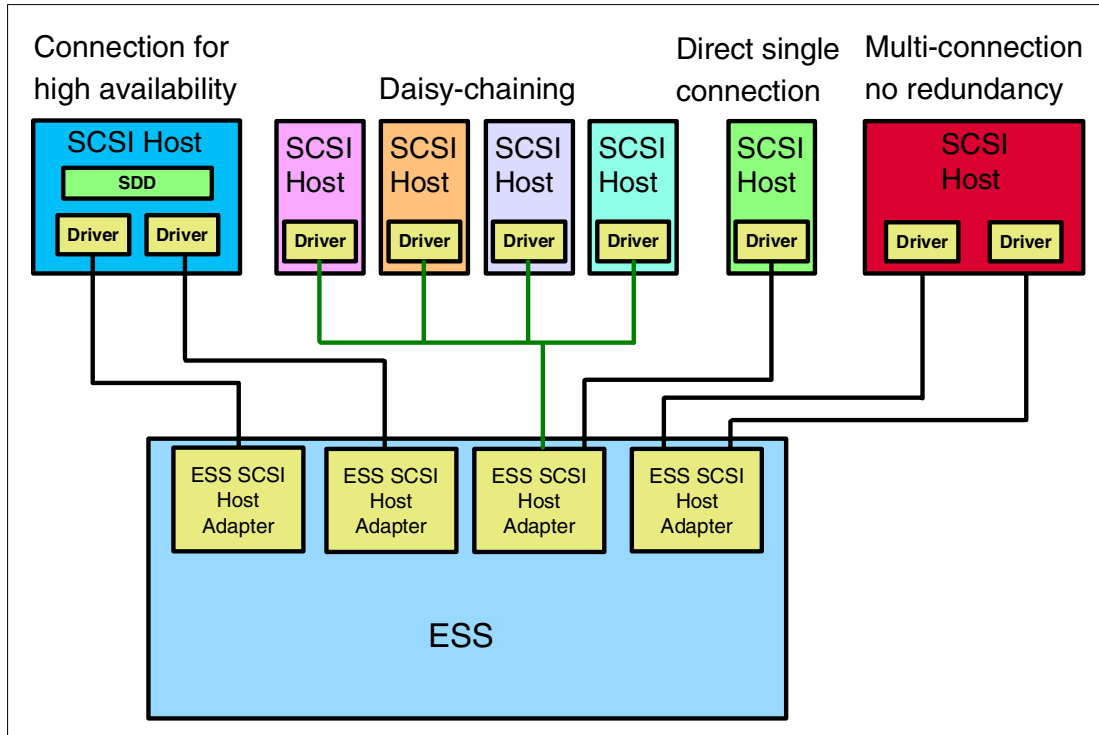


Figure 4-33 SCSI connectivity

4.30.1 Single host connection

Figure 4-33 shows different possibilities for attaching SCSI hosts to the IBM TotalStorage Enterprise Storage Server. The simplest of these possible attachments is the single host connected to only one SCSI host adapter port. In this type of connection, the server has only one path to its logical volumes in the ESS. If the path fails, then the server loses all access to its data because no redundancy was provided.

4.30.2 SCSI connection for availability

For availability purposes, you can configure logical devices in the ESS as shared devices. To do that, you must assign them to two different SCSI ports in the ESS. This allows you to connect your host to two (or more) separate SCSI host adapter ports (preferably located in different HA bays), both seeing the same set of shared logical devices. You can then use the IBM Subsystem Device Driver, which comes standard with the ESS to distribute the I/O activity among the SCSI adapters in the host, and it will automatically recover failed I/Os on an alternate path. This is valid for any cause of connection failure, such as SCSI interface failures, SCSI host adapter failures, or even ESS host adapter port failures. Another advantage of using the SDD is the capability of having concurrent maintenance of the SCSI host adapter cards. In such cases, SDD offers commands that allow you to deactivate the I/Os through a specific adapter and return it back to operation once the maintenance action has finished. One last consideration of the SDD benefits is that it will automatically balance the I/O over the available paths, hence improving overall server I/O performance. See 5.8, “Subsystem Device Driver” on page 157 for more details on the Subsystem Device Driver.

4.30.3 Multi-connection without redundancy

Figure 4-33 on page 134 also illustrates a multi-connection setup without path redundancy. This connection can be done by having multiple SCSI adapters in the host and having each SCSI adapter connected to a different SCSI port in the ESS and not sharing the logical volumes. But in this case, having no SDD software in the server to fail over to the alternate path, if one SCSI adapter fails then all the logical volumes associated with it become unavailable. The absence of SDD also means there will be no I/O load balancing between SCSI adapters.

4.30.4 Daisy-chaining host SCSI adapters

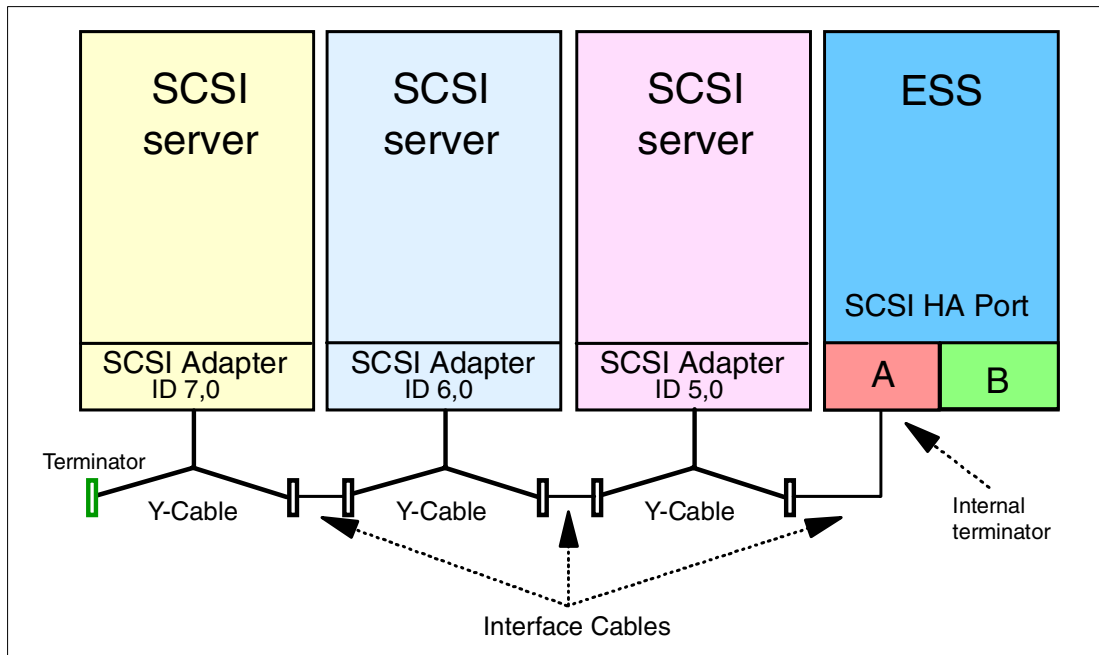


Figure 4-34 SCSI daisy-chaining

As you can see from Figure 4-34, the IBM TotalStorage Enterprise Storage Server allows daisy-chaining of several host adapters. Although it is not the most efficient connection, whenever you need to do this, follow these rules:

- ▶ A maximum of four host initiators is recommended on a single ESS host adapter SCSI port. The SCSI ID priority order is 7 – 0 then 15 – 8. The first host system that you add is assigned to SCSI ID 7, while the second is assigned to SCSI ID 6. You must verify that these assignments match the SCSI ID setting in each host system SCSI adapter card, and make adjustments to the map of SCSI IDs as necessary.
- ▶ The number of hosts that can be daisy-chained is host and adapter dependent. You should check with the server adapter provider for this support. The iSeries does not allow daisy-chaining with the adapters used to connect the external 9337 devices, as the interface is not designed for that.
- ▶ The SCSI adapters are daisy-chained with Y-Cables. Both ends of the cables must be terminated. The ESS must be at one end of the interface, because it has internal terminators on the SCSI host adapter cards.
- ▶ Avoid mixing host SCSI adapters of different types in the chain. The best results are obtained when running the chain with the same type of adapter.

- ▶ The cables must be 2-byte differential SCSI cables and must match the requirements for the host SCSI adapters. For more details about the supported host SCSI adapters, see the Web site:
<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>
- ▶ When more than one server is daisy-chained from the ESS, the length of the cables in the chain must be added together and the sum must not exceed 25 meters (82 feet). This includes the length of the cable branches (Y-cables) to each server. Queries on the Y-cable requirements should be referred to the provider of the server adapter to which the ESS will attach, since daisy-chaining support varies from vendor to vendor.
- ▶ Daisy-chaining should be avoided whenever possible because it creates an overhead of SCSI arbitration on the interface, which may result in performance degradation. Note that this is a SCSI limitation and not an ESS limitation.

Remember that when the ESS is daisy-chained to multiple servers, all of these SCSI servers on the same bus can address any LUN (logical device) defined on that port of the ESS.

4.31 Fibre Channel host connectivity

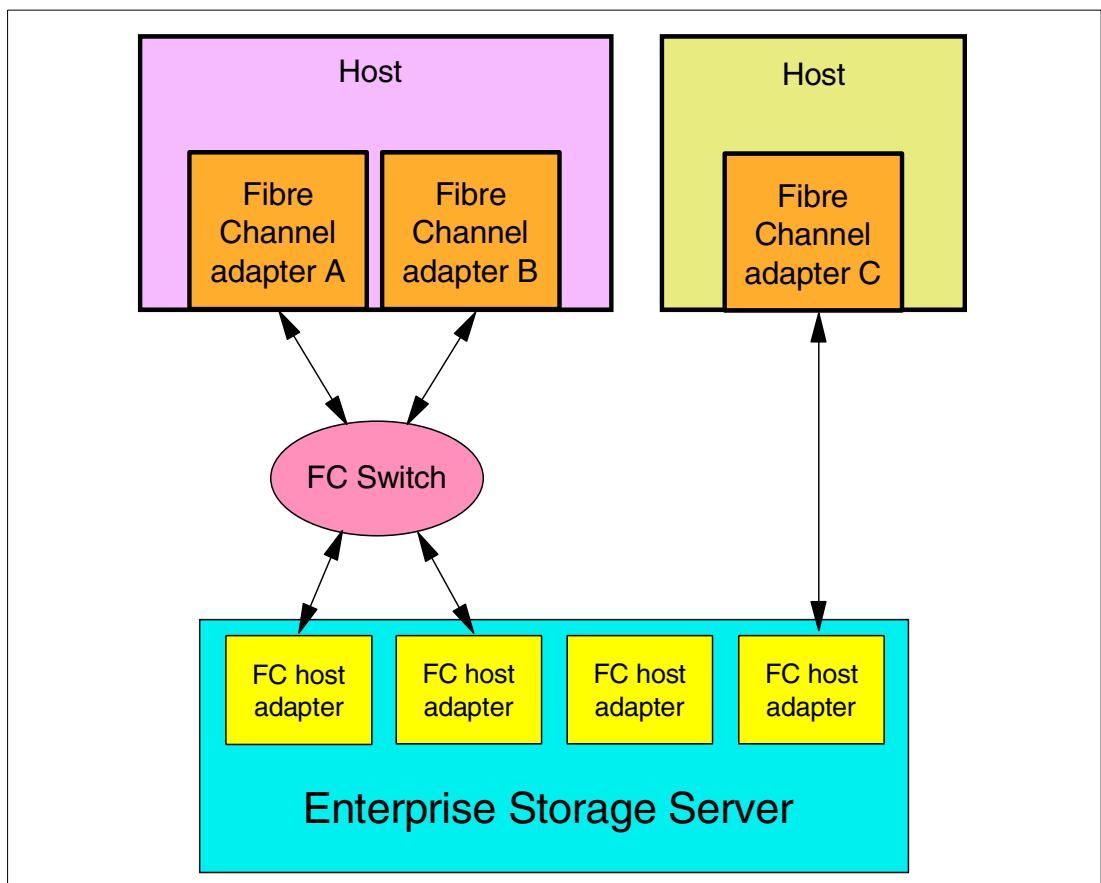


Figure 4-35 Fibre Channel connectivity

With Fibre Channel, the attachment limitations seen on SCSI in terms of distance, addressability, and performance are overcome. Fibre Channel is not just a replacement of parallel SCSI by a serial-based interface, but is more the ability to build *Storage Area Networks* (SANs) of interconnected host systems and storage servers.

4.31.1 Fibre Channel topologies

Three different topologies are defined in the Fibre Channel architecture. All of the three topologies are supported by the IBM TotalStorage Enterprise Storage Server. The three topologies are discussed briefly below.

Point-to-point

This is the simplest of all the topologies. By using just a fiber cable, two Fibre Channel adapters (one host and one ESS) are connected. Fibre Channel host adapter card C in Figure 4-35 on page 136 is an example of a point-to-point connection. This topology supports the maximum bandwidth of Fibre Channel, but does not exploit any of the benefits that come with SAN implementations. When using the ESS Specialist to connect directly to a host HBA, set the ESS Fibre Channel port attribute dependent on the host HBA configuration.

Arbitrated Loop

Fibre Channel Arbitrated Loop (FC-AL) is a uni-directional ring topology very much like token ring. Information is routed around the loop and repeated by intermediate ports until it arrives at its destination. If using this topology, all other Fibre Channel ports in the loop must be able to perform these routing and repeating functions in addition to all the functions required by the point-to-point ports. Up to a maximum of 127 FC ports can be interconnected via a looped interface. All ports share the FC-AL interface and therefore also share the bandwidth of the interface. Only one connection may be active at a time, and the loop must be a private loop.

When using the ESS Specialist to configure a loop, always use Arbitrated Loop as the Fibre Channel port attribute. Specifically for the iSeries, remember that it has to connect via FC-AL and that the ESS Fibre Channel port cannot be shared with any other platform type.

Switched fabric

Whenever a switch or director is used to interconnect Fibre Channel adapters, we have a switched fabric. A switched fabric is an intelligent switching infrastructure that delivers data from any source to any destination. Figure 4-35 on page 136, with Fibre Channel adapters A and B, shows an example of a switched fabric. A switched fabric is the basis for a Storage Area Network (SAN). When using the ESS Specialist to configure a switched fabric, always use Point to Point as the Fibre Channel port attribute, as this is the protocol used in fabrics.

Refer to the following document for further information on Fibre Channel topologies:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essfcwp.pdf>

The distance between the host and the ESS depends on the speed of the host adapters (currently 2 Gb or 1 Gb) and whether short-wave or long-wave host adapters are being used. Long-wave adapters support greater distances than short-wave, while higher link speeds reduce the maximum distance. See Table 2-2 on page 43 for the actual distances.

4.31.2 Fibre Channel connection for availability

In Figure 4-35 on page 136, the attachment for the host that contains Fibre Channel adapters A and B does not alone provide for redundant access to the data. Besides configuring the LUNs to *both* Fibre Channel HBAs, in order for the host to take advantage of both paths, you must also run the Subsystem Device Driver (SDD) program that is distributed with the ESS. This program runs in the host, and besides giving automatic switching between paths in the event of a path failure, it will also balance the I/Os across them.

See 5.8, “Subsystem Device Driver” on page 157 for further information.

4.32 ESCON host connectivity

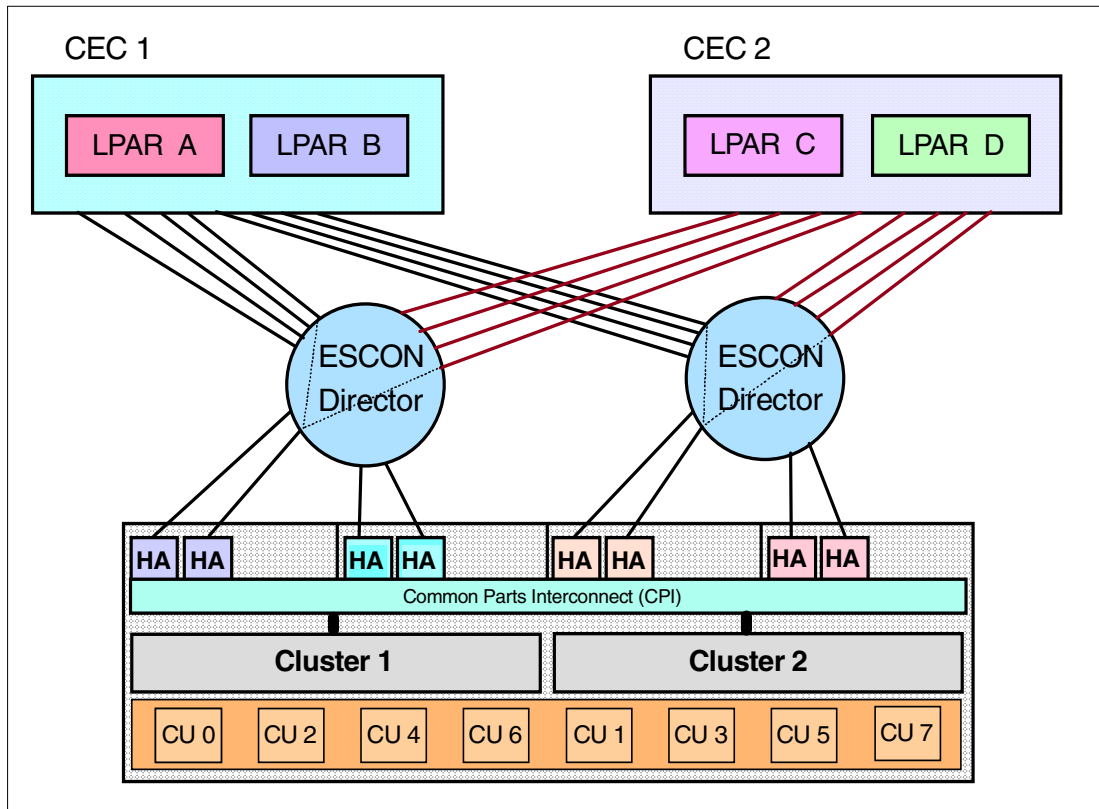


Figure 4-36 ESCON connectivity example

Figure 4-36 shows an example of how an IBM TotalStorage Enterprise Storage Server can be attached using ESCON via ESCON directors to different CECs and LPARs to provide high availability. For the best availability, you should spread all ESCON host adapters across all available host bays. These basic connectivity concepts remain similar for ESCON and FICON. What is not similar is the characteristics that they provide and the resulting attachment capabilities.

4.32.1 ESCON control unit images

The IBM TotalStorage Enterprise Storage Server allows you to configure up to 16 LSSs that will represent a matching number of CKD CU images or logical control units (LCUs) in the ESS. The CU images allow the ESS to handle the following connections:

- ▶ Up to 256 devices (base and alias) per CU image.
- ▶ Up to 4096 devices (base and alias) on the 16 CU images of the ESS. This is the 16 LCUs, each capable of addressing 256 devices, making a total of 4096 addressable devices within the ESS.
Note: Each ESCON host channel is only capable of addressing a maximum of 1024 devices, so multiple path groups will be required to address more than 1024 devices.
- ▶ A maximum of 64 logical paths per port.
- ▶ Up to 128 logical paths, and up to 64 path groups, for each CU image.
- ▶ A total of 2048 logical paths in the ESS (128 logical paths per CU image, multiplied by 16 CU images, makes 2048 logical paths).

4.32.2 ESCON logical paths establishment

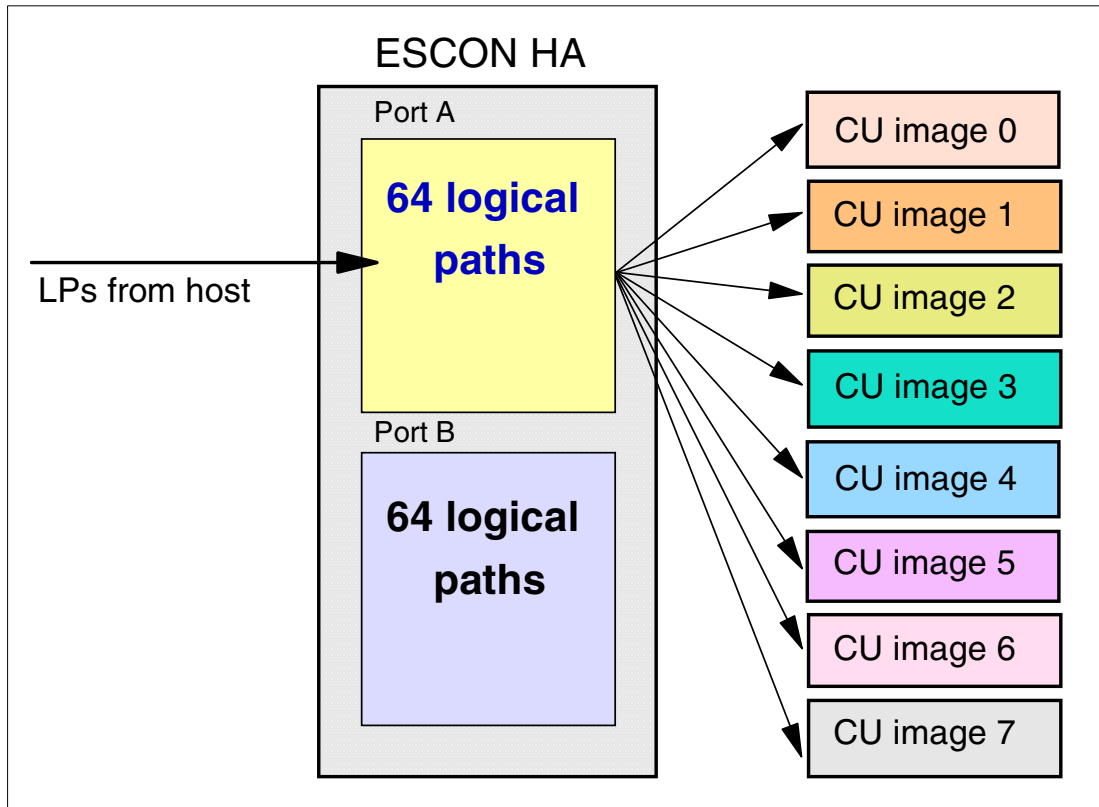


Figure 4-37 Establishment of logical paths for ESCON attachment

Figure 4-37 shows how logical paths are established in the ESS with the ESCON host adapters. This example shows a single port on an ESCON HA card and an ESS with eight CU images (LSSs) configured.

4.32.3 Calculating ESCON logical paths

For the following explanation, refer to Figure 4-36 on page 138:

- ▶ The ESS is configured for eight CU images
- ▶ All four LPARs have access to all eight CU images
- ▶ All LPARs have eight paths to each CU image

This results in:

$$4 \text{ LPARs} \times 8 \text{ CU images} = 32 \text{ logical paths}$$

So there will be 32 logical paths per ESCON adapter port, which does not exceed the 64 LPs per port ESCON maximum.

Under the same assumptions, each CU image must handle:

$$4 \text{ LPARs} \times 8 \text{ CHPIDs} = 32 \text{ LPs}$$

This will not exceed the 128 LPs a single ESS CU image can manage. These calculations may be needed if the user is running large sysplex environments. In such a case, it is also recommended to have many more channel paths attached to the ESS, to spread the CU images across several different channel sets.

4.33 FICON host connectivity

FICON channel connectivity brings some differences and provides a list of benefits over ESCON channel connectivity. Among the benefits you should consider:

- ▶ Increased addressing per channel (from 1024 device addresses for ESCON to up to 16,384 for FICON).
- ▶ Reduced number of channels, and hence fibers, with increased bandwidth and I/O rate per FICON channel.
- ▶ FICON channel to ESS supports multiple concurrent I/Os (ESCON supports only one I/O operation at a time).
- ▶ Greater channel and link bandwidth: FICON has up to 10 times the link bandwidth of ESCON (1 Gbps full duplex, compared to 200 Mbps half duplex). FICON has at least four times the effective channel bandwidth (70 MBps compared to 17 MBps). This advantage will grow as faster FICON link speeds are made available.
- ▶ FICON path consolidation using switched point-to-point topology.
- ▶ Greater unrepeated fiber link distances (from 3 km for ESCON to up to 10 km, or for FICON 20 km with an RPQ). FICON also sustains performance over greater distances than ESCON.

The configuration shown in Figure 4-38 is an example of connectivity of a zSeries server using eight FICON channel paths to half of the possible CU images of an ESS.

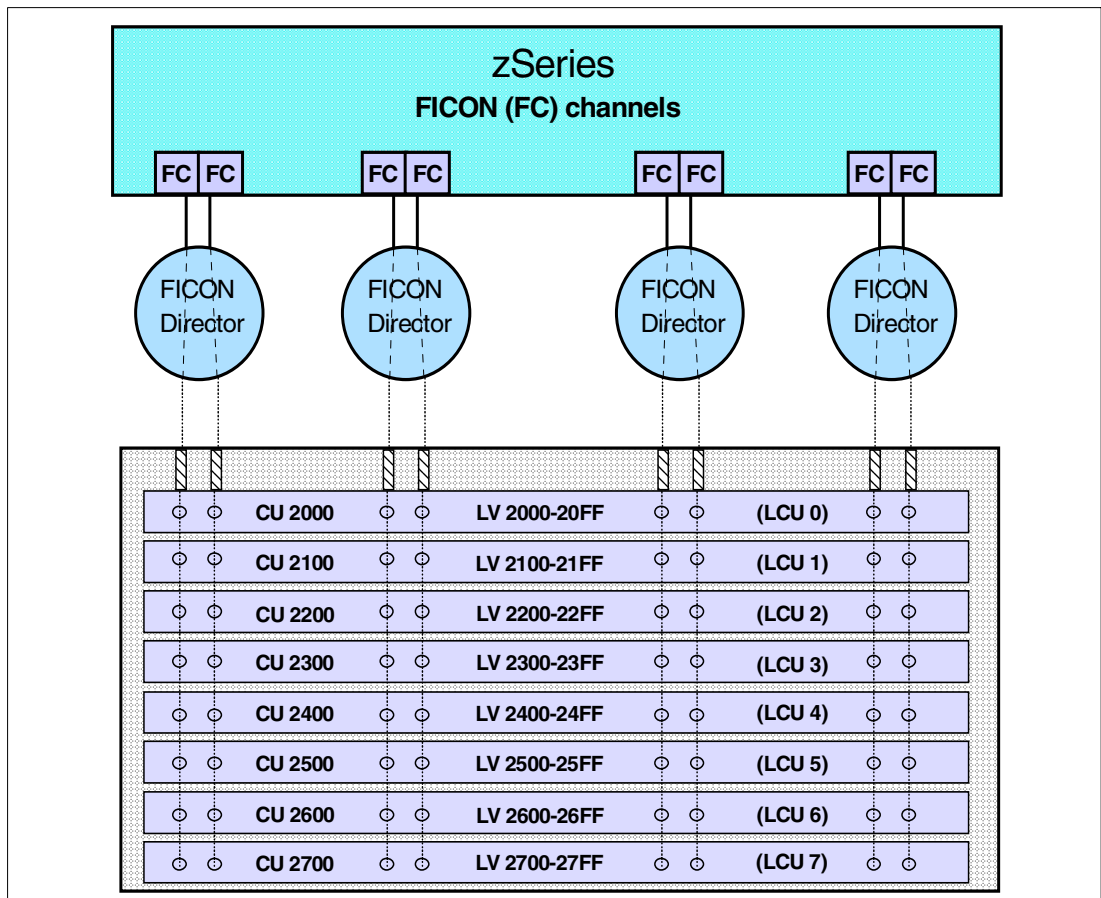


Figure 4-38 FICON connectivity

Refer to *FICON Native Implementation and Reference Guide*, SG24-6266 for more detailed information on ESS FICON connectivity.

4.33.1 FICON control unit images

The CU images allow the ESS to handle the following connections:

- ▶ Up to 256 devices (base and alias) per CU image (same as ESCON).
- ▶ Up to 4096 devices (base and alias) on the 16 CU images of the ESS. The 16 LCUs, each capable of addressing 256 devices, provide a total of 4096 addressable devices within the ESS. This is the same number of devices as ESCON, but all devices are addressable by a single FICON host channel.
- ▶ A maximum of 256 logical paths per port.
- ▶ Up to 256 logical paths for each CU image (double that of ESCON).
- ▶ A total of 4096 logical paths in the ESS (256 logical paths per CU image multiplied by 16 CU images, makes 4096 logical paths).

When you plan your ESS connectivity layout, you realize the dramatic benefits you get from the FICON implementation because of:

- ▶ The increased channel device addresses supported by the FICON implementation
- ▶ The increased number of concurrent connections
- ▶ The increased number of hosts that can share volumes

Increased channel device-address support

From 1024 devices on an ESCON channel to 16,384 devices for a FICON channel. This makes it possible for any FICON channel connected to the ESS to address all of the 4096 devices you can have within the ESS. This extra flexibility will simplify your configuration setup and management.

Increased number of concurrent connections

FICON provides an increased number of channel-to-control unit concurrent I/O connections. ESCON supports one I/O connection at any one time while FICON channels support multiple concurrent I/O connections. While an ESCON channel can have only one I/O operation at a time, the FICON channel can have I/O operations to multiple LCUs at the same time, even to the same LCU, by using the FICON protocol frame multiplexing.

All these factors, plus the increased bandwidth, allow you to take advantage of FICON and allow you to create redundant configurations more easily, accessing more data with even better performance than is possible with ESCON. For further considerations on FICON system attachment, refer to the document at:

<http://www.storage.ibm.com/hardsoft/products/ess/support/essficonwp.pdf>

FICON logical path establishment

Figure 4-39 on page 142 shows how logical paths are established in the IBM TotalStorage Enterprise Storage Server with the FICON host adapters. The FICON host adapter port will handle a maximum of 256 logical paths. This example shows a single port on a FICON host adapter card and an ESS with eight CU images (LSSs) configured.

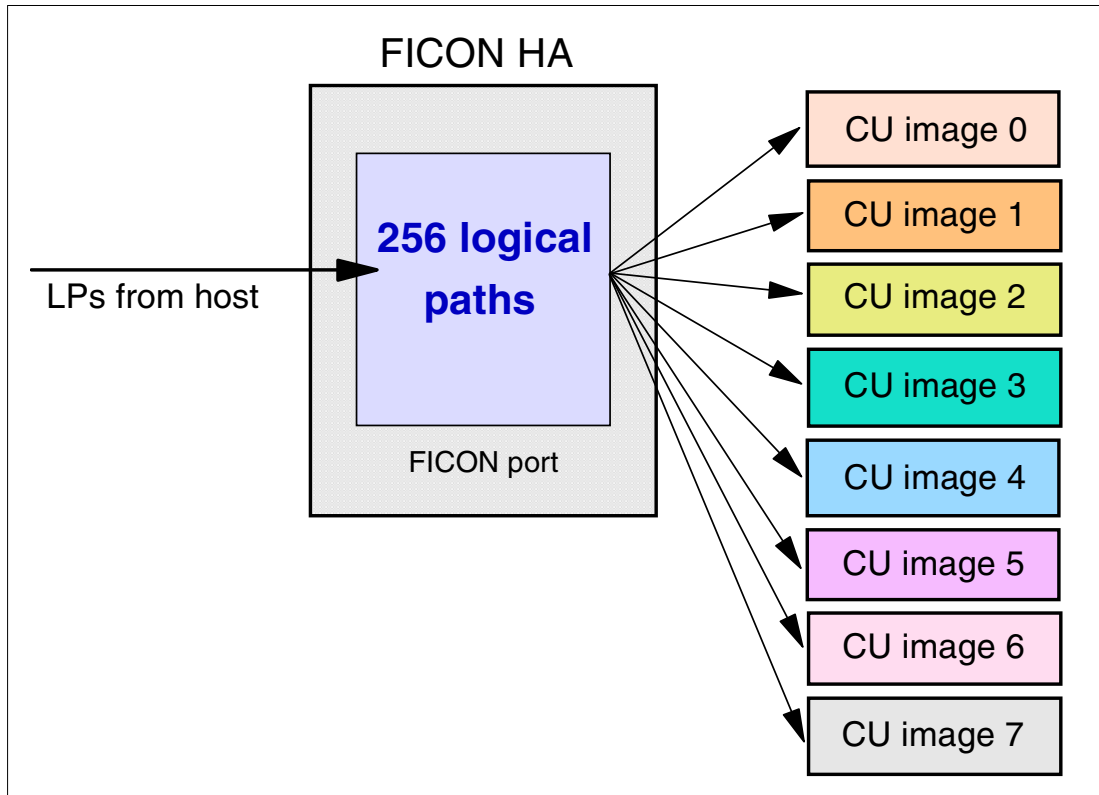


Figure 4-39 Establishment of logical paths for FICON attachment

4.33.2 Calculating FICON logical paths

In the example shown in Figure 4-40 on page 143, there are two ESSs, each with eight FICON host adapters (two per host adapter bay) and each with eight logical control units configured. All host adapters can be used by all logical control units within each ESS (conforming to the S/390 and z/Architecture maximum of eight paths from a processor image to a control unit image), and each logical control unit has 256 device addresses.

The two directors are each connected to four FICON channels on each CEC, resulting in 16 ports on each director for host channel connectivity. Each director is also connected to four host adapters on each of the two ESSs, resulting in eight ports on each director for ESS connectivity. This gives a total of 24 used ports for each director.

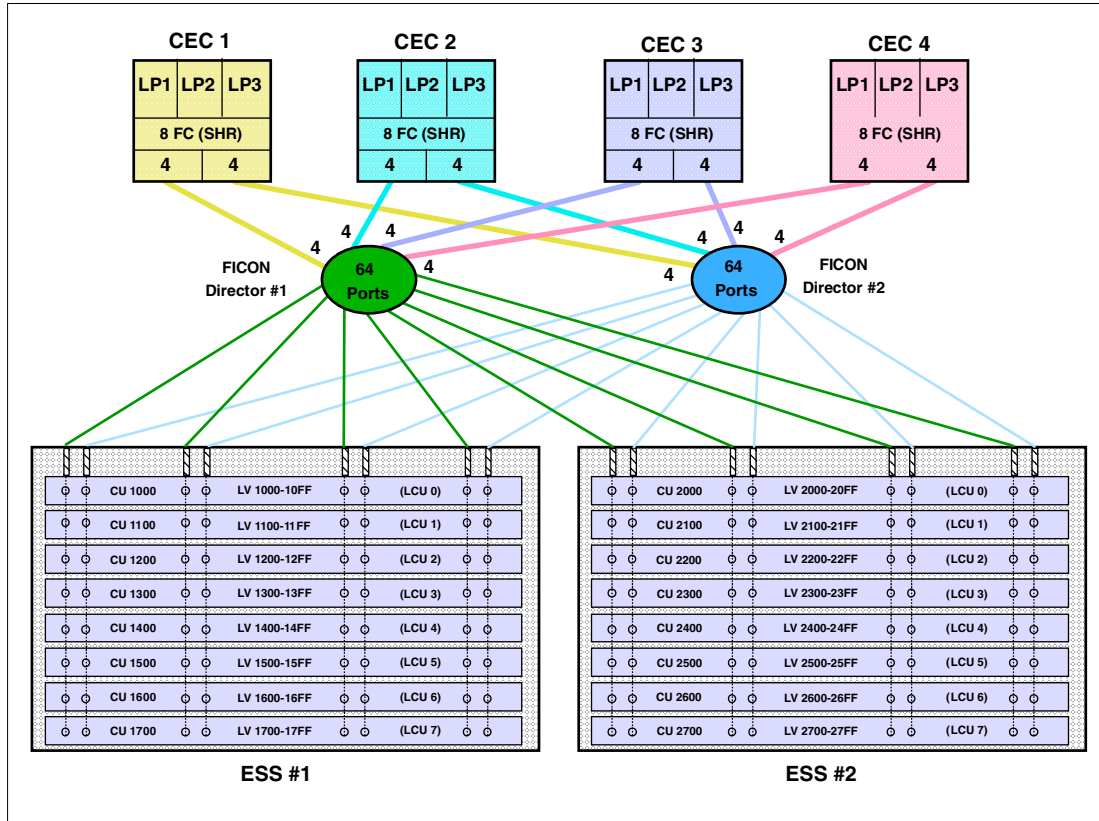


Figure 4-40 FICON connectivity example

The resources used by this configuration are:

- ▶ Eight FICON channels per CEC.
- ▶ 24 FICON ports per director (16 for host channels, and eight for ESS host adapters).
- ▶ 4096 subchannels per processor image.

Two ESSs, each one configured with eight logical control units, and 256 devices (the maximum) per logical control unit, makes 4096 subchannels per processor image ($2 \times 8 \times 256 = 4096$). The maximum number of subchannels per image is CEC dependent.

- ▶ 12,288 subchannels per CEC.

Three images per CEC, each with 4096 subchannels to access all the devices in both ESSs, makes 12,288 subchannels per CEC ($3 \times 4096 = 12,288$). Note that this number does not distinguish between base and alias devices. An alias device requires a subchannel just like a base device.

The maximum number of subchannels per CEC is CEC dependent.

- ▶ 4096 subchannels per FICON channel.

As each FICON channel is connected to all eight logical control units on both ESSs, and each logical control unit has 256 devices configured (the maximum), the number of subchannels per FICON channel is 4096 ($2 \times 8 \times 256 = 4096$).

The maximum number of devices per FICON channel is 16,384.

- ▶ Four Fibre Channel N_Port logins per FICON host adapter.

There are four CECs and all control unit host adapters are accessed by a channel to all CECs, so there are four N_Port logins per ESS host adapter.

The maximum number of N_Port logins for the ESS is 128 per FICON host adapter. This means that the maximum number of FICON channels that can be attached to a FICON port (using a director) is 128.

- ▶ 96 logical paths per FICON host adapter.

There are 12 images in total (4 CECs x 3 LPARs), and each image has eight logical paths through each of the FICON host adapters (one logical path per logical control unit within the ESS). This makes 96 logical paths per FICON host adapter in the example (12 x 8 = 96).

The maximum number of logical paths per ESS host adapter is 256 for FICON attachment (vs. 64 for ESCON).

- ▶ 96 logical paths per logical control unit.

There are eight paths per logical control unit to each processor image. In all four CECs there are 12 images, so there are 96 (8 x 12) logical paths per logical control unit.

The maximum number of logical paths per logical control unit is 256 (vs. 128 for ESCON).

So the example ESS configuration shown in Figure 4-40 on page 143 is within the FICON resources limit, and you can see that it requires significantly fewer channel and connectivity resources than an equivalent ESCON configuration would require.

FICON resources exceeded

When planning for your configuration, take care not to over-define the configuration. One of the most common situations is to over-define the number of logical paths per logical control unit. For the ESS, you cannot have more than 256 logical paths online to any logical control unit. Any attempt to vary more logical paths online will fail.

The problem with the over-defined configuration may not surface until an attempt is made to vary online paths to the devices beyond the already established limit of logical paths for the logical control unit or host adapter.

The z/OS message in Figure 4-41 is issued to reject the vary path processing:

```
VARY PATH(dddd,cc), ONLINE
IEE714I PATH(dddd,cc) NOT OPERATIONAL
```

Figure 4-41 Over-defined paths — system message

Note that there is no additional indication of the cause of the not-operational condition. For this situation, you can run the ICKDSF logical path report to identify which channel images have established logical paths to the logical control unit. For more detailed information on running and interpreting the ICKDSF logical paths report, refer to the FICON problem determination chapter in *FICON Native Implementation and Reference Guide*, SG24-6266.

4.34 ESCON and FICON connectivity intermix

Intermixing ESCON channels and FICON native channels to the same CU from the same operating system image is supported as a transitional step for migration only.

Intermixing ESCON channels, FICON Bridge channels, and FICON native channels to the same control unit from the same processor image is also supported, either using

point-to-point, switched point-to-point or both. IBM recommends that FICON native channel paths only be mixed with ESCON channels and FICON Bridge channel paths to ease migration from ESCON channels to FICON channels using dynamic I/O configuration.

The coexistence is very useful during the transition period from ESCON to FICON channels. The mixture allows you to dynamically add FICON native channel paths to a control unit while keeping its devices operational. A second dynamic I/O configuration change can then remove the ESCON channels while keeping devices operational. The mixing of FICON native channel paths with native ESCON and FICON Bridge channel paths should only be for the duration of the migration to FICON.

This migration process is illustrated in 9.11, “Migrating from ESCON to FICON” on page 256.

4.35 Standard logical configurations

There are two ways that the IBM TotalStorage Enterprise Storage Server can be configured when it is first installed

- ▶ Using just the ESS Specialist
- ▶ Using both the ESS Batch Configuration Tool and the ESS Specialist.

The latter is a way to configure the ESS in a simplified fashion that utilizes standard configurations and standard volume sizes. This tool is designed to perform the initial configuration of the ESS. Subsequent changes to the configuration should be done with the ESS Specialist. The ESS Batch Configuration Tool can also be used to configure additional eight-packs if you add capacity to the ESS. The ESS Batch Configuration Tool is a tool that is run by the IBM System Service Representative, whereas the ESS Specialist is available for anyone to use, if authorized.

For CKD servers, refer to Appendix A, “S/390 standard logical configuration”, in *IBM TotalStorage Enterprise Storage Server Configuration Planner for S/390 and zSeries Hosts*, SC26-7476.

For FB and iSeries servers, refer to Appendix A, “Open-systems standard configuration”, in *IBM TotalStorage Enterprise Storage Server Configuration Planner for Open-Systems Hosts*, SC26-7477.

These standard options speed the logical configuration process, and the only setup you must do is the assignment to the host adapter ports, which is a quick process. The effective capacity of each standard configuration depends on the disk array capacity.



Performance

The IBM TotalStorage Enterprise Storage Server Model 800 is an unparalleled performing storage subsystem. This chapter describes the performance features and characteristics that position the ESS as the performance leader for disk storage solutions.

The ESS also delivers a multi-feature synergy with the zSeries server operating systems. These features are presented in Chapter 6, “zSeries performance” on page 165.

5.1 Performance accelerators

The ESS design based upon the Seascape Architecture uses the latest IBM technology that includes advanced RISC symmetrical multiprocessing (SMP) microprocessors, Serial Storage Architecture (SSA) disk adapters, high-performance disk drives, and microcode intelligent algorithms. These features gathered together under a superb architectural design deliver the best performance you could presently expect from a disk storage server solution.

Figure 5-1 lists the performance accelerator features and functions that make the ESS the leader in performance for all the heterogeneous environments it attaches to.

- Third-generation hardware technology
- Faster RISC SMP processors
 - with Turbo option
- 2 GB non-volatile storage (NVS)
- Double CPI bandwidth
- 64 GB cache
- Efficient caching algorithms
- Efficient I/O commands
- Serial Storage Architecture (SSA) back end
 - New generation, more powerful disk adapters
- High performance 15,000 rpm disks drives
- 2 Gb Fibre Channel/FICON adapters

Figure 5-1 ESS Performance accelerators

The zSeries and S/390 platform-specific features are presented in Chapter 6, “zSeries performance” on page 165.

5.2 Third-generation hardware

The IBM TotalStorage Enterprise Storage Server Model 800 is the third generation of the ESS and builds upon the functionality, reliability, and proven track record of the earlier models that set the standard for others to follow.

The ESS Model 800 integrates a new generation of hardware from top to bottom, including host adapters, through the processors and NVS and cache, increased internal bandwidth, then down to the device adapters and 15,000 rpm disks. These updates provide up to two times the throughput and performance of the previous generation F models, and up to two and a half times with the optional Turbo processor feature. Not only do normal I/O workloads benefit, but advanced ESS functions, such as FlashCopy, also have improved performance.

Since workload, I/O profile, and requirements differ, guidance on the performance achievable in your environment using the ESS Model 800 can be supplied by your local IBM Storage Specialist. The specialist can, after discussing details of a user’s specific storage workload, undertake performance modeling with a tool called Disk Magic.

5.3 RISC SMP processors

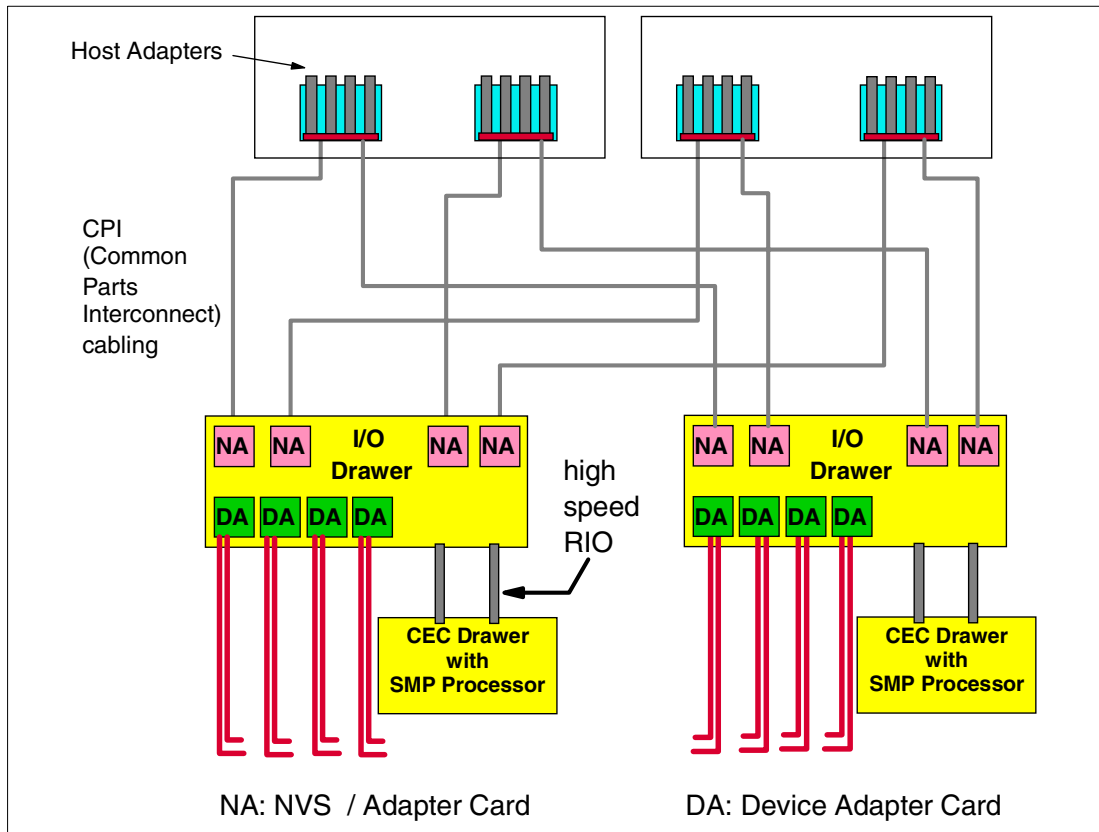


Figure 5-2 RISC SMP processors (simplified diagram)

Each cluster contains high-performance RISC processors, configured as symmetric multiprocessors (SMP). The processors are 64-bit RISC based, and utilize innovative copper and silicon-on-insulator (SOI) technology. There is the option of a Turbo processor feature, for the more demanding workloads. Each SMP can be configured with 4 GB to 32 GB of cache (totaling 8 GB to 64 GB per ESS).

The standard processors provide better performance and throughput over the previous ESS F models (up to two times), with the Turbo processor feature delivering even higher performance (two and a half times) for the heavy-duty workload environments.

Generally, the Turbo processors will be beneficial for high-throughput applications with very high I/O per second requirements (OLTP, some database applications, TPF). The Turbo processors will also benefit heavy copy services workloads with high I/O rates coupled with high storage capacity. Your IBM Field Technical Support Specialist (FTSS) should be consulted to assist you in selecting the best cluster processor option to meet your performance requirements within your specific workload mix.

CPI bandwidth

The CPI bandwidth between the host adapters and the clusters has been significantly improved (doubled) with more powerful buses. Together, the I/O drawers provide 16 PCI slots for logic cards, eight for Device Adapter cards (DA) and eight for NVS adapter cards (NA). Each NA card has NVS memory and a CPI bus to the host adapter. Figure 5-2 presents a simplified diagram of the internal connections of the ESS.

5.4 Caching algorithms

With its effective caching algorithms, the IBM TotalStorage Enterprise Storage Server Model 800 is able to minimize wasted cache space, reduce disk drive utilization, and consequently reduce its back-end traffic. The ESS Model 800 has a maximum cache size of 64 GB, and the NVS standard size is 2 GB.

5.4.1 Optimized cache usage

The ESS manages its cache in 4 KB segments, so for small data blocks (4 KB and 8 KB are common database block sizes) minimum cache is wasted. In contrast, large cache segments could exhaust cache capacity while filling up with small random reads. Thus the ESS, having smaller cache segments, is able to avoid wasting cache space for situations of small record sizes that are common in the interactive applications.

This efficient cache management, together with the ESS Model 800 powerful back-end implementation that integrates new (optional) 15,000 rpm drives, enhanced SSA device adapters, and twice the bandwidth (as compared to previous models) to access the larger NVS (2 GB) and the larger cache option (64 GB), all integrate to give greater throughput while sustaining cache speed response times.

5.4.2 Sequential pre-fetch I/O requests

Storage subsystem cache has proven to be of enormous benefits for the CKD (z/OS) servers. It is often of less value for the FB (open systems) servers because of the way these servers use the server processor memory as cache. In the zSeries environments, storage system cache usually offers significant performance benefits for two main reasons:

- ▶ zSeries operating systems tend not to keep the most frequently referenced data in processor memory.
- ▶ zSeries servers store data in a way that allows disk systems to potentially predict sequential I/O activity and pre-fetch data into system cache.

The ESS monitors the channel program patterns to determine if the data access pattern is sequential or not. If the access is sequential, then contiguous data is pre-fetched into cache in anticipation of the next read requests. It is common that z/OS set a bit into the channel program notifying the disk subsystem that all subsequent I/O operations will be sequential read requests. ESS supports these bits in the channel program and helps to optimize its pre-fetch process.

The ESS uses its sequential prediction algorithms, and its high back-end bandwidth, to keep the cache pre-loaded and ready with data for the upcoming I/O operations from the applications. For detailed information, refer to the sections in Chapter 3, "Architecture" on page 49 where cache operations and algorithms are described.

5.5 Efficient I/O operations

Both for the fixed block (FB) environments and for the count-key-data (CKD) environments, I/O operations are solved in the most efficient manner.

5.5.1 SCSI command tag queuing

Servers connecting to an ESS using the SCSI command set (over Fibre Channel or over parallel SCSI) may use SCSI command tag queuing. This function supports multiple outstanding requests to the same LUNs at the same time.

Such requests are processed by the ESS concurrently if possible. Some of these I/O requests may then be solved by the ESS with a cache hit, and others may be solved on the disk array where the logical volumes are striped.

5.5.2 z/OS enhanced CCWs

For the z/OS environments, the ESS supports channel command words (CCWs) that reduce the characteristic overhead associated to the previous (3990) CCW chains. Basically, with these CCWs, the ESS can read or write more data with fewer CCWs. CCW chains using the old CCWs are converted to the new CCWs whenever possible. Again, the cooperation of IBM z/OS software and the IBM ESS provides the best benefits for the application's performance.

Adding the performance benefits gained from FICON together with the enhanced CCW function delivers front-end I/O bandwidth and performance to the ESS that was not achievable in the past.

z/VM itself does not use the new CCWs; however, it allows a guest to use the new CCWs.

5.6 Back-end high performance design

A performance design based only on cache size and its efficiency may not be properly addressing the workload characteristics of all the FB servers. For these type of servers, the back-end efficiency is the key to the unbeatable performance of the ESS.

The performance of the disk portion of the disk storage subsystem (disk drives, buses, and disk adapters, generally designated as *backstore* or *back end*) has a major impact on performance. In the ESS, these characteristics make the difference for the fixed block environments where the servers usually do not get so many cache hits in the storage subsystem.

5.6.1 Serial Storage Architecture (SSA)

The ESS uses Serial Storage Architecture (SSA) loops for its back-end implementation.

SSA loop bandwidth

SSA is physically configured as a loop. A single SSA loop has a maximum bandwidth of 160 MBps. This bandwidth is multiplied by the number of adapters attached to the loop. ESS has a total of eight SSA adapters and one adapter connects two distinct loops. Therefore, a total of $8 \times 2 \times 160 \text{ MBps} = 2560 \text{ MBps}$ nominal bandwidth is available in the backstore.

The ESS Model 800 has also improved the SSA device adapters, which are able to achieve more operations per second than the previous F models.

Paths to the disks

The ESS has up to 64 (SSA) internal data paths to disks, each with a bandwidth of 40 MBps (the SSA loop bandwidth of 160 MBps per adapter consists of two read paths and two write paths at 40 MBps delivered at each adapter connection). It is possible for every path to be transferring data at the same time.

SSA implementation is described in detail in 2.8, “Device adapters” on page 31 and 2.9, “SSA loops” on page 33.

5.6.2 Striping

Data striping means storing logically contiguous data across multiple physical disk drives, which provides significant performance benefits, including:

- ▶ Balanced disk drive utilization
- ▶ I/O parallelism for cache misses
- ▶ Higher bandwidth for the logical volumes
- ▶ Avoidance of disk drive hot spots

All of these lead to higher sustained performance and reduced manual tuning. The following diagrams show how the ESS implements data striping.

RAID 5 striping

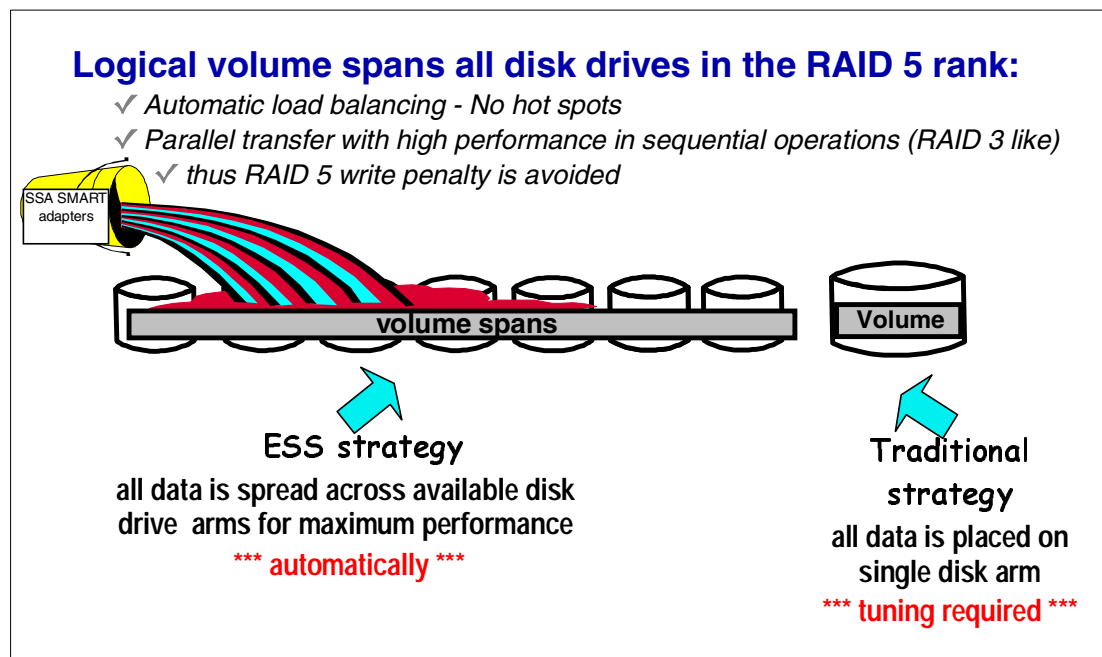


Figure 5-3 RAID 5 logical volume striping

The ESS stripes every RAID 5 protected logical volume across multiple disk drives of the RAID rank. Additionally, to neutralize the RAID 5 write penalty, for sequential operations the ESS does the rank writes in a RAID 3 style (parallel transfer of all stripes). This avoids the read and recalculation overheads associated with the RAID 5 write operations.

RAID 10 striping

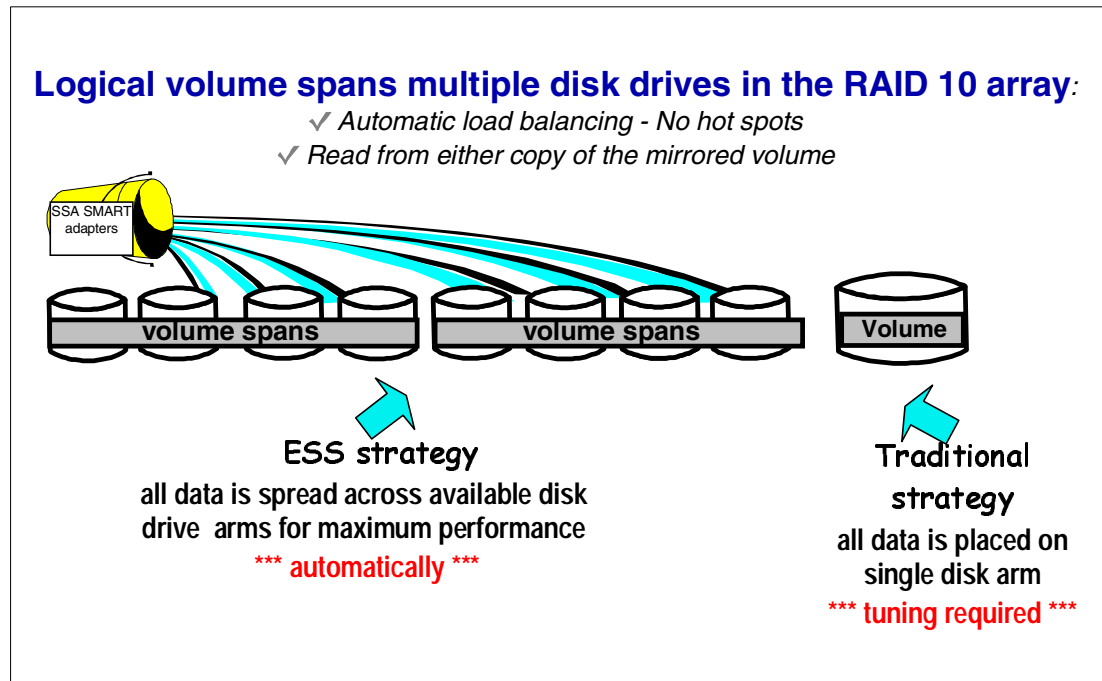


Figure 5-4 RAID 10 logical volume striping

The ESS also stripes every RAID 10 protected logical volume across multiple disk drives of the RAID rank. This provides up to four times performance improvement over a single disk.

RAID 10 allows a write to both data and mirrored copy of data, this operation is done in one *striped* write. As RAID 10 enables a read from either copy of the data, this ensures that whenever a copy is not available (busy, or with errors) then the alternate copy will be accessed.

5.6.3 RAID 10 vs RAID 5

RAID 5 optimizes storage far better than RAID 10 (less capacity overhead). While RAID 5 remains a price/performance leader, offering excellent performance for most applications, RAID 10 can offer better performance for selected applications, in particular high random write content applications in the open systems environment. This is so because of the I/O processing characteristics of the RAID 10 operations:

- ▶ Reads can be satisfied from either of the mirrored copies in a RAID 10 rank, so the I/O read request will be satisfied from the available copy.
- ▶ Writes don't need parity generation when done upon a RAID 10 rank.

The decision about whether to use RAID 5 or RAID 10 will depend principally upon the user's application's performance requirements balanced against the extra cost of using RAID 10 over RAID 5. Although RAID 10 arrays will potentially deliver higher performance than RAID 5 arrays, it must be considered that the disk arrays are front-ended by a powerful pair of clusters with considerable amounts of cache memory and NVS. For some applications, the majority of its reads and writes can be satisfied from cache and NVS, so these applications will see no discernible difference whether the ranks are RAID 5 or RAID 10.

You can estimate that RAID 10 reads will be very similar to RAID 5 reads, from the performance perspective. It is high levels of random writes that will show benefits for the

RAID 10 operations. Random reads, sequential reads, and sequential write performance will be very similar between RAID 5 and RAID 10. In fact, you should consider that for sequential writes RAID 5 should be faster, because it writes less data as compared to RAID 10. Since writes on the ESS are asynchronous, whether it is RAID 10 or RAID 5 does not really affect the I/O response time seen by the application. However, RAID 10 allows more write operations (asynchronous destages) to the disks, thus allowing higher throughput levels for write-intensive workloads as compared to RAID 5 with the same number of disks.

The ESS has already proven that RAID 5 is suitable for the majority of the commercial applications and it would be wise to continue to evaluate specific application performance requirements and I/O characteristics to decide whether RAID 5 or RAID 10 is most appropriate. For instance, RAID 5 is very well suited for database logging since it is generally sequential in nature and enables the SSA adapter to do full stripe writes.

The Disk Magic modeling tool can be used by the IBM Storage Specialist for determining whether any benefit is to be gained using one RAID format over another with your specific workload mix.

5.6.4 High-performance disk drives



Figure 5-5 Disk drive

When configuring the IBM TotalStorage Enterprise Storage Server Model 800, you have the option of choosing different capacity and different speed disk drives: 18.2 GB and 36.4 GB 15,000 rpm and 10,000 disk drives, and the 72.8 GB 10,000 rpm drives. These different capacities and speed disk drives can be intermixed within the same loop of an ESS (certain limitations apply; please refer to “Disk speed (RPM) intermix considerations” on page 28).

The latest 15,000 rpm family of drives, which are available in 18.2 GB and 36.4 GB capacities, provide levels of throughput and performance that can translate into substantial price/performance benefits for the entire system. An ESS populated with RAID 5 ranks of 15,000 rpm drives can provide up to 80% greater total system random throughput for cache standard workload (typical database workload) than a comparably configured ESS with 10,000 rpm drives.

These drives can drive workloads at significantly higher access densities (reads/writes per second per GB disk capacity) without worrying about performance degradation, and fewer disk drives may be required to achieve high disk utilization rates, which can lead to cost savings. Reduced response times may also be realized, providing shorter batch processing windows or improved productivity because transactions complete more quickly. For example, at a typical access density of 1.0, a response time reduction of up to 25% can be achieved. And in more demanding environments, an online transaction processing (OLTP) workload at a relatively stressed access density of 2.5 can enjoy a reduction of up to 40% in response time. These improvements may be even more significant for cache-hostile OLTP workloads.

The Disk Magic modeling tool can be used by the IBM Storage Specialist for determining the disks configuration that best meets the specific performance requirements of a particular I/O workload.

In line with the Seascape architecture of the ESS, the disk drive models used within the ESS are often changed. This allows you to utilize the latest high-performance and capacity disk drives available.

5.7 Configuring for performance

This section contains general recommendations for configuring the ESS for performance. In addition, white papers with information related to the ESS are available at:

<http://www.storage.ibm.com/hardsoft/products/ess/whitepaper.htm>

Important: Despite the general recommendations given in this section, remember that performance generalizations are often dangerous and the best way to decide on the ESS configuration options most appropriate for your processing environment is for an IBM Storage Specialist or Business Partner Storage Specialist to discuss your requirements in detail and to model your workload with a tool such as Disk Magic.

5.7.1 Host adapters

Always spread the host connections across all the host adapter bays. Distribute the connections to the host adapters across the bays in the following sequence: Bay 1 - Bay 4 - Bay 2 - Bay 3. This recommendation is for the following reasons:

1. The bays are connected to different PCI buses in each cluster, and by spreading the adapter connections to the host adapters across the bays, you also spread the load and improve overall performance and throughput.
2. If you need to replace or upgrade a host adapter in a bay, then you have to quiesce all the adapters in that bay. If you spread them evenly, then you will only have to quiesce a quarter of your adapters. For example, for an ESCON configuration with eight ESCON links spread across the four bays, then the loss of two ESCON links out of eight may have only a small impact, compared with all eight if they were all installed in one bay.
3. Take into consideration that HAs in Bays 1 and 3 share the same internal bus in each cluster, and HAs in Bays 2 and 4 share a different internal bus in the cluster. This is

especially important for open systems to avoid the situation where all the activity for a cluster comes from Bays 1 and 3, or from Bays 2 and 4.

4. For optimal availability avoid the situation where all connections to a single host system are from Bays 1 and 2, or from Bays 3 and 4.

5.7.2 RAID ranks

The RAID implementation of the ESS ensures the best performance and throughput for back-end disk I/O read and write activity. This applies to both open systems servers and zSeries servers environments. The ESS has proven that the common perception that RAID 5 has poor performance characteristics is not true, and now you have the choice of RAID 5 or RAID 10 as well (see discussion in 5.6.3, “RAID 10 vs RAID 5” on page 153).

For both the zSeries and the open systems servers, the *fast write* capability, the *sequential processing* capability, together with the back-end SSA device adapters processing characteristics and the new fastest (rpm) disk drives, they all contribute to have high performance I/O operations upon the RAID 5 arrays.

As already discussed, for selected applications environments RAID 10 will normally deliver better performance because its particular implementation characteristics (see previous discussions in page 153) can make a difference for certain types of application workloads. The Disk Magic modeling tool can be run by the IBM Storage Specialist to determine how RAID 10 improves the throughput/performance for a particular user workload.

For the best performance, both with RAID 5 and RAID 10 ranks, spread your I/O activity across all the RAID ranks. It is also beneficial to evenly balance the RAID ranks across clusters and between all available LSS/LCUs.

5.7.3 Cache

Cache size can be 8, 16, 24, 32 or 64 GB. With different cache sizes, the ESS can be configured for the proper balance between disk capacity (with an associated access density) and cache size. Alternatively, if you don't increase disk capacity (access density remaining the same) but double the cache size, then the cache hit ratio may improve resulting in better I/O response times to the application.

The large ESS cache size, together with the ESS advanced cache management algorithms, provide high performance for a variety of workloads. The Disk Magic modeling tool can be run by the IBM Storage Specialist to determine how the cache size impacts the throughput and performance behavior for a particular user workload.

5.7.4 Disk drives

The ESS offers a high degree of choice when selecting disk drive sizes, with either 10,000 rpm or 15,000 rpm 18.2 GB and 36.4 GB capacity disk drives being available, as well as 10,000 rpm 72.8 GB capacity disk drives.

Analysis of workload trends over recent years shows that access densities (I/O operations per second per GB disk space) continue to decline, so workloads can often be migrated to higher capacity disk drives without any significant impact in the performance. For example, the greater capacity 72.8 GB drives provide superior seek and throughput characteristics, thereby allowing each drive to satisfy big demanding I/O loads. As access densities continue to decline, workloads migrated to 72.8 GB drives will often not notice any material difference in performance compared to smaller capacity older generation disk drives.

The faster (rpm) disks have better performance, and so some applications (more than others) will be able to obtain noticeable gains from the disk speed difference. For example, RAID 5 ranks populated with 15,000 rpm drives can provide up to 80% greater total throughput for *cache standard* workloads (typical database workload) as compared to similar configurations with 10,000 rpm drives.

Some users have very high demand workloads that may be very cache-unfriendly or have a very high random write content. These workloads may still require the lower capacity drives (more drives are required for a given capacity). For these workloads, users may want to consider lower capacity disk drives or, for environments with mixed workloads with different characteristics, an intermix of drive capacities.

As each customer's workload (I/O profile) and requirements differ, guidance on the performance achievable in your environment using the ESS Model 800 with various disk capacities and speeds (rpm) can be supplied by your local IBM Storage Specialist. The specialist can, after discussing details of the user's specific storage workload, undertake performance modeling with a tool called Disk Magic.

5.8 Subsystem Device Driver

The IBM Subsystem Device Driver (SDD) software is a host-resident software package that manages redundant connections between the host server and the IBM TotalStorage Enterprise Storage Server Model 800, providing enhanced performance and data availability.

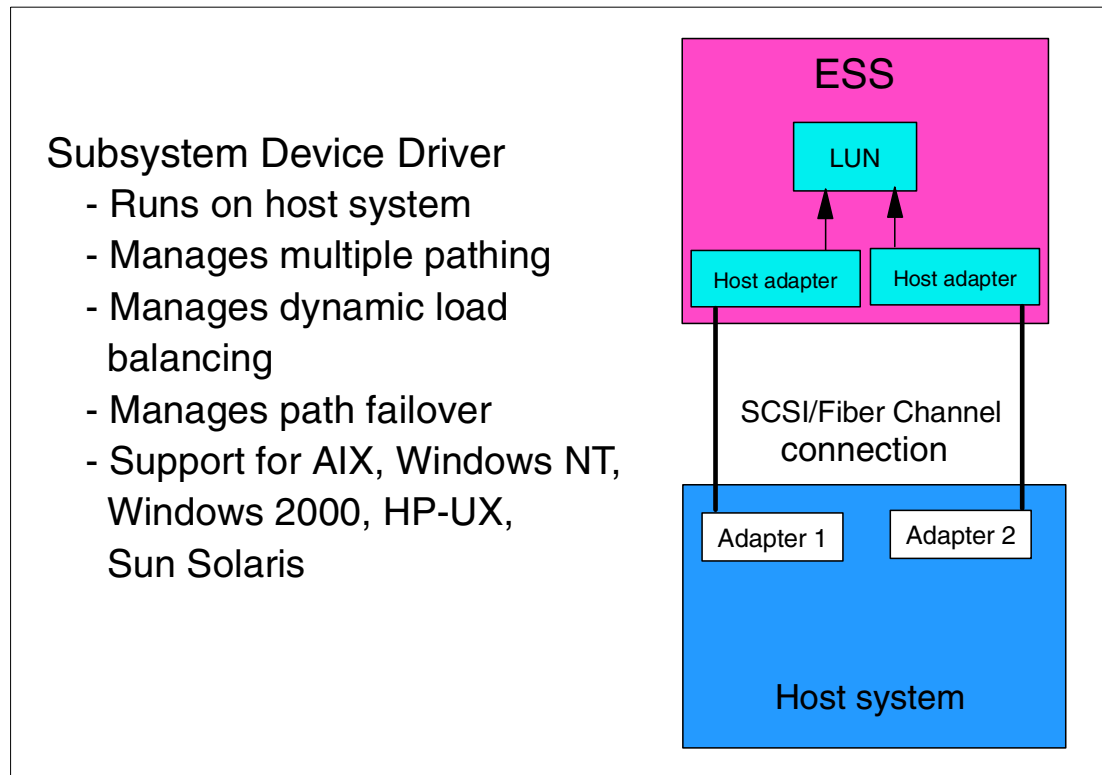


Figure 5-6 Subsystem Device Driver (SDD)

SDD provides ESS users in Windows NT and 2000, AIX, HP/UX, and Sun Solaris environments with:

- ▶ Load balancing between the paths when there is more than one path from a host server to the ESS. This may eliminate I/O bottlenecks that occur when many I/O operations are directed to common devices via the same I/O path.
- ▶ An enhanced data availability capability for customers that have more than one path from a host server to the ESS. It eliminates a potential single point of failure by automatically rerouting I/O operations to the remaining active path from a failed data path.

The Subsystem Device Driver may operate under different modes/configurations:

- ▶ Concurrent data access mode - A system configuration where simultaneous access to data on common LUNs by more than one host is controlled by system application software such as Oracle Parallel Server, or file access software that has the ability to deal with address conflicts. The LUN is not involved in access resolution.
- ▶ Nonconcurrent data access mode - A system configuration where there is no inherent system software control of simultaneous access to the data on a common LUN by more than one host. Therefore, access conflicts must be controlled at the LUN level by a hardware-locking facility such as SCSI Reserve/Release.

Some operating systems and file systems natively provide similar benefits provided by SDD, for example z/OS, OS/400, NUMA-Q Dynix, and HP/UX.

It is important to note that the IBM Subsystem Device Driver does not support boot from a Subsystem Device Driver pseudo device. Also, the SDD does not support placing a system paging file in a SDD pseudo device.

For more information, refer to the *IBM TotalStorage Subsystem Device Driver User's Guide*, SC26-7478. This publication and other information are available at:

<http://www.ibm.com/storage/support/techsup/swtechsup.nsf/support/sddupdates>

Load balancing

SDD automatically adjusts data routing for optimum performance. Multipath load balancing of data flow prevents a single path from becoming overloaded, causing input/output congestion that occurs when many I/O operations are directed to common devices along the same input/output path. The path selected to use for an I/O operation is determined by the policy specified for the device. The policies available are:

- ▶ Load balancing (default) - The path to use for an I/O operation is chosen by estimating the load on the adapter to which each path is attached. The load is a function of the number of I/O operations currently in process. If multiple paths have the same load, a path is chosen at random from those paths.
- ▶ Round robin - The path to use for each I/O operation is chosen at random from those paths not used for the last I/O operation. If a device has only two paths, SDD alternates between the two.
- ▶ Failover only - All I/O operations for the device are sent to the same (preferred) path until the path fails because of I/O errors. Then an alternate path is chosen for subsequent I/O operations.

Normally, path selection is performed on a global rotating basis; however, the same path is used when two sequential write operations are detected.

Concurrent LIC load

With SDD you can concurrently install and activate the ESS Licensed Internal Code while applications continue running if multiple paths from the server have been configured. During the activation process, the host adapters inside the ESS might not respond to host I/O requests for up to 30 seconds. SDD makes this process transparent to the host system through its path-selection and retry algorithms.

Path failover and online recovery

SDD automatically and nondisruptively can redirect data to an alternate data path. In most cases, host servers are configured with multiple host adapters with either SCSI or Fibre Channel connection to an ESS that in turn would provide internal component redundancy. With dual clusters and multiple host adapters, the ESS provides more flexibility in the number of input/output paths that are available.

When a path failure occurs, the IBM SDD automatically reroutes the I/O operations from the failed path to the other remaining paths. This eliminates the possibility of a data path being a single point of failure.

5.9 Measurement tools

There are a number of tools available for monitoring the performance of an ESS. There are common and specific tools available for both the zSeries environment and for the open systems environments.

IBM TotalStorage Expert (ESS Expert feature)

- Can monitor both open systems and zSeries ESS performance
- Agent in ESS to collect data
- Performance data
 - Number of I/Os, bytes transferred
 - Cache hit rates, response times

RMF

- DASD reports
- CRR
- SMF data

IDCAMS LISTDATA

- RAID rank performance data

Note: RMF and IDCAMS are for S/390 environments
IBM TotalStorage ESS Expert is for all environments.

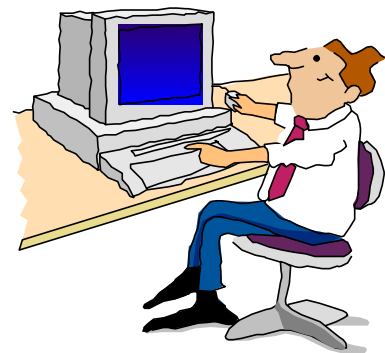


Figure 5-7 Performance measurement tools

Today's storage subsystems have very complex features such as multiple host adapters (Fiber Channel, ESCON, FICON, SCSI), SSA device adapters, large cache memory sizes, multiple clusters, RAID protected disk arrays, all tied together to provide large bandwidth to satisfy performance requirements. This multiplicity of features is especially true with the ESS. Working with more than one tool may be more useful when managing today storage servers.

5.10 IBM TotalStorage Expert

The IBM TotalStorage Expert (ESS Expert feature), a product of the IBM TotalStorage family, enables you to gather a variety of information about your IBM TotalStorage Enterprise Storage Server. The ESS Expert then presents you with information that can significantly help you monitor and optimize performance, and manage the capacity and asset information related to your ESS.

For more information on the ESS Expert (5648-TSE), refer to *IBM TotalStorage Expert Hands-On Usage Guide*, SG24-6102 or visit the following sites. For general features and functions:

<http://www.storage.ibm.com/software/storwatch/ess/index.html>

For installation and product prerequisite information, visit:

<http://www.storage.ibm.com/software/expert/>

The IBM TotalStorage software family site is:

<http://www.storage.ibm.com/storwatch>

There is also an IBM TotalStorage Expert online demonstration site available where you can get a hands-on look at the functions and reports available. The Web address is:

<http://www.storwatch.com/>

5.10.1 ESS Expert overview

The ESS Expert helps you with the following activities:

- ▶ Asset management
- ▶ Capacity management
- ▶ Performance management

This section gives an introduction and brief description of the tasks the ESS Expert offers.

Asset management

When storage administrators have to manage multiple storage systems, it can be time consuming to track all of the storage system names, microcode levels, model numbers, and trends in growth. The asset management capabilities of ESS Expert provide the capability to:

- ▶ Discover (in the enterprise network) all of the ESS storage systems, and identify them by serial number, name, and model number
- ▶ Identify the number of clusters and expansion features on each
- ▶ Track the microcode level of each cluster

This information can save administrators time in keeping track of their storage system hardware environment.

Capacity management

ESS Expert provides information that storage administrators can use to manage the ESS total, assigned, and free storage capacity. Some of the information that ESS Expert provides includes:

- ▶ Storage capacity including:
 - Storage that is assigned to application server hosts
 - Storage that is free space
- ▶ Capacity assigned to each application server, capacity shared with other application servers, and the names and types of each application server that can access that ESS.
- ▶ A view of volumes per host. This view lists the volumes accessible to a particular SCSI or Fiber Channel-attached host for each ESS.
- ▶ A volume capability that provides information about a particular volume. This function also identifies all the application server hosts that can access it.
- ▶ Trends and projections of total, assigned, and free space over time for each ESS.
- ▶ Trends in growth.

A LUN-to-host disk mapping feature provides complete information on the logical volumes, including the physical disks and adapters from the host perspective.

The capacity tracking and reporting capabilities of ESS Expert help identify when additional space must be added to the ESS. This can help administrators to be proactive and avoid application outages due to out-of-space conditions.

Performance management

The ESS Expert gathers performance information from the ESS and stores it in a relational database. Administrators can generate and view reports of performance information to help them make informed decisions about volume placement and capacity planning, as well as identifying ways to improve ESS performance. The administrator can schedule when performance data collection commences and the length of time that data is gathered. The information is summarized in reports that can include:

- ▶ Number of I/O requests for the entire storage server in total and separated among various physical disk groupings
- ▶ Read and write cache hit ratio
- ▶ Cache to and from disk operations (stage/destage)
- ▶ Disk read and write response time
- ▶ Disk utilization
- ▶ Number of disk I/O requests for each array, and for each device adapter (the internal adapter that provides I/O to the physical disks)

Storage administrators can use this information to determine ways to improve ESS performance, make decisions about where to allocate new space, and identify time periods of heavy usage.

SNMP ALERTS

ESS Expert provides a Simple Network Management Protocol (SNMP) agent to send an alert to an SNMP management application when an exception event occurs. The threshold values for key performance metrics are set by the Expert user or administrator and an SNMP alert is triggered when the threshold value is exceeded. Additionally, the ability to enable or disable the SNMP alert delivery for each ESS threshold setting is provided.

5.10.2 How does the ESS Expert work

Figure 5-8 shows where the ESS Expert resides and how it interfaces with your ESSs.

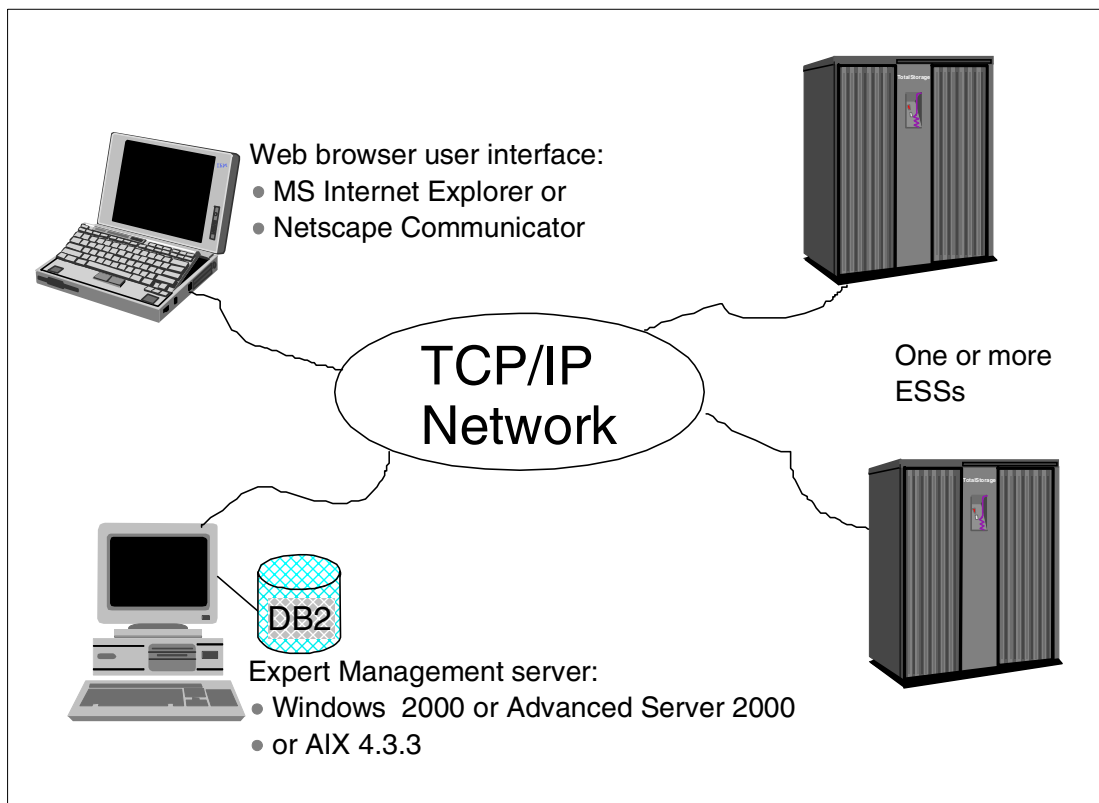


Figure 5-8 How the Expert communicates with the ESS

The ESS Expert itself runs on a Windows 2000 or AIX V4.3.3-based server. However, you do not have to operate the ESS Expert on the server where it is installed, because the user interface to the ESS Expert is via a Web browser interface. In other words, a user can operate the ESS Expert remotely using Netscape Navigator or Microsoft Internet Explorer.

The ESS Expert communicates with your ESSs through a TCP/IP network. Therefore, you can gather information on ESSs across your local or wide area network as long as you have a communication path between the ESS Expert and the ESSs.

The ESS Expert collects information from your ESSs about their capacity and/or performance. It then stores this information into a DB2 database and produces a variety of reports.

5.10.3 ESS Expert performance reports

The IBM TotalStorage Expert (ESS Expert feature) provides three types of reports:

- ▶ Disk Utilization
 - Number of I/O requests, average response time, average utilization
- ▶ Disk to/from Cache
 - Number of tracks/second transferred from disk to cache, and cache to disk

- ▶ Cache
 - Cache hit ratios, time that data remains in cache, percentage of I/Os delayed due to NVS space unavailable, percentage of each type of I/O completed (record or block mode reads, standard non-sequential (partial track) and sequential reads and writes)

Each report type can then display the performance information from four ESS hardware levels or viewpoints:

- ▶ Cluster reports show entire ESS subsystem performance.
- ▶ SSA Device Adapter reports show I/O distribution across clusters.
- ▶ Disk Group reports show I/O distribution in a device adapter.
- ▶ Logical Volume reports show I/O distribution in a disk group.

The ability to report on performance from different hardware viewpoints allows you to “drill down” to find potential problem areas. For example, you may want to know what the disk drive utilization for a given RAID disk group is, and which logical volume in the group has the highest activity, to determine whether or not some of its files should be moved to other arrays, in order to alleviate excess I/O activity.

Cluster reports

An ESS has two clusters, and each cluster independently provides major functions for the storage subsystem. Some examples include directing host adapters for data transferring to/from host processors, managing cache resources, and directing lower device interfaces for data transferring to/from back-end disk drives.

Thus the Cluster level report gives you a performance overview from the viewpoint of the ESS subsystem. Please note that Disk Utilization reports do not have Cluster level reports, because Disk Utilization reports deal with physical device performance.

SSA Device adapter reports

Each cluster on an ESS has four device adapters. Each device adapter in a cluster is paired with the other cluster’s device adapter, and each device adapter pair has up to two SSA disk loops. Though they make a pair, they usually work independently, and each manages separate groups of DDMs under normal operation. Device Adapter level reports help you understand the I/O workload distribution among device adapters/loops on either cluster of the ESS.

Disk Group reports

A disk group is an array of DDMs that are on the same SSA loop. RAID ranks are configured on the disk groups. The rank belongs to a device adapter, and the device adapter controls it under normal operation. The Disk Group level of reports helps you understand I/O workload distribution among the various disk groups in a loop connected to a certain device adapter.

Logical volume reports

A logical volume belongs to a disk group. Host systems view logical volumes as their logical devices. A disk group has multiple logical volumes, depending on your definitions. So this level of report helps you understand I/O workload distribution among logical volumes in a disk group. Please note that Disk Utilization reports do not have this level of report, because Disk Utilization reports deal with physical device performance.

5.11 z/OS environment tools

This section discusses the performance tools specifically available for z/OS users.

RMF

If you are in a z/OS environment, RMF will help you look at the ESS and discover how well logical volumes and control units are performing. You will get information such as I/O rate at the device and logical control unit level, and response time for specific volumes, as well as information on the number of I/O requests that are completed within cache memory (cache hits).

RMF will report on ESS performance in a similar way to a 3990. Some specific considerations apply for ESS. Device addresses are reported with response time, connect, disconnect, PEND and IOSQ times as usual. Alias addresses for PAVs are not reported, but RMF will report the number of aliases (or in RMF terms, exposures) that have been used by a device and whether the number of exposures has changed during the reporting interval (the MX field shows the number of exposures, which is the number of aliases + 1, the base). RMF cache statistics are collected and reported by logical subsystem (*LCU* in zSeries terminology). So a fully configured ESS will produce 16 sets of cache data.

Note: For FICON attachment, there are changes in the way the components of the I/O operations add up to the values reported by RMF. You may refer to 6.6.2, “FICON I/O operation” on page 186 for detailed explanation of the FICON I/O operation components.

RMF works together with the ESS to retrieve performance statistics data and reports on the cache workload and operations. However, these cache statistics will be related to the activity of the z/OS images only, and will not show the activity of any other zSeries (VM, VSE, Linux) or open systems servers that may be attached to the same ESS. To get a global view of the ESS performance, the IBM TotalStorage Expert (ESS Expert feature) will be required as it gathers information from the ESS components regardless of host server type or operating system.

ESS cache and NVS reporting

RMF cache reporting and the results of a LISTDATA STATUS command report a cache size of 4, 8, 12, 16, 32 GB and an NVS size of 1 GB — half the actual size. This is because the information returned represents only the cluster to which the logical control unit is attached. Each LCU on the cluster reflects the cache and NVS size of that cluster. z/OS users will find that only the SETCACHE CFW ON | OFF command is supported, while other SETCACHE command options (for example, DEVICE, SUBSYSTEM, DFW, NVS) are not accepted.

FICON

RMF has been changed to support FICON channels. With APAR OW35586, RMF extends the information in the Channel Path Activity report of all monitors by reporting about data transfer rates and bus utilization values for FICON channels. There are five new measurements reported by RMF in a FICON channel environment:

- ▶ Bus utilization
- ▶ Read bandwidth for a partition in MBps
- ▶ Read bandwidth total (all logical partitions on the processor) in MBps
- ▶ Write bandwidth for a partition in MBps
- ▶ Write bandwidth total in MBps

Also there has been some SMF record changes for FICON. For more information about the new fields reported by RMF for FICON channels and SMF record changes, refer to *FICON Native Implementation and Reference Guide*, SG24-6266.



zSeries performance

The IBM TotalStorage Enterprise Storage Server Model 800 delivers a multi-feature synergy with the zSeries servers running the z/OS operating system, allowing an unprecedented breakthrough in performance for those environments. These performance features are presented in detail in this chapter.

6.1 Overview

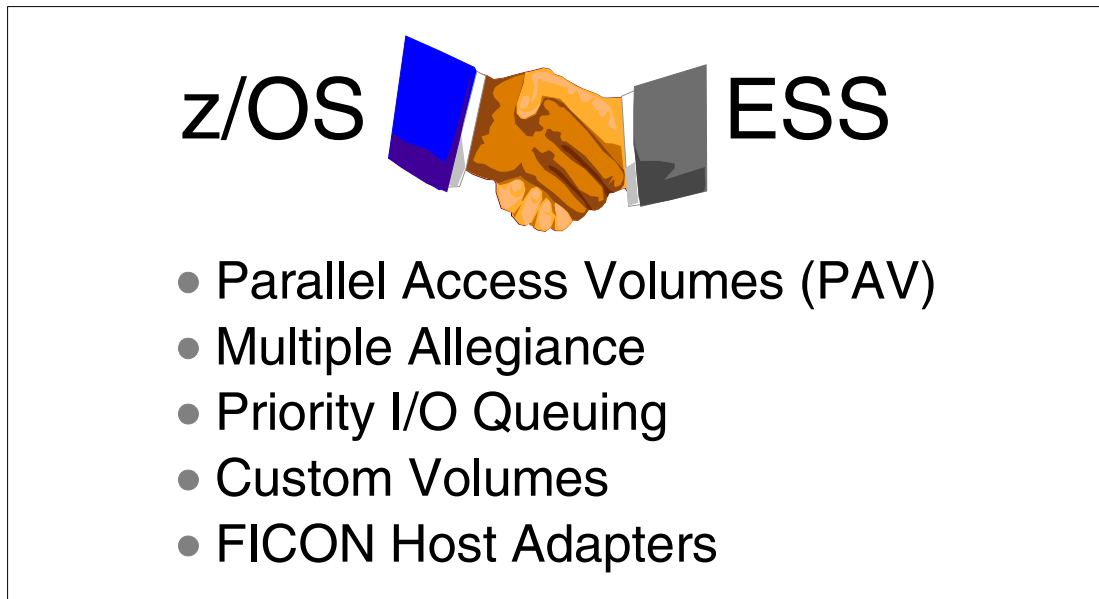


Figure 6-1 zSeries specific performance features

The IBM TotalStorage Enterprise Storage Server Model 800, with its architecture that efficiently integrates all the performance features and functions we presented in Chapter 5, “Performance” on page 147, provides the highest levels of performance across all the sever platforms it attaches.

The IBM TotalStorage Enterprise Storage Server Model 800 also keeps from the previous models the set of performance features specifically available for the zSeries users. The cooperation between these ESS unique features and the zSeries operating systems (mainly the z/OS operating system) makes the ESS performance even more outstanding in the zSeries environment.

These are the features that are described in this chapter:

- ▶ Parallel Access Volumes (PAV)
- ▶ Multiple Allegiance
- ▶ I/O Priority Queuing
- ▶ Custom volumes
- ▶ FICON Host Adapters

6.2 Parallel Access Volume

Parallel Access Volume (PAV) is one of the original features that the IBM TotalStorage Enterprise Storage Server brings specifically for the z/OS and OS/390 operating systems, helping the zSeries running applications to concurrently share the same logical volumes.

The ability to do multiple I/O requests to the same volume nearly eliminates IOSQ time, one of the major components in z/OS response time. Traditionally, access to highly active volumes has involved manual tuning, splitting data across multiple volumes, and more. With PAV and the Workload Manager, you can almost forget about manual performance tuning. WLM manages PAVs across all the members of a sysplex too. The ESS in conjunction with z/OS has the ability to meet the performance requirements by its own.

Before we see PAV in detail, let us review how z/OS traditionally behaves when more than one application needs concurrently to access a logical volume.

6.2.1 Traditional z/OS behavior

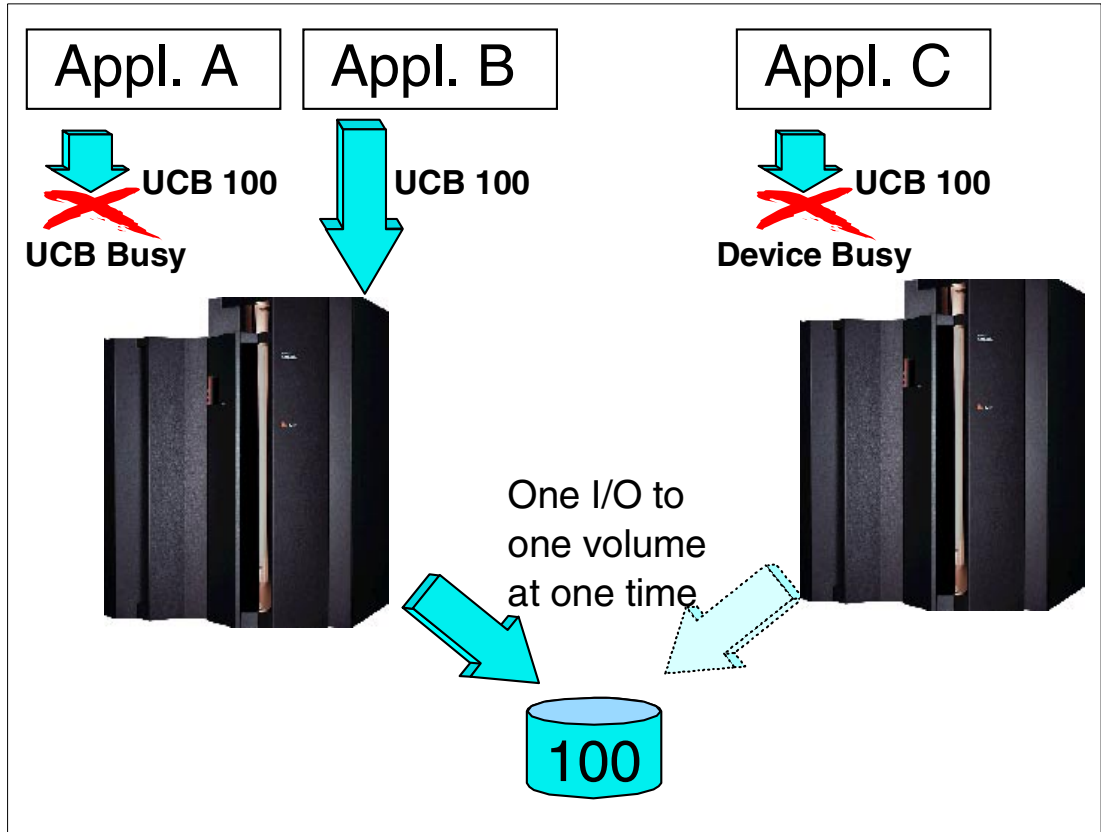


Figure 6-2 Traditional z/OS behavior

Traditional DASD subsystems (here, we use the term DASD - Direct Access Storage Device - instead of logical volume since the term DASD is more common among zSeries users) have allowed for only one channel program to be active to a DASD volume at a time, in order to ensure that data being accessed by one channel program cannot be altered by the activities of some other channel program.

From a performance standpoint, it did not make sense to send more than one I/O at a time to the storage subsystem, because the DASD hardware could process only one I/O at a time.

Knowing this, the z/OS systems did not try to issue another I/O to a DASD volume — in z/OS represented by a Unit Control Block (UCB) — while an I/O was already active for that volume, as indicated by a UCB busy flag (see Figure 6-2).

Not only were the z/OS systems limited to processing only one I/O at a time, but also the storage subsystems accepted only one I/O at a time from different system images to a shared DASD volume, for the same reasons mentioned above.

6.2.2 Parallel I/O capability

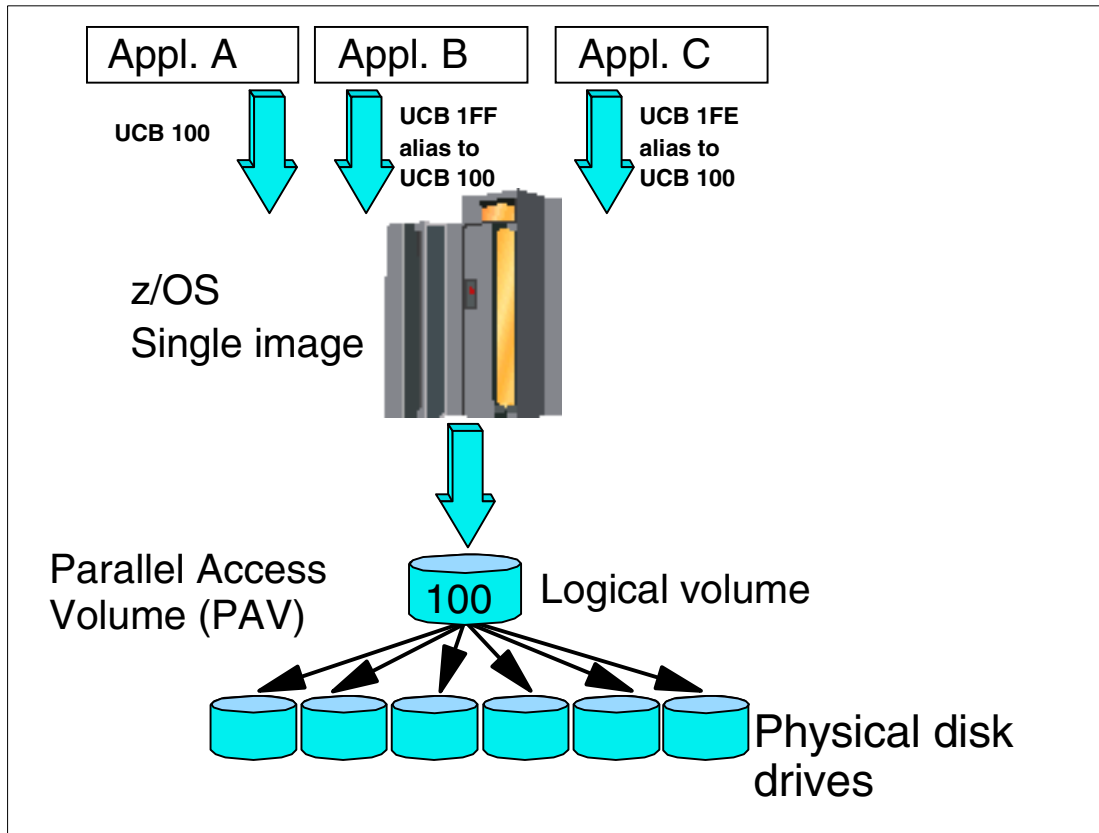


Figure 6-3 Parallel I/O capability using PAV

The IBM TotalStorage Enterprise Storage Server Model 800 is a modern storage subsystem with large cache sizes and disk drives arranged in RAID arrays. Cache I/O is much faster than disk I/O, no mechanical parts (actuator) are involved. And I/Os can take place in parallel, even to the same volume. This is true for reads, and it is also possible for writes, as long as different extents on the volume are accessed.

The ESS emulates zSeries ECKD volumes over RAID 5 or RAID 10 disk arrays. While the zSeries operating systems continue to work with these logical DASD volumes as a unit, its tracks are spread over several physical disk drives. So parallel I/O from different applications to the same logical volume would also be possible for cache misses (when the I/Os have to go to the disk drives, involving mechanical movement of actuators) as long as the logical tracks are on different physical disk drives.

The ESS has the capability to do more than one I/O to an emulated CKD volume. The ESS introduces the new concept of alias address, in addition to the conventional base address. This allows a z/OS host to now use several UCBs for the same logical volume instead of one UCB per logical volume. For example, base address 2C00 may now have alias addresses 2CFD, 2CFE and 2CFF. This allows for four parallel I/O operations to the same volume.

This feature that allows parallel I/Os to a volume from one host is called Parallel Access Volume (PAV).

6.2.3 Benefits of Parallel Access Volume

- Multiple UCBs per logical volume
- PAVs allow simultaneous access to logical volumes by multiple users or jobs from one system
- Reads are simultaneous
- Writes to different domains are simultaneous
- Writes to same domain are serialized
- Eliminates or sharply reduces IOSQ

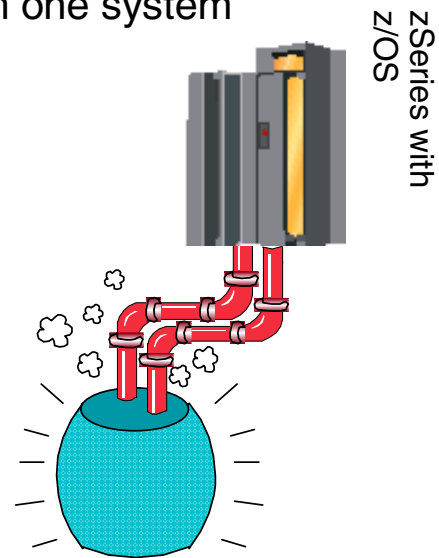


Figure 6-4 Parallel Access Volume (PAV) benefits

z/OS systems queue I/O activity on a Unit Control Block (UCB) that represents the logical device. High I/O activity, particularly to large volumes (3390 model 9), could adversely affect performance, because the volumes are treated as a single resource and serially reused. This could result in large IOSQ times (IOSQ shows these queued I/Os average waiting time). This is because traditionally the operating system does not attempt to start more than one I/O operation at a time to the logical device. If an application initiates an I/O to a device that is already busy, the I/O request will be queued until the device completes the other I/O and becomes available.

The IBM TotalStorage Enterprise Storage Server is capable of handling parallel I/Os with PAV. The definition and exploitation of ESS's Parallel Access Volumes requires the operating system software support to define and manage alias device addresses and subchannels. The z/OS and OS/390 operating systems have this support and can issue multiple channel programs to a volume, allowing simultaneous access to the logical volume by multiple users or jobs. Reads can be satisfied simultaneously, as well as writes to different domains. The domain of an I/O consists of the specified extents to which the I/O operation applies. Writes to the same domain still have to be serialized to maintain data integrity.

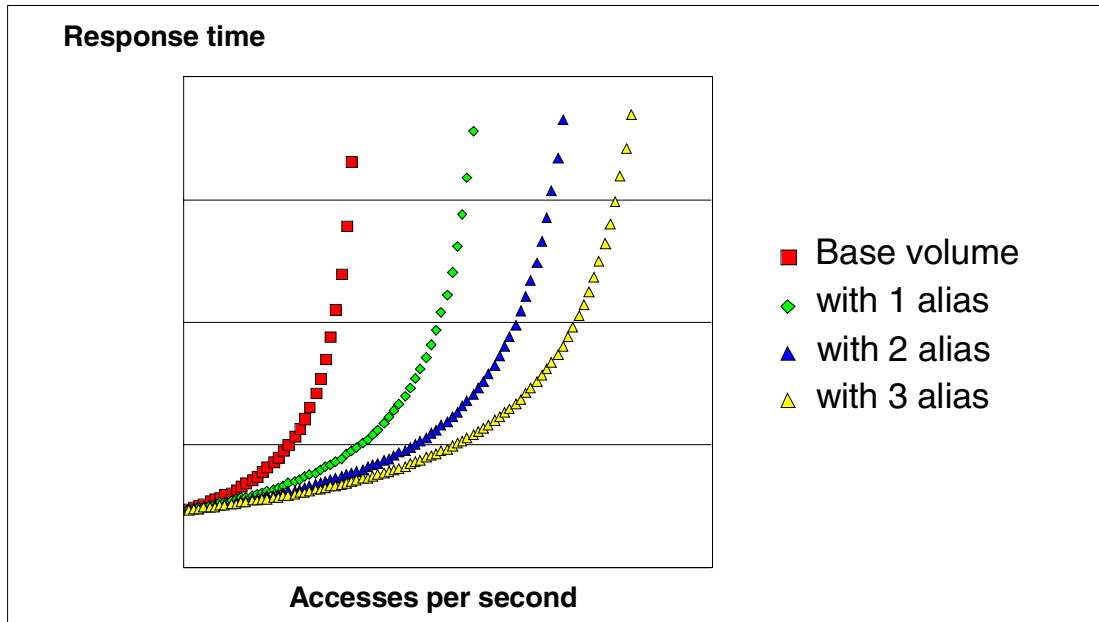


Figure 6-5 Potential performance impact of PAV

ESS's parallel I/O capability can drastically reduce or eliminate IOSQ time in the operating system, allowing for much higher I/O rates to a logical volume, and hence increasing the overall throughput of an ESS (see Figure 6-5). This cooperation of the IBM TotalStorage Enterprise Storage Server and the IBM operating systems z/OS and OS/390 provides additional value to your business.

6.2.4 PAV base and alias addresses

<p><u>Alias address</u></p> <ul style="list-style-type: none"> • Maps to base address • IO operation runs against the base • No physical space reserved • Alias are visible to z/OS only 	<p><u>Base address</u></p> <ul style="list-style-type: none"> • Actual unit address of the volume • One base address per volume • Volume space is associated to base only • Reserve/release only to base
---	---

IOCP

```

CNTLUNIT CUNUMBR=9A0,PATH=(A8,AA,B7,B8,B9,BA,BB),
UNITADD=((00,128)),UNIT=2105,CUADD=0
IODEVICE UNIT=3390B,ADDRESS=(B00,40),CUNUMBR=(9A0),
UNITADD=00,STADET=YES,FEATURE=SHARED
IODEVICE UNIT=3390A,ADDRESS=(B28,88),CUNUMBR=(9A0),
UNITADD=28,STADET=YES,FEATURE=SHARED
  
```

Base Alias

Figure 6-6 PAV base and alias addresses

The IBM TotalStorage Enterprise Storage Server implementation of PAV introduces two new unit address types: *base* device address and *alias* device address. Several aliases can be assigned to one base address. Therefore, there can be multiple unit addresses (and hence UCBs) for a volume in z/OS, and z/OS can use all these addresses for I/Os to that logical volume.

Base address

The base device address is the conventional unit address of a logical volume. There is only one base address associated with any volume. Disk storage space is associated with the base address. In commands, where you deal with unit addresses (for example, when you set up a PPRC pair) you use the base address.

Alias address

An alias device address is mapped to a base address. I/O operations to an alias address run against the associated base address storage space. There is no physical space associated with an alias address. You can define more than one alias per base. The ESS Specialist allows you to define up to 255 aliases per base, and the maximum devices (alias plus base) is 256 per LCU (see “Configuring CKD base and alias addresses” on page 129). Alias addresses are visible only to the I/O Supervisor (IOS). Alias UCBs have the same memory storage requirements as base addresses.

Alias addresses have to be defined to the ESS and to the zSeries host IODF file. The number of base and alias addresses in both definitions, in the ESS and in the IODF, must match. Base and alias addresses are defined to the host system using the IOCP IODEVICE macro specifying UNIT=3390B and UNIT=3390A respectively (see Figure 6-6 on page 170).

6.2.5 PAV tuning

Alias reassignment

- Association between base and alias is predefined
- Predefinition can be changed

Automatic PAV tuning

- Association between base and its aliases is automatically tuned
- WLM in Goal mode manages the assignment of alias addresses
- WLM instructs IOS when to reassign an alias

Figure 6-7 PAV tuning

The association between base addresses and alias addresses is predefined in the ESS by using the ESS Specialist. Adding new aliases can be done nondisruptively. The ESS Specialist allows the definition of 0 to 255 aliases per base. So you can have any combination

of the number of alias and base for a given LSS, within the 256 device number limit of the LCU.

The association between base and alias addresses is not fixed. Alias addresses can be assigned to different base addresses by the z/OS Workload Manager. If the Workload Manager is not being used, then the association becomes a static definition. This means that for each base address you define, the quantity of associated alias addresses will be what you already specified in the IODF and the ESS Specialist when doing the logical configuration procedure.

Dynamic PAV tuning

It will not always be easy to predict which volumes should have an alias address assigned, and how many. Your software can automatically manage the aliases according to your goals. z/OS can exploit automatic PAV tuning if you are using the Workload Manager (WLM) in Goal mode. The WLM can dynamically tune the assignment of alias addresses. The Workload Manager monitors the device performance and is able to dynamically reassign alias addresses from one base to another if predefined goals for a workload are not met. The WLM instructs the IOS to reassign an alias (WLM dynamic tuning is explained later in 6.2.9, “WLM support for dynamic PAVs” on page 174).

6.2.6 Configuring PAVs

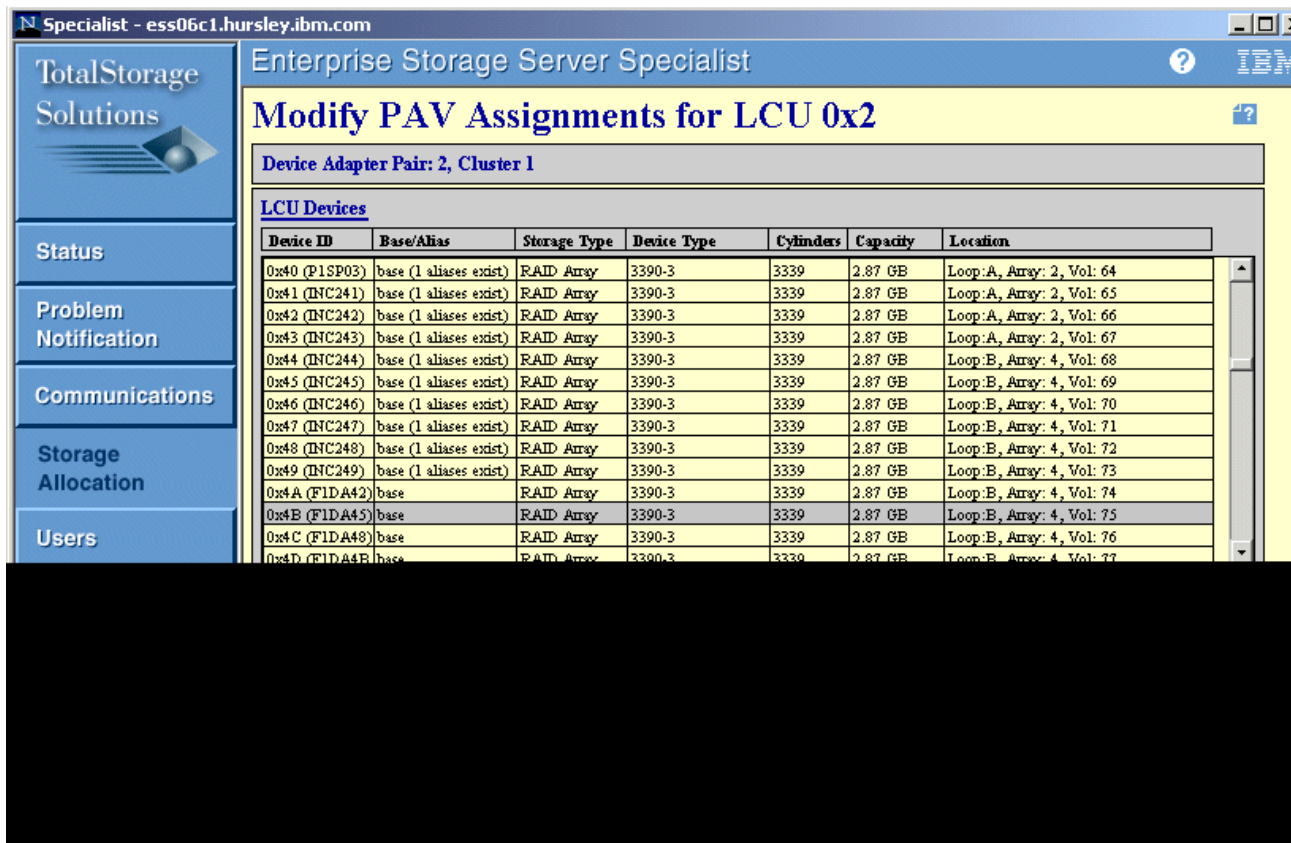


Figure 6-8 Modify PAV Assignments window

Before PAVs can be used, they must first be defined to the ESS using the ESS Specialist (see Figure 6-8), and to the host IODF file using the Hardware Configuration Definition (HCD)

utility. Both the ESS and the z/OS HCD definitions must match; otherwise you will get an error message.

When defining PAV volumes to the ESS, and to HCD, you specify a new device type: the 3390B (or 3380B) for a PAV base, and 3390A (or 3380A) for a PAV alias. Device support UIMs support these new PAV device types.

You can enable or disable the use of dynamic PAVs. In your HCD definition, you can specify WLMPAV = YES | NO. If you stay with the default (WLMPAV=YES), dynamic PAV management by WLM is enabled. In a Parallel Sysplex, if dynamic PAV management is specified for one of the systems, then it is enabled for all the systems in the sysplex, even if they specify WLMPAV=NO.

An alias must not be defined for MIH in the IECIOS member. An alias should not be initialized, because they are only known to the I/O Supervisor routines of the z/OS.

The association of alias address (one or more) to base address in the ESS is done using the ESS Specialist Modify PAV Assignment window (see Figure 6-8 on page 172). You can modify PAV assignments only after you have created the LCU, defined disk groups, and created base volumes.

6.2.7 Querying PAVs

```

DS QPAVS,D222,VOLUME

IEE459I 08.20.32 DEVSERV QPATHS 591
      Host                               Subsystem
Configuration                           Configuration
-----
UNIT                                     UNIT   UA
NUM. UA  TYPE          STATUS          SSID  ADDR.  TYPE
-----
D222 22  BASE
D2FE FE  ALIAS-D222
D2FF FF  ALIAS-D222
***          3 DEVICE(S) MET THE SELECTION CRITERIA

```

Figure 6-9 Querying PAVs

You can use the DEVSERV QPAVS command to verify the PAV definitions (see Figure 6-9). This command shows you the unit addresses currently assigned to a base. If the hardware and software definitions do not match, the STATUS field will not be empty, but rather will contain a warning, such as INV-ALIAS for an invalid alias or NOT-BASE if the volume is not a PAV volume.

Note that the DEVSERV command shows the unit number and unit address. The unit number is the address, used by z/OS. This number could be different for different hosts accessing the same logical volume. The unit address is an ESS internal number used to unambiguously identify the logical volume.

In a reverse way, if you are needing to know which base address is associated to a given alias address, then you may find the D M=DEV(alias-address) command more useful.

Note: Remember that alias addresses cannot be used in such commands as D U, VARY, and so on. For z/VM, where you have PAV guest support, you also have a Q PAV command (see “PAV for VM/ESA guests” on page 178).

6.2.8 PAV assignment

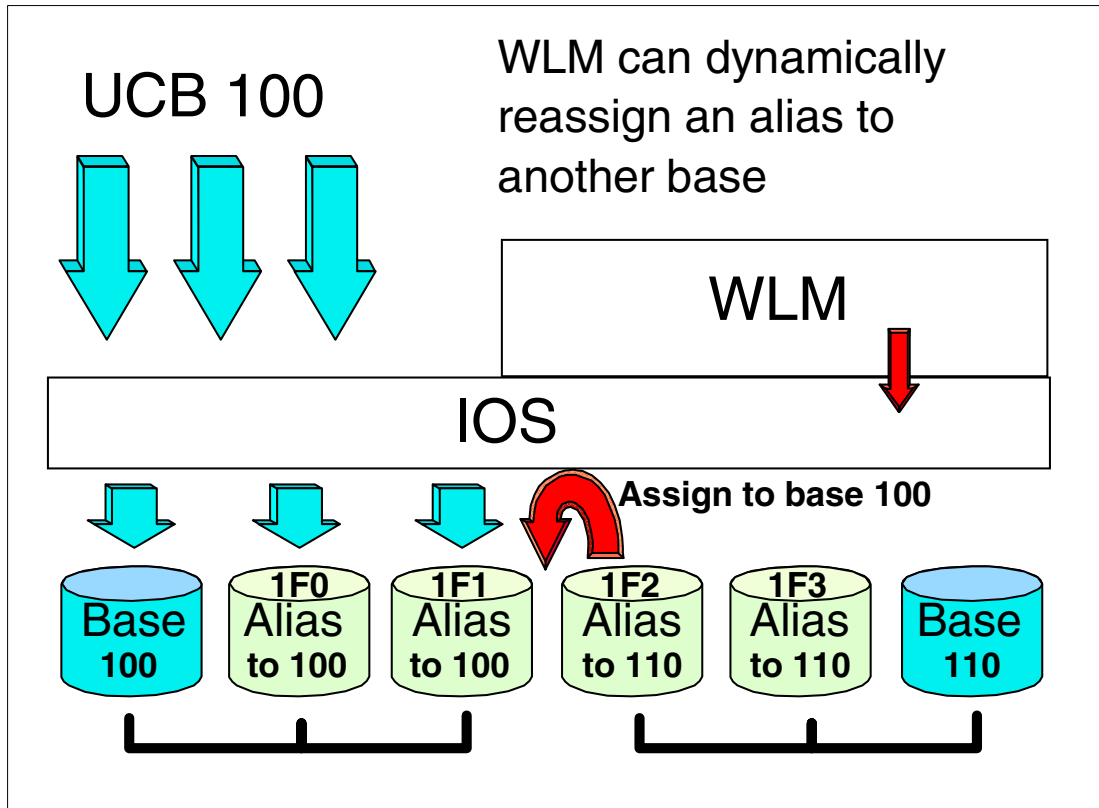


Figure 6-10 Assignment of alias addresses

z/OS recognizes the aliases that are initially assigned to a base during the NIP (Nucleus Initialization Program) phase. If dynamic PAVs are enabled, the WLM can reassign an alias to another base by instructing the IOS to do so when necessary (see Figure 6-10).

6.2.9 WLM support for dynamic PAVs

z/OS's Workload Manager in Goal mode tracks the system workload and checks if the workloads are meeting their goals established by the installation.

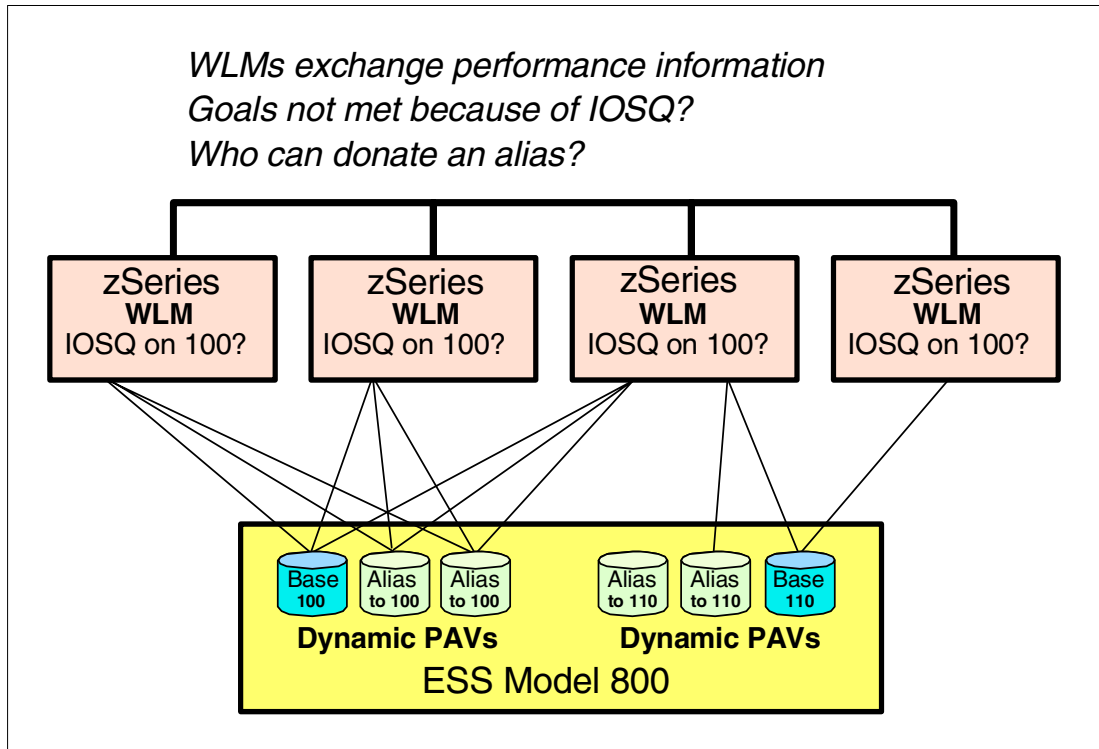


Figure 6-11 Dynamic PAVs in a sysplex

WLM also keeps track of the devices utilized by the different workloads, accumulates this information over time, and broadcasts it to the other systems in the same sysplex. If WLM determines that any workload is not meeting its goal due to IOSQ time, WLM will attempt to find an alias device that can be reallocated to help this workload achieve its goal (see Figure 6-11).

Actually there are two mechanisms to tune the alias assignment:

1. The first mechanism is goal based. This logic attempts to give additional aliases to a PAV-enabled device that is experiencing IOS queue delays and is impacting a service class period that is missing its goal. To give additional aliases to the receiver device, a donor device must be found with a less important service class period. A bitmap is maintained with each PAV device that indicates the service classes using the device.
2. The second is to move aliases to high-contention PAV-enabled devices from low-contention PAV devices. High-contention devices are identified by having a significant amount of IOS queue time (IOSQ). This tuning is based on efficiency rather than directly helping a workload to meet its goal. Because adjusting the number of aliases for a PAV-enabled device affects any system using the device, a sysplex-wide view of performance data is important, and is needed by the adjustment algorithms. Each system in the sysplex broadcasts local performance data to the rest of the sysplex. By combining the data received from other systems with local data, WLM can build the sysplex view.

Aliases of an offline device will be considered unbound. WLM uses unbound aliases as the best donor devices. If you run with a device offline to some systems and online to others, you should make the device ineligible for dynamic WLM alias management in HCD.

RMF reports the number of exposures for each device in its Monitor/DASD Activity report and in its Monitor II and Monitor III Device reports. RMF also reports which devices had a change in the number of exposures.

RMF reports all I/O activity against the base address — not by the base and associated aliases. The performance information for the base includes all base and alias activity.

As mentioned before, the Workload Manager (WLM) must be in Goal mode to cause PAVs to be shifted from one logical device to another.

Further information regarding WLM can be found at the WLM/SRM site at:

<http://www.ibm.com/s390/wlm/>

6.2.10 Reassignment of a PAV alias

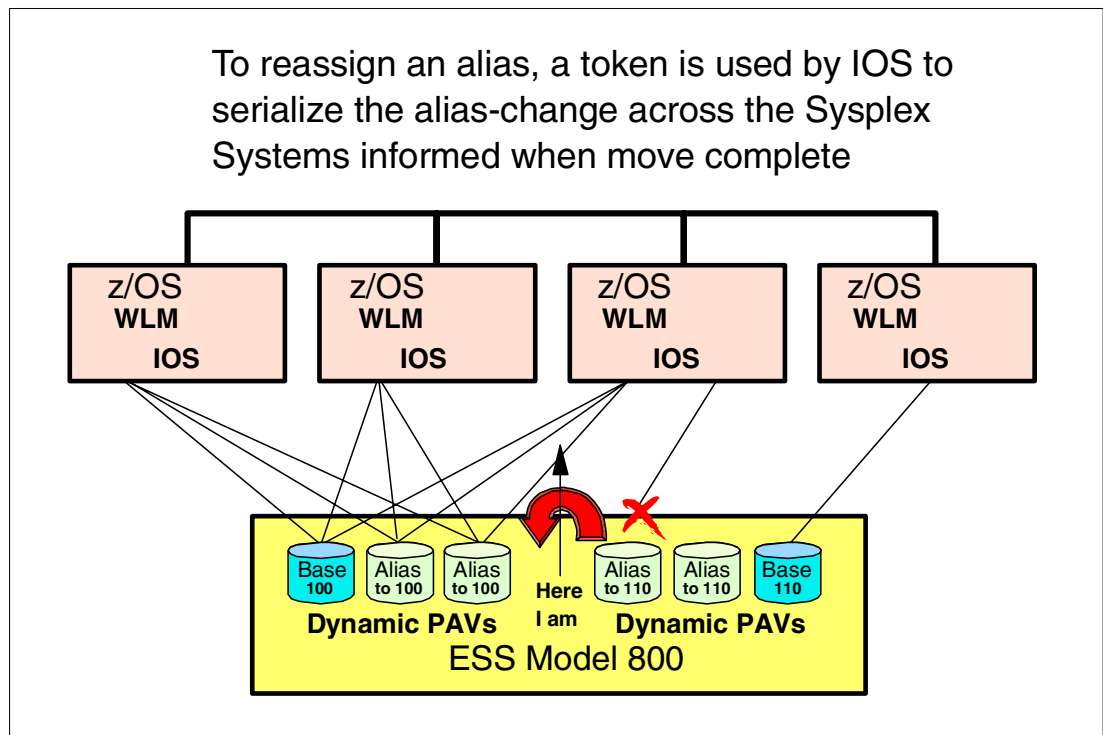


Figure 6-12 Reassignment of dynamic PAV alias

The movement of an alias from one base to another is serialized within the sysplex. IOS tracks a token for each PAV-enabled device. This token is updated each time an alias change is made for a device. IOS and WLM exchange the token information. When the WLM instructs IOS to move an alias, WLM also presents the token. When IOS has started a move and updated the token, all affected systems are notified of the change through an interrupt.

6.2.11 Mixing PAV types

```

Coefficients/Options  Notes  Options  Help
-----
                Service Coefficients/Service Definition Options
Command ==>_____

Enter or change the Service Coefficients:

CPU . . . . . _____ (0.0-99.9)
IOC . . . . . _____ (0.0-99.9)
MSO . . . . . _____ (0.0000-99.9999)
SRB . . . . . _____ (0.0-99.9)

Enter or change the service definition options:

I/O priority management . . . . . NO (Yes or No)
Dynamic alias tuning management. . . . . YES (Yes or No)

```

Figure 6-13 Activation of dynamic alias tuning for the WLM

With HCD, you can enable or disable dynamic alias tuning on a device-by-device basis. On the WLM's Service Definition ISPF window, you can globally (sysplex-wide) enable or disable dynamic alias tuning by the WLM as Figure 6-13 shows. This option can be used to stop WLM from adjusting the number of aliases in general, when devices are shared by systems that do not support dynamic alias tuning (WLM not operating in Goal mode).

If you enable alias tuning (this is the default in WLM) for devices shared by zSeries hosts with WLM not in Goal mode, and therefore only supporting static PAVs, these systems still recognize the change of an alias and use the new assigned alias for I/Os to the associated base address. However, the WLMs on the systems that do support the dynamic alias tuning will not see the I/O activity to and from these static PAV-only systems to the shared devices. Therefore, the WLMs can not take into account this hidden activity when making their judgements.

Without a global view, the WLMs could make a wrong decision. Therefore, you should not use dynamic PAV alias tuning for devices from one system and static PAV for the same devices on another system.

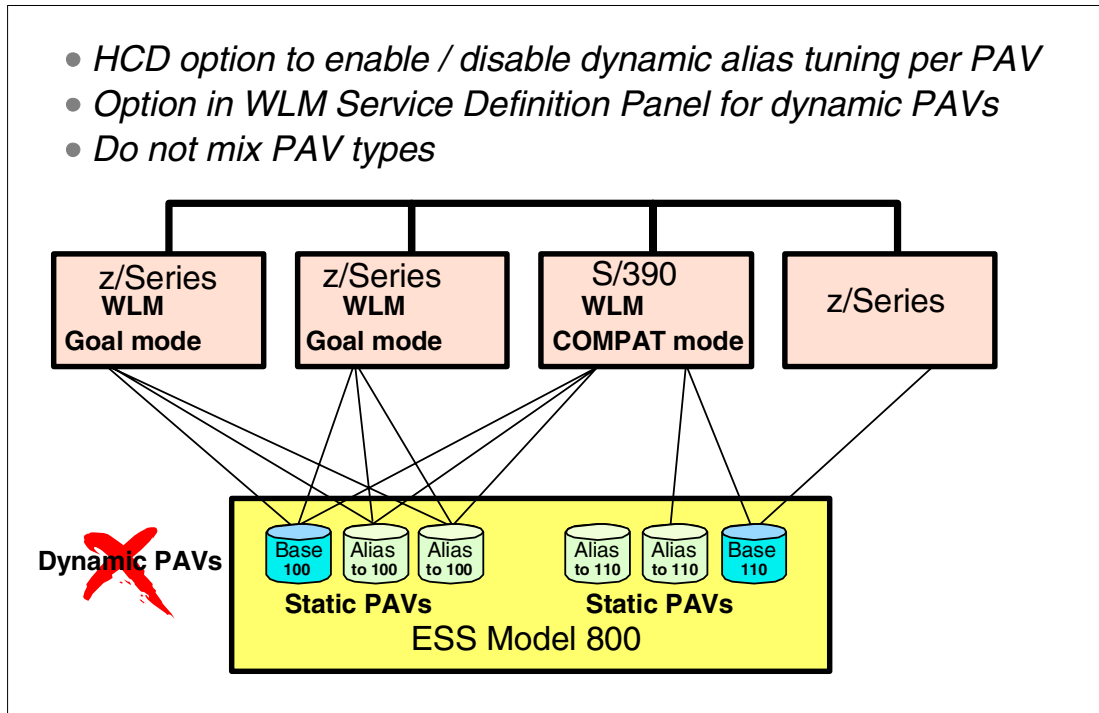


Figure 6-14 No mixing of PAV types

If at least one system in the sysplex specifies dynamic PAV management, then it is enabled for all the systems in the sysplex. There is no consistency checking for this parameter. It is an installation's responsibility to coordinate definitions consistently across a sysplex. WLM will not attempt to enforce a consistent setting of this option (see Figure 6-14).

6.2.12 PAV support

PAV is supported by the zSeries servers running z/OS or OS/390, natively or under VM.

PAV for z/OS

z/OS and OS/390 support the PAV function. The *Preventive Service Planning* (PSP) buckets contain operating system support and planning information that includes *application program analysis reports* (APARs) and *program temporary fixes* (PTFs) required for PAV support in the z/OS environment. They are available from your IBM Service Representative or by contacting the IBM Software Support Center — for the ESS ask for the 2105DEVICE PSP bucket, subset name for z/OS is 2105MVS/ESA. Refer to 8.3, “z/OS support” on page 232 for information on z/OS and related product support.

PAV for VM/ESA guests

z/VM has not implemented the exploitation of PAV for itself. However, with VM/ESA 2.4.0 with an enabling APAR, z/OS and OS/390 guests can use PAV volumes and dynamic PAV tuning. Alias and base addresses must be attached to the z/OS guest. You need a separate ATTACH for each alias. You should attach the base and its aliases to the same guest.

A base cannot be attached to SYSTEM if one of its aliases is attached to that guest. This means that you cannot use PAVs for Full Pack minidisks.

There is a new QUERY PAV command available for authorized (class B) users to query base and alias addresses: QUERY PAV rdev and QUERY PAV ALL.

The response for a PAV base is:

```
Device 01D2 is a base Parallel Access Volume device with the following aliases:  
01E6 01EA 01D3
```

The response for a PAV alias is:

```
Device 01E7 is an alias Parallel Access Volume device whose base device is 01A0
```

Refer to 8.4, “z/VM support” on page 235 support information.

6.3 Multiple Allegiance

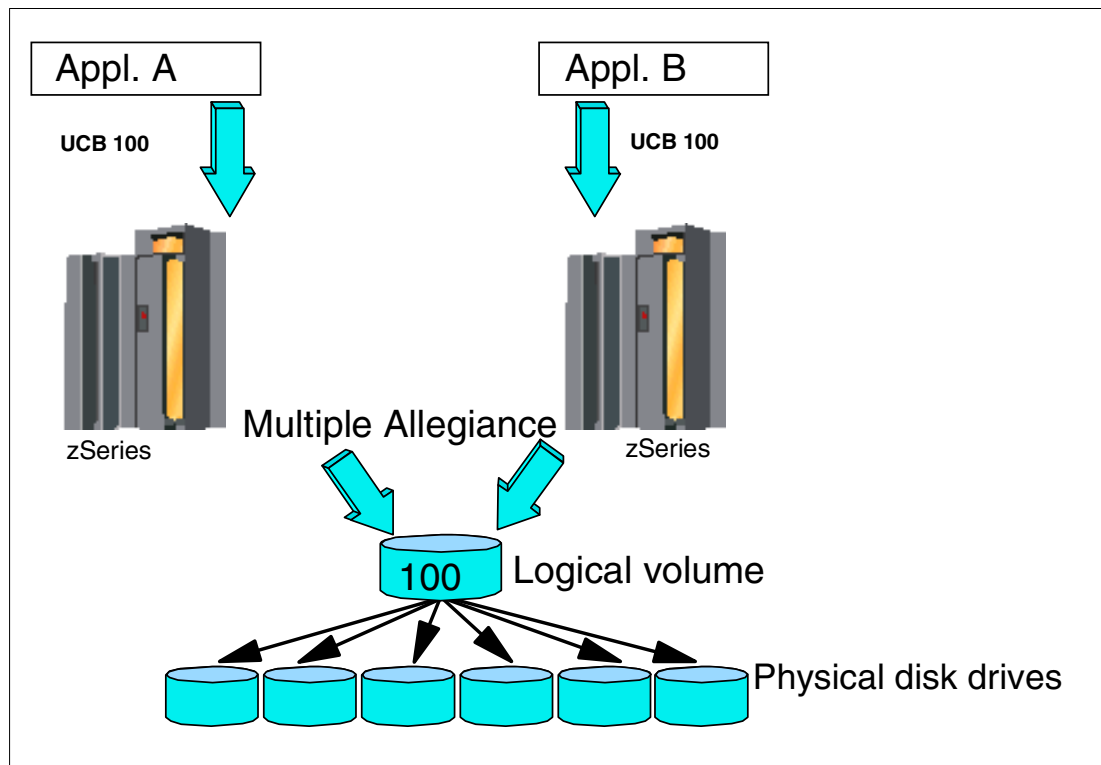


Figure 6-15 Parallel I/O capability with Multiple Allegiance

Normally, if any zSeries host image (server or LPAR) does an I/O request to a device address for which the storage subsystem is already processing an I/O that came from another zSeries host image, then the storage subsystem will send back a device-busy indication. This delays the new request and adds to processor and channel overhead (this delay is shown in the RMF Pend time column).

6.3.1 Parallel I/O capability

The IBM TotalStorage Enterprise Storage Server accepts multiple I/O requests from different hosts to the same device address, increasing parallelism and reducing channel overhead.

In previous storage subsystems, a device had an implicit *allegiance*, that is, a relationship created in the control unit between the device and a channel path group when an I/O operation is accepted by the device. The allegiance causes the control unit to guarantee

access (no busy status presented) to the device for the remainder of the channel program over the set of paths associated with the allegiance.

With Multiple Allegiance, the requests are accepted by the ESS and all requests will be processed in parallel, unless there is a conflict when writing data to a particular extent of the CDK logical volume. Still, good application software access patterns can improve the global parallelism by avoiding reserves, limiting the extent scope to a minimum, and setting an appropriate file mask, for example, if no write is intended.

In systems without Multiple Allegiance, all except the first I/O request to a shared volume were rejected, and the I/Os were queued in the zSeries channel subsystem, showing up as PEND time in the RMF reports.

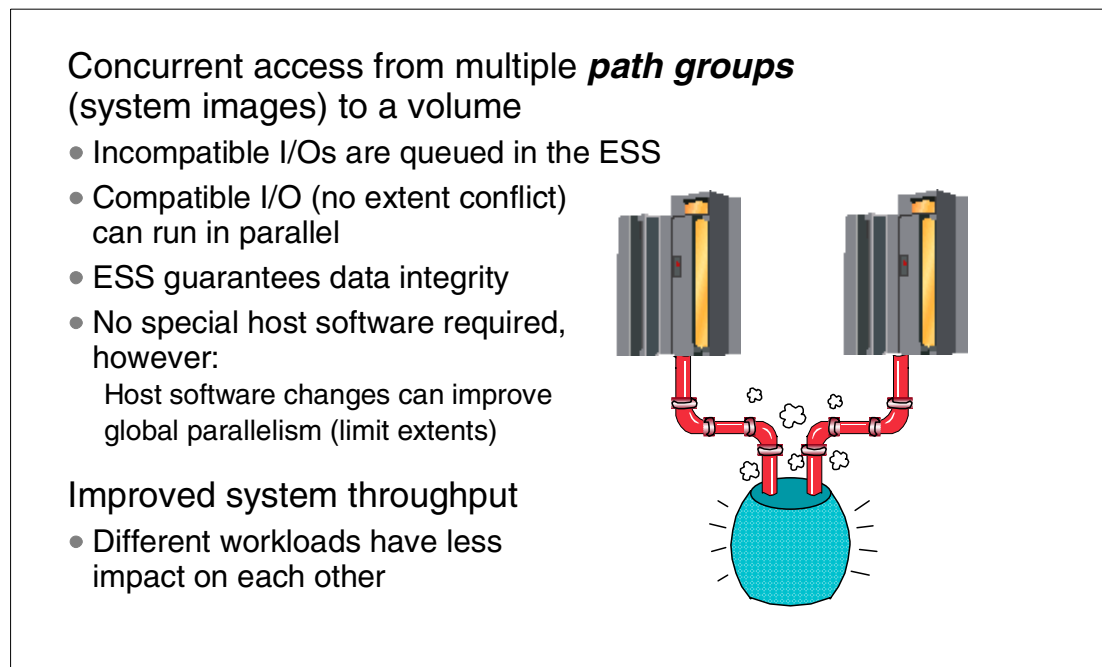


Figure 6-16 Multiple Allegiance

Multiple Allegiance provides significant benefits for environments running a sysplex, or zSeries systems sharing access to data volumes (Figure 6-16). Multiple Allegiance and PAV can operate together to handle multiple requests from multiple hosts.

6.3.2 Eligible I/Os for parallel access

ESS distinguishes between compatible channel programs that can operate concurrently and incompatible channel programs that have to be queued to maintain data integrity. In any case the ESS ensures that, despite the concurrent access to a volume, no channel program can alter data that another channel program is using.

6.3.3 Software support

Basically, there is no software support required for the exploitation of ESS's Multiple Allegiance capability. The ESS storage subsystem looks at the extent range of the channel program and whether it intends to read or to write. Whenever possible, ESS will allow the I/Os to run in parallel.

6.3.4 Benefits of Multiple Allegiance

Performance Environment	Host 1 Online (4K read hits)	Host 2 Data Mining (32 record 4K read chains)
Max ops/sec Isolated	767 SIOs/SEC	55.1 SIOs/sec
Concurrent	59.3 SIOs/SEC	54.5 SIOs/sec
Concurrent with Multiple Allegiance	756 SIOs/SEC	54.5 SIOs/sec

✓ Database, utility programs and extract programs can run in parallel

✓ One copy of data

Figure 6-17 Benefits of Multiple Allegiance for mixing workloads

The ESS ability to run channel programs to the same device in parallel can dramatically reduce IOSQ and pending times in shared environments.

In particular, different workloads—for example, batch and online—running in parallel on different systems can have an unfavorable impact on each other. In such cases, ESS’s Multiple Allegiance can improve the overall throughput. Figure 6-17 shows an example of a DB2 data mining application running in parallel with normal database access.

The application running long CCW chains (Host 2) drastically slows down the online application in the example when both applications try to access the same volume extents.

ESS’s support for parallel I/Os lets both applications run concurrently without impacting each other adversely.

6.4 I/O Priority Queuing

The concurrent I/O capability of the IBM TotalStorage Enterprise Storage Server allows it to execute multiple channel programs concurrently, as long as the data accessed by one channel program is not altered by another channel program.

6.4.1 Queuing of channel programs

When the channel programs conflict with each other and must be serialized to ensure data consistency, the ESS will internally queue channel programs.

This subsystem I/O queuing capability provides significant benefits:

- ▶ Compared to the traditional approach of responding with device-busy status to an attempt to start a second I/O operation to a device, I/O queuing in the storage subsystem eliminates the overhead associated with posting status indicators and re-driving the queued channel programs.
- ▶ Contention in a shared environment is eliminated. Channel programs that cannot execute in parallel are processed in the order they are queued. A fast system cannot monopolize access to a volume also accessed from a slower system. Each system gets a fair share.

6.4.2 Priority queuing

I/Os from different z/OS system images can be queued in a priority order. It is the z/OS Workload Manager that makes use of this priority to privilege I/Os from one system against the others. You can activate I/O Priority Queuing in WLM's Service Definition settings.

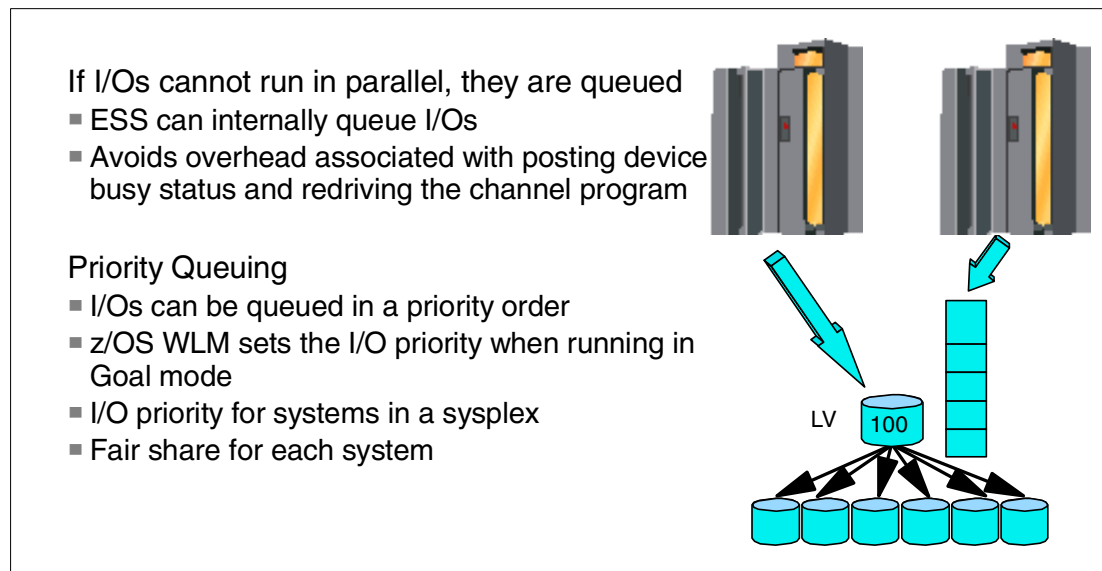


Figure 6-18 I/O queuing

WLM has to run in Goal mode. I/O Priority Queuing is available for z/OS and OS/390.

When a channel program with a higher priority comes in and is put in front of the queue of channel programs with lower priority, the priority of the low-priority programs is also increased. This prevents high-priority channel programs from dominating lower priority ones and gives each system a fair share (see Figure 6-19 on page 183).

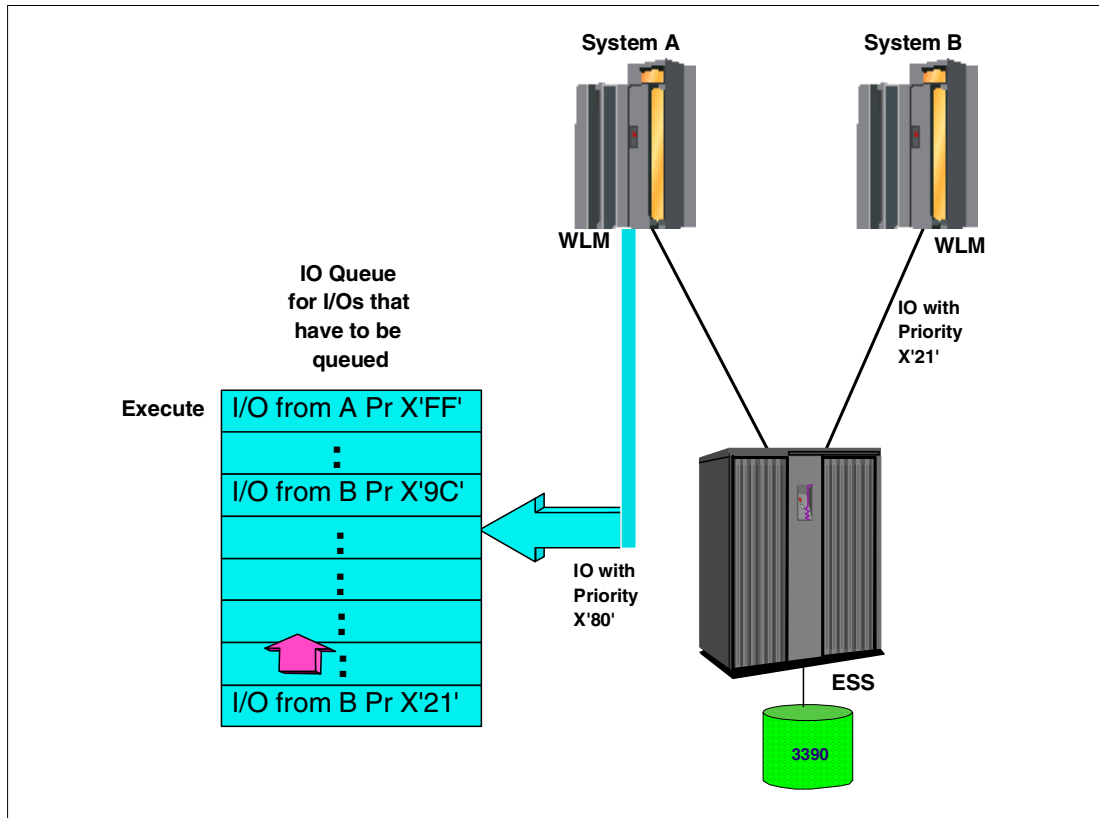


Figure 6-19 I/O Priority Queuing

6.5 Custom volumes

As we have seen, the IBM TotalStorage Enterprise Storage Server is able to do several concurrent I/O operations to a logical volume with its PAV and MA functions. This drastically reduces or eliminates IOS queuing and pending times in the z/OS environments.

Even if you cannot benefit from these functions, for example, because you did not order the PAV optional function, you have another option to reduce contention to volumes and hence reduce IOS queue time.

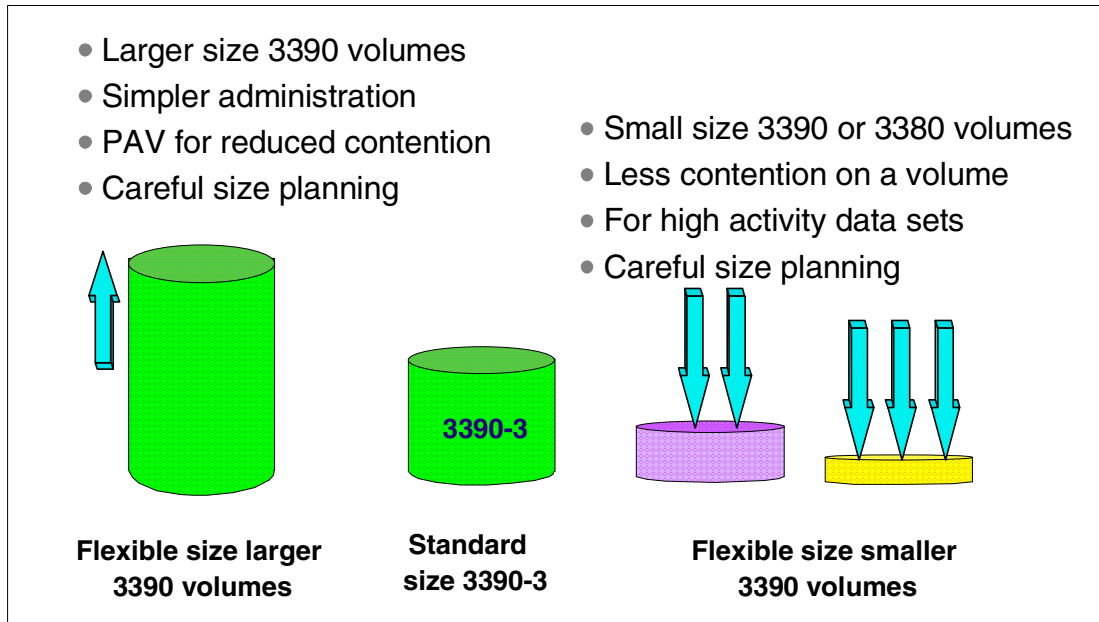


Figure 6-20 CKD custom volumes

When configuring your CKD volumes in the ESS, you have the option to define custom volumes. You can define logical 3390 or 3380 type volumes that do not have the standard number of cylinders of a Model 3 or Model 9, for example, but instead have a flexible number of cylinders that you can choose — in fact any number from 1 to 32,760 cylinders.

If you do not have PAV, then you probably want to define small volume sizes to reduce contention to the volume. You can spread high-activity data sets on separate smaller custom volumes, or even give each high-activity data set its own custom volume (plan carefully so that you do not exceed the 256 device addressing limit per LSS).

You should carefully plan the size of the custom volumes, and consider the potential growth of the data sets. You can adjust the size of each custom volume to the data set that you plan to put on this volume. But you might also come to the conclusion that you just need some standard small volume sizes of, perhaps 50, or 100, or 500 cylinders. You have the choice.

Large volume support

Alternatively you may want to define custom volumes larger than the standard size 3390 volumes. This may simplify the administration when having large databases. For this particular situation, PAV will allow you to avoid the I/O contention on the UCB queues.

6.6 FICON host adapters

FICON extends the IBM TotalStorage Enterprise Storage Server Model 800's ability to deliver bandwidth potential to the volumes needing it, when they need it.

6.6.1 FICON benefits

For the z/Architecture and s/390 servers, the FICON attachment provides many improvements:

- ▶ Increased number of concurrent I/O connections over the link. FICON provides channel-to-ESS multiple concurrent I/O connections. ESCON supports only one I/O connection at any one time (these are logical connections, not links).
- ▶ Increased distance. With FICON, the distance from the channel to the ESS, or channel to switch, or switch to ESS link is increased. The distance for ESCON of 3 km is increased to up to 10 km (20 km with RPQ) for FICON channels using long wavelength lasers with no repeaters.
- ▶ Increased link bandwidth. FICON has up to 10 times the link bandwidth of ESCON (1 Gbps full duplex, compared to 200 MBps half duplex). FICON has up to more than four times the effective channel bandwidth (70 MBps compared to 17 MBps for ESCON). The FICON adapter on the ESS will also transmit/receive at 2 Gb if the switch/director supports 2 Gb links.
- ▶ No data rate droop effect. For ESCON channels, the droop effect started at 9 km. For FICON, there is no droop effect even at a distance of 100 km.
- ▶ Increased channel device-address support, from 1,024 devices for an ESCON channel to up to 16,384 for a FICON channel.
- ▶ Greater exploitation of Parallel Access Volume (PAV). FICON allows for greater exploitation of PAV in that more I/O operations can be started for a group of channel paths.
- ▶ Greater exploitation of I/O Priority Queuing. FICON channels use frame and Information Unit (IU) multiplexing control to provide greater exploitation of the I/O Priority Queuing mechanisms within the FICON-capable ESS.
- ▶ Better utilization of the links. Frame multiplexing support on FICON channels, switches, and FICON control units such as the ESS provides better utilization of the links.

These improvements can be realized in more powerful and simpler configurations with increased throughput. The z/OS user will notice improvements over ESCON channels, with reduced bottlenecks from the I/O path, allowing the maximum control unit I/O concurrency exploitation:

- ▶ IOSQ time (UCB busy) can be reduced by configuring more alias device addresses, using Parallel Access Volumes (PAVs). This is possible because FICON channels can address up to 16,384 devices (ESCON channels address up to 1,024; the ESS has a maximum of 4096 devices).
- ▶ Pending time can also be reduced:
 - Channel busy conditions are reduced by FICON channel's multiple starts.
 - Port busy conditions are eliminated by FICON switches' frame multiplexing
 - Control unit busy conditions are reduced by FICON adapters' multiple starts
 - Device-busy conditions are also reduced, because FICON's multiple concurrent I/O operations capability can improve the Multiple Allegiance (MA) function exploitation

FICON channels allow a higher I/O throughput, using fewer resources. FICON's architecture capability to execute multiple concurrent I/O operations, which can be sustained by the FICON link bandwidth, allows for:

- ▶ A single FICON channel can have I/O operations to multiple logical control units at the same time, by using the FICON protocol frame multiplexing.

- ▶ FICON's CCW and data prefetching and pipelining, and protocol frame multiplexing also allows multiple I/O operations to the same logical control unit. As a result, multiple I/O operations can be done concurrently to any logical control unit. By using IBM ESS's Parallel Access Volumes (PAV) function, multiple I/O operations are possible even to the same volume.

6.6.2 FICON I/O operation

An I/O operation executed over a FICON channel has differences with an I/O operation executed over an ESCON channel. Both will be started by the same set of I/O routines of the operating system, but the mechanics of the operation differ whether it develops over a FICON channel or over an ESCON channel. In this section we go deeper into the I/O operation components, to understand how FICON improves these components to make the complete I/O more efficient.

Note: The explanations developed in this section are based on the view that z/OS has of the I/O operation sequence.

Let us first understand some of these basic components that are part of an I/O operation.

- ▶ I/O Supervisor Queue time (IOSQ), measured by the operating system

The application I/O request may be queued in the operating system if the I/O device, represented by the UCB, is already being used by another I/O request from the same operating system image (UCB busy). The I/O Supervisor (IOS) does not issue a Start Subchannel (SSCH) command to the channel subsystem until the current I/O operation to this device ends, thereby freeing the UCB for use by another I/O operation. The time while in this operating system queue is the *IOSQ time*.
- ▶ Pending time (PEND), measured by the channel subsystem
 - After IOS issues the Start Subchannel command, the channel subsystem may not be able to initiate the I/O operation if any path or device-busy condition is encountered:
 - Channel busy, with another I/O operation from another z/OS system image (from the same or from a different CEC).
 - Switch port busy, with another I/O operation from another or the same CEC. This can only occur on an ESCON channel. The use of buffer credits on a FICON native channel eliminates switch port busy.
 - Control unit adapter busy, with another I/O operation from another or the same CEC.
 - Device busy, with another I/O operation from another CEC.
- ▶ Connect time, measured by the channel subsystem

This is the time that the channel is connected to the control unit, transferring data for this I/O operation.
- ▶ Disconnect time

The channel is not being used for this I/O operation, since the control unit is disconnected from the channel, waiting for access to the data or to reconnect. Disconnect time often does not exist when a cache hit occurs because data is transferred directly from cache without the need for access to the device.

ESCON cache hit I/O operation

Figure 6-21 on page 187 shows the components of an ESCON I/O operation when a cache hit occurs. In this case, there is no disconnect time and the I/O operation ends at the end of the connect time, after transferring the requested data.

The ESCON I/O operation may accumulate IOSQ time if the device (UCB) is already in use for another I/O request from this system. The operating system IOS (I/O Supervisor) routines only initiate one I/O operation at a time for a device with the channel subsystem. The new I/O operation cannot be started with the channel subsystem until the I/O interrupt signalling completion of the current outstanding I/O operation has been processed by IOS.

Parallel Access Volume (PAV) reduces IOSQ time and device-busy conditions by allowing multiple I/O requests per UCB for access to the same logical volume.

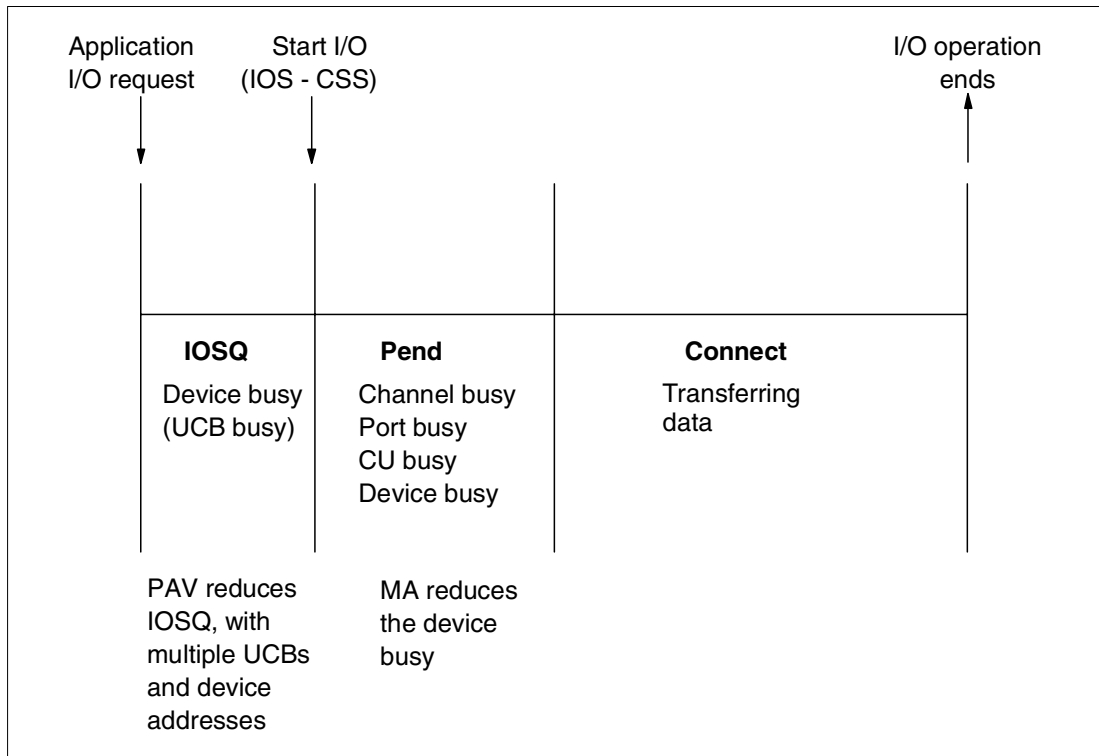


Figure 6-21 ESCON cache hit I/O operation sequence

Once the request has been accepted by the channel subsystem, it may accumulate PEND time if the channel subsystem is unable to start the request because the condition indication is either channel busy, port busy, control unit (CU) busy, or device busy.

With the ESS, some control unit busy conditions can be alleviated with I/O queuing by the control unit. In the case of a cache hit, the ESS may queue an I/O request for conditions that in other subsystems would result in CU busy, such as destaging, extent conflict resolution, and so on. This control unit I/O queuing time is accumulated in disconnect time, but reported later in pend time.

In the ESS, Multiple Allegiance alleviates some device-busy conditions, since this function enables different operating system images to perform concurrent I/O operations at the same logical volume as long as no extent conflict exists (note that a device-busy condition still occurs if the device is reserved by another operating system image).

When the I/O operation is accepted by the control unit, connect time is accumulated as the channel transfers data to/from cache.

The I/O operation completes when the data transfer into cache is complete. No access to the physical volume is required before the end of the I/O operation is signalled in the case of a cache hit.

ESCON cache miss I/O operation

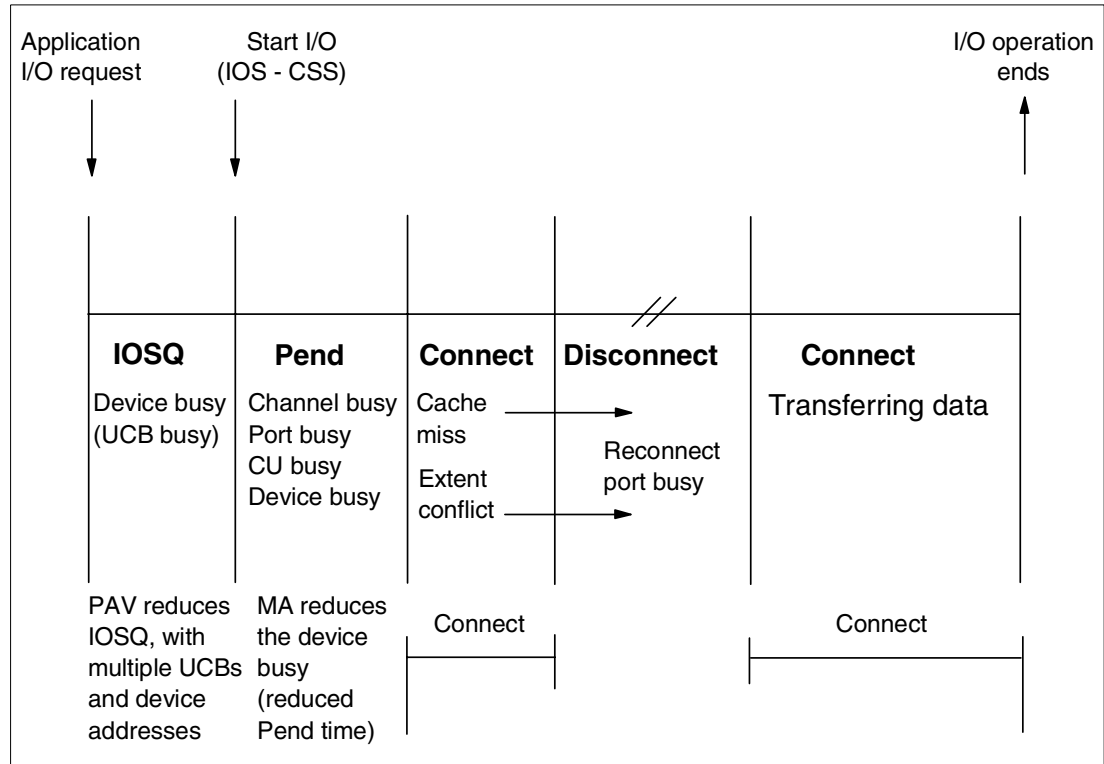


Figure 6-22 ESCON cache miss I/O operation sequence

Figure 6-22 shows an ESCON I/O operation sequence when a cache miss occurs. In this case, connect time is accumulated as the positioning CCWs are transferred to the ESS. For the ESS, this connect time component also includes the extent conflict checking time.

Disconnect time is then accumulated as the physical disk drive positioning operations are performed. Then the control unit must reconnect to the channel for the transfer of data. It is during the attempt to reconnect that a port busy condition may occur.

Further connect time is accumulated as data is transferred over the channel. For ESCON I/O operations, the total connect time component is predictable, since the data transfer is directly related to the speed of the channel and the number of bytes transferred.

The I/O operation ends after the requested data has been transferred and the terminating interrupt has been presented by the channel.

FICON cache hit I/O operation

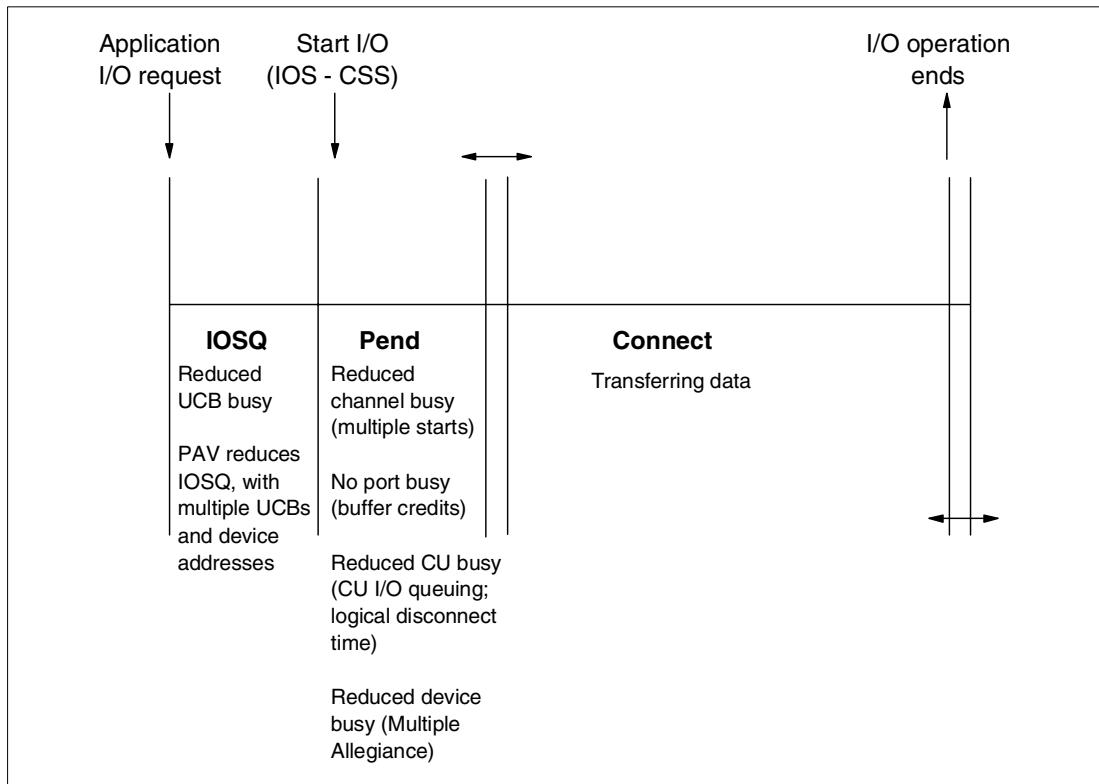


Figure 6-23 FICON cache hit I/O operation sequence

Figure 6-23 shows a FICON I/O operation sequence when a cache hit occurs. When using FICON, some busy conditions will be reduced or eliminated:

- ▶ More devices (and consequently, more UCBs) can be configured (up to 16,384 devices per FICON channel), allowing a higher PAV exploitation. This reduces the number of device-busy and UCB-busy conditions that are accumulated in the IOSQ time parameter of the z/OS operating systems.
- ▶ Channel busy is reduced with FICON's capability of multiple starts to the same channel path, thereby reducing pend time conditions.
- ▶ Port busy does not exist on FICON switches. The FICON switch uses switch port buffer credits.
- ▶ Control unit busy is reduced with CU I/O queuing in the ESS, also reducing the pend time.
- ▶ Device-busy conditions are reduced by further exploitation of the ESS Multiple Allegiance (MA) function, due to FICON's multiple concurrent I/O operations capability.

As Fibre Channel frame multiplexing is used by FICON links, some connect time is less predictable for individual I/Os. Figure 6-23 shows this effect when showing the start and end points of the connect component.

FICON cache miss I/O operation

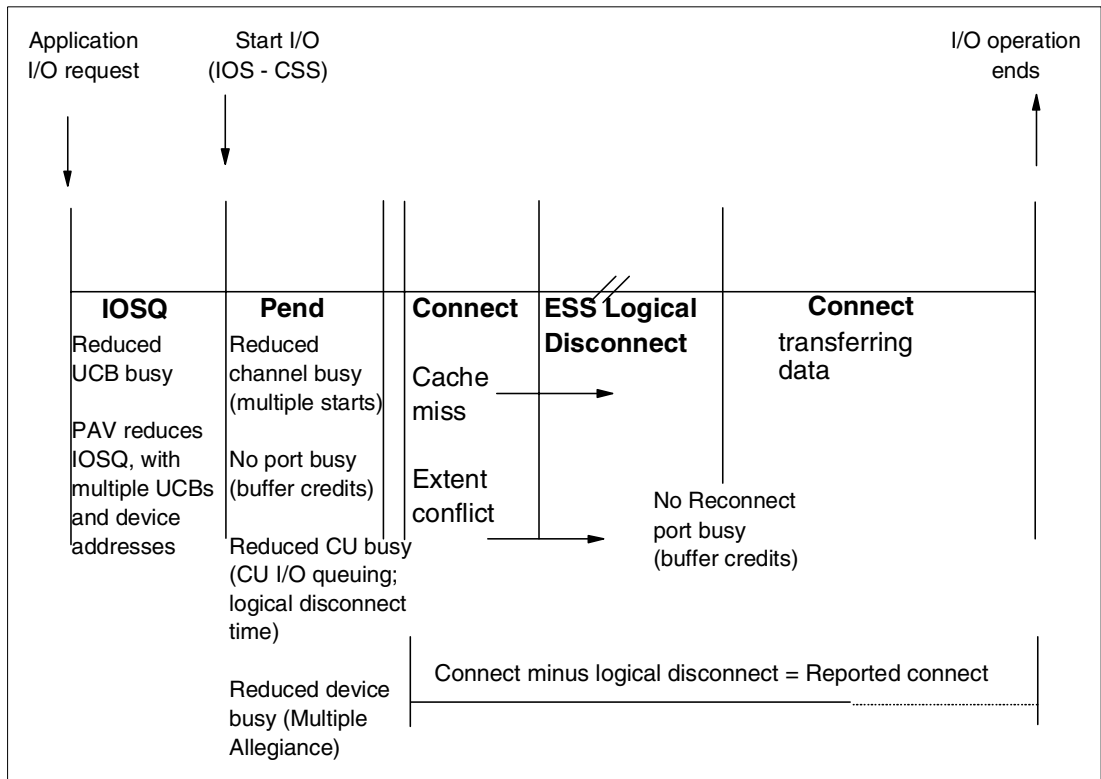


Figure 6-24 FICON cache miss I/O operation sequence

Figure 6-24 shows a FICON I/O operation sequence when a cache miss occurs.

Having all the benefits about reducing busy conditions and times as shown in the previous cache example, a new condition takes place in this kind of I/O operation, removing another busy condition.

In the ESCON cache miss operation, a disconnect time component is expected. Because the ESS FICON adapter can handle multiple concurrent I/O operations at a time, it will not disconnect from the channel when a cache miss occurs for a single I/O operation. So the ESS adapter remains connected to the channel, being able to transfer data from other I/O operations. This condition is called “logical disconnect”.

With no disconnect time, the port-busy condition during ESCON channel reconnect time does not exist also, and this is another improvement over ESCON channels. Note that the channel subsystem reported connect times are not affected by logical disconnect times. The logical disconnect time is accumulated by the ESS as a component of connect time, but the connect time reported by the channel is calculated by excluding the logical disconnect time. The logical disconnect time is reported as part of the pend time.

6.6.3 FICON benefits at your installation

The most obvious benefits of FICON connectivity are the increased per channel bandwidth and greater simplicity of channel fabric. This speed allows an approximately 4:1 or more reduction in the number of channels when converting from ESCON channels to FICON channels.

Greater simplicity of configuration is also achieved because the FICON architecture allows more devices per channel and an increased bandwidth.

Single-stream sequential operations benefit from improvements in throughput, so that elapsed time for key batch, data mining, or dump operations dramatically improves. This provides relief to those installations where the batch or file maintenance windows are constrained today.

Response-time improvements may accrue for some customers, particularly for data stored using larger block sizes. The data transfer portion of the response time is greatly reduced because the data rate during transfer is six times faster than ESCON. This improvement leads to significant connect time reduction. The larger the transfer, the greater the reduction as a percentage of the total I/O service time.

Pend time caused by director port busy is totally eliminated because there are no more collisions in the director with FICON architecture.

PAV and FICON work together to allow multiple data transfers to the same volume at the same time over the same channel, providing greater parallelism and greater bandwidth while simplifying the configurations and the storage management activities.

6.7 Host adapters configuration

For the zSeries environments, there are certain specific recommendations in order to fully take advantage of the ESS performance capacity, or you will be using just part of the total ESS performance capability.

When configuring for ESCON, consider these recommendations:

- ▶ Configure at least 16 ESCON channels to the ESS
- ▶ Use 8-path groups
- ▶ Plug channels for an 8-path group into four HAs (that is, use one HA per bay)
- ▶ Each 8-path group should access its LCU on one cluster
- ▶ One 8-path group is better than two 4-path groups

This way, both the channels connected to the same HA will serve only even or only odd LCUs, which is the best, and access will be distributed over the four HA bays.

When configuring for FICON, consider the following recommendations:

- ▶ Define a minimum of four channel paths per CU. Fewer channel paths will not allow exploitation of full ESS bandwidth. A more typical configuration would have eight FICON channels.
- ▶ Spread FICON host adapters across all adapter bays. This should result in minimally one host adapter per bay, or in a typically configured ESS, two host adapters per bay.
- ▶ Define a minimum of four FICON channels per path group.

See 6.6, “FICON host adapters” on page 184 and 4.33, “FICON host connectivity” on page 140 for more details on ESS FICON attachment characteristics.



Copy functions

The IBM TotalStorage Enterprise Storage Server Model 800 has a rich set of copy functions suited for the different data protection and recovery implementations that enterprises need to set up for their business continuance solutions, as well as for data migration and offsite backups.

This chapter gives an introductory overview of the copy functions available with the ESS Model 800. For more detailed information on the topics presented in this chapter, refer to the following redbooks:

- ▶ *Implementing ESS Copy Services on UNIX and Windows NT/2000*, SG24-5757
- ▶ *Implementing ESS Copy Services on S/390*, SG24-5680
- ▶ *IBM TotalStorage Enterprise Storage Server PPRC Extended Distance*, SG24-6568

7.1 ESS Copy Services functions

The ESS Model 800 provides a powerful set of copy functions for the diverse servers it attaches. This ample set of copy functions is introduced in this chapter. For the open systems servers and for the zSeries servers, the ESS provides the following copy functions (refer to Figure 7-1 and Figure 7-2 on page 195):

- ▶ FlashCopy
- ▶ Peer-to-Peer Remote Copy (PPRC)
- ▶ Peer-to-Peer Remote Copy Extended Distance (PPRC XD)

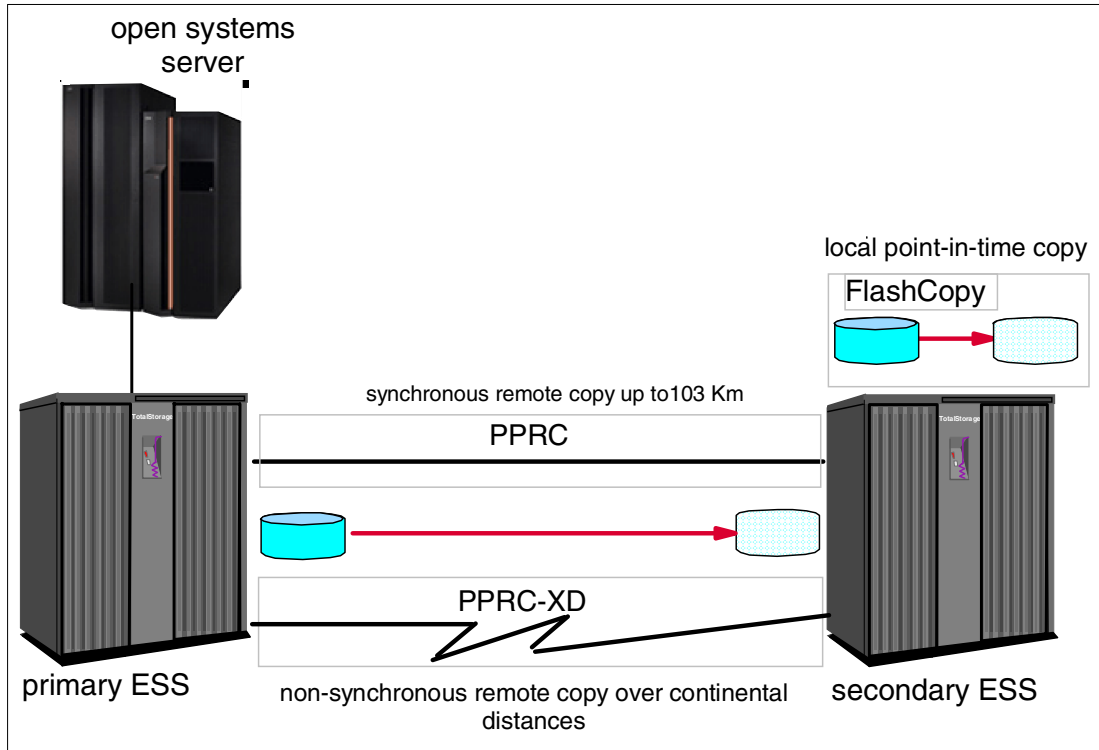


Figure 7-1 ESS Copy Services for open systems

Additionally for the zSeries servers, the ESS provides the following copy functions (refer to Figure 7-2 on page 195):

- ▶ Extended Remote Copy (XRC)
- ▶ Concurrent Copy

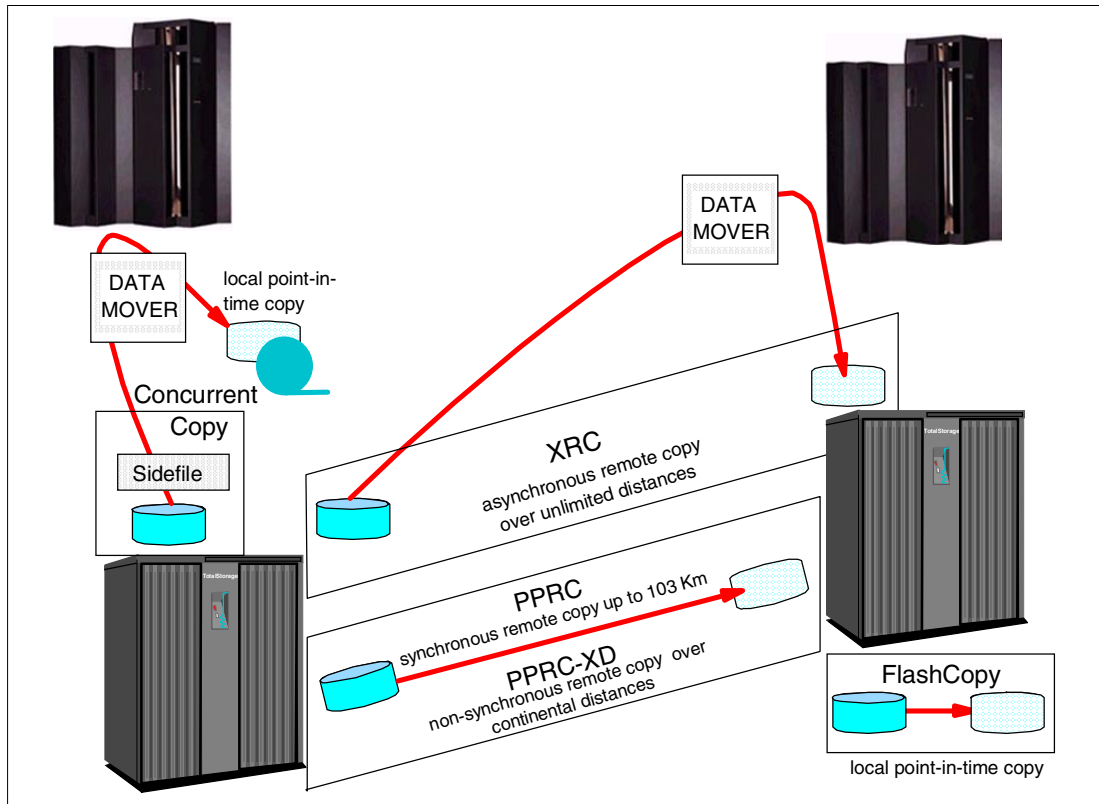


Figure 7-2 ESS Copy Services for zSeries

7.2 Managing ESS Copy Services

There are several ways of invoking and managing the various ESS Copy Services functions, for the different operating environments. For PPRC and FlashCopy, the most commonly used are:

- ▶ ESS Copy Services Web user interface (WUI)
- ▶ ESS Copy Services command-line interface (CLI)
- ▶ TSO Commands (only for z/OS and OS/390)

This section describes these interfaces for PPRC and FlashCopy management. The WUI and the CLI can be used with open systems servers as well as with the zSeries servers. In addition to them, for the z/OS and OS/390 systems, the TSO commands can also be used for invoking and managing PPRC and FlashCopy (and XRC).

The other ways of invoking and managing the ESS Copy Services that are available for the zSeries servers are the following:

- ▶ The ANTRQST macro provided with the DFSMSdftp component, for PPRC, XRC, and FlashCopy invocation and management under z/OS and OS/390
- ▶ The DFSMSdss utility with its ADRSSU program, for FlashCopy and Concurrent Copy invocation and management under z/OS and OS/390
- ▶ The ICKDSF utility for PPRC (excluding PPRC-XD) for the MVS, VM, VSE, and stand-alone operating environments
- ▶ The TSO commands for XRC invocation and management under z/OS and OS/390

- ▶ Native CP commands for FlashCopy invocation under z/VM

7.2.1 ESS Copy Services Web user interface

The ESS provides a Web user interface to invoke and manage ESS Copy Services. With the ESS Copy Services Web user interface, users of open system servers and zSeries servers can invoke and control PPRC and FlashCopy. This interface requires one of the following browsers:

- ▶ Netscape Communicator 4.6 (or above)
- ▶ Microsoft Internet Explorer (MSIE) 5.1 (or above)

Note: the IBM SSR has to enable the ESS Copy Services Web interface for CKD volumes.

The Web browser interface allows you to easily control the ESS copy functionality from the network from any platform for which the browser is supported via a graphical user interface (GUI). The ESS Master Console (feature 2717 of the ESS) comes with a browser (refer to 4.14.1, “ESSNet and ESS Master Console” on page 108 for more information).

To invoke the ESS copy services using the Web user interface, click the **Copy Services** button on the Welcome window of the IBM TotalStorage Enterprise Storage Server Specialist (see Figure 7-3).



Figure 7-3 ESS Specialist Welcome window

Clicking the **Copy Services** button on the left side of the IBM TotalStorage Enterprise Storage Server Specialist Welcome window connects you to the ESS Copy Services Welcome window (see Figure 7-4 on page 197).



Figure 7-4 ESS Copy Services Welcome window

This is the starting point for managing the copy functions of the ESS by means of the ESS Copy Services Web user interface (WUI). On the left side (refer to Figure 7-4) the following options can be selected:

► **Volumes**

- Get information and status about volumes defined in a logical storage subsystem (LSS) of the ESS
- Select source and target volume for a PPRC or FlashCopy task
- Filter the output of the volume display to a selected range
- Search for a specific volume based on its unique volume ID
- Establish, terminate and suspend PPRC copy pairs and optionally save the task
- Establish and withdraw FlashCopy pairs and optionally save the task
- Enter the multiple selection mode for PPRC and FlashCopy

► **Logical Subsystems**

- View all ESSs within the storage network
- View all logical subsystems (LSSs) within the storage network
- Get information about a logical subsystem and its status
- View and modify the copy properties of a logical subsystem
- Filter the output of a selected range
- Search for a specific logical subsystem based on its unique address
- Establish, terminate, and suspend PPRC copy pairs, and optionally save the task

► **Paths**

- Establish PPRC paths
- Remove PPRC paths
- View information about PPRC paths

► **Tasks**

By using the Volumes, Logical Subsystems, and Paths windows of the ESS Copy Services, you can create, save, and run one or more tasks, which perform particular

functions when you choose to run them. When you run a task, the ESS Copy Services server tests whether there are existing conflicts. If conflicts exist, then the server ends this task. The Task window is used for organizing a series of related tasks into a group, and this group should be named. Then because the group has a name, the entire series of related tasks (group) can be run by selecting the group and running it. When a group task is run, all the tasks within the group are submitted to the ESS Copy Services server in parallel.

For a complete description of the specific windows and the use of the ESS Copy Services Web user interface, refer to the publication *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448. To find this publication go to:

<http://ssddom02.storage.ibm.com/techsup/webnav.nsf/support/2105>

and click the **Documentation** link.

7.2.2 ESS Copy Services command-line interface (CLI)

Using the ESS Copy Services command-line interface (CLI), users of selected open systems servers are able to communicate with the ESS Copy Services server from the host's command line to manage PPRC and FlashCopy. An example would be to automate tasks, such as FlashCopy, by invoking the ESS Copy Services commands within customized scripts.

The ESS Copy Services command-line interface (CLI) is available for selected open servers. To see the list of supported hosts, refer to the corresponding server page documentation in the interoperability matrix found at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

The ESS Copy Services command-line interface is a Java-based application that runs on the host server. The host system does not necessarily need to be connected to storage assigned to one of the host adapter ports of the ESS. The only requirement is that the host from where you want to invoke the commands is connected via the network to the ESS that is defined as the primary ESS Copy Services server.

Detailed and complete information on the ESS Copy Services command-line interface can be found in the publication *IBM TotalStorage Enterprise Storage Server Copy Services Command-Line Interface Reference*, SC26-7449. This publication can be found at:

<http://ssddom02.storage.ibm.com/disk/ess/documentation.html>

ESS Copy Services commands

For both the UNIX and Windows based operating systems, the same ESS Copy Services commands are available. The command set for UNIX operating systems consists of shell scripts with the *.sh ending; the commands for the Windows operating systems are batch files with the *.bat ending. Functionally they are identical, but there may be some differences in specified parameters when invoking the commands.

► **rsExecuteTask** (.sh .bat)

Accepts and executes one or more predefined ESS Copy Services tasks. Waits for these tasks to complete execution.

► **rsList2105s** (.sh .bat)

Displays the mapping of host physical volume name to 2105 (ESS) volume serial number.

► **rsPrimeServer** (.sh .bat)

Notifies the ESS Copy Services server of the mapping of host disk name to 2105 (ESS) volume serial number. This command is useful when the ESS Copy Services Web windows are used to perform FlashCopy and/or PPRC functions. Collects the mapping of the host physical volume names to the ESS Copy Services server. This permits a host volume view from the ESS Copy Services Web window.

► **rsQuery** (.sh .bat)

Queries the FlashCopy and PPRC status of one or more volumes.

► **rsQueryComplete** (.sh .bat)

Accepts a predefined ESS Copy Services server task name and determines whether all volumes defined in that task have completed their PPRC copy initialization. If not, this command waits for that initialization to complete.

► **rsTestConnection** (.sh .bat)

Determines whether the ESS Copy Services server can successfully be connected.

Scripting the ESS Copy Services CLI

You can enhance the functionality of the ESS Copy Services command-line interface by incorporating its use in your own customized scripts. Common applications of the CLI might include batch, automation, and custom utilities.

7.2.3 TSO commands

The z/OS and OS/390 TSO/E commands can be used to control the PPRC, XRC and FlashCopy functions on the z/OS and OS/390 systems. These commands are extremely powerful, so it is important that they are used correctly and directed to the correct devices. It is recommended that the access to these commands be restricted by placing them in a RACF-protected library for use by the authorized storage administrators only.

Detailed and complete information on the TSO commands for the ESS Copy Services, PPRC, XRC, and FlashCopy can be found in the publication *z/OS DFSMS Advanced Copy Services*, SC35-0428. This publication can be located at:

<http://www.storage.ibm.com/software/sms/sdm/sdmtech.html>

The TSO commands to control FlashCopy and PPRC operations are described in 7.4.3, “FlashCopy management on the zSeries” on page 202 and 7.5.4, “PPRC management on the zSeries” on page 207 respectively. The TSO commands used to control XRC sessions are not listed in this chapter, but can be found in *z/OS DFSMS Advanced Copy Services*, SC35-0428.

7.3 ESS Copy Services setup

ESS Copy Services provides a Web-based interface for establishing and managing PPRC and FlashCopy. To use ESS Copy Services, one cluster of any ESS is defined as the primary ESS Copy Services Server (CSS). A second ESS cluster should be defined as the backup CSS. In case the primary CSS fails, then ESS Copy Services can be managed via the backup. Connection between primary and backup CSS is made by an Ethernet connection (see Figure 7-5 on page 200).

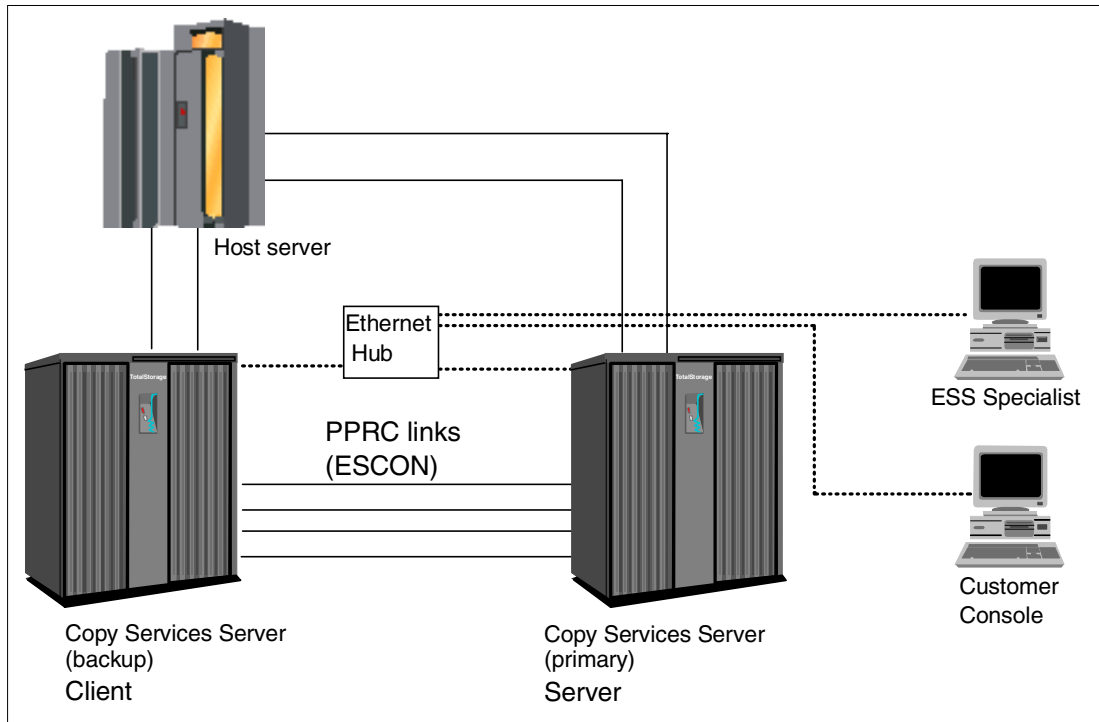


Figure 7-5 ESS Copy Services setup

Setting up the CSS in the ESS is done by the IBM Service Support Representative. You must provide the Communication Resources Worksheet, which must be filled in during installation planning according to *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444.

Additional information on ESS Copy Services setup and implementation can be found in the publication *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448.

7.4 FlashCopy

Today, more than ever, organizations require their applications to be available 24 hours per day, seven days per week (24x7). They require high availability, minimal application downtime for maintenance, and the ability to perform data backups with the shortest possible application outage.

FlashCopy provides an immediate point-in-time copy of data. FlashCopy creates a physical point-in-time copy of data and makes it possible to access both the source and target copies immediately. FlashCopy is an optional feature on the ESS Model 800 (refer to Appendix A, "Feature codes" on page 263).

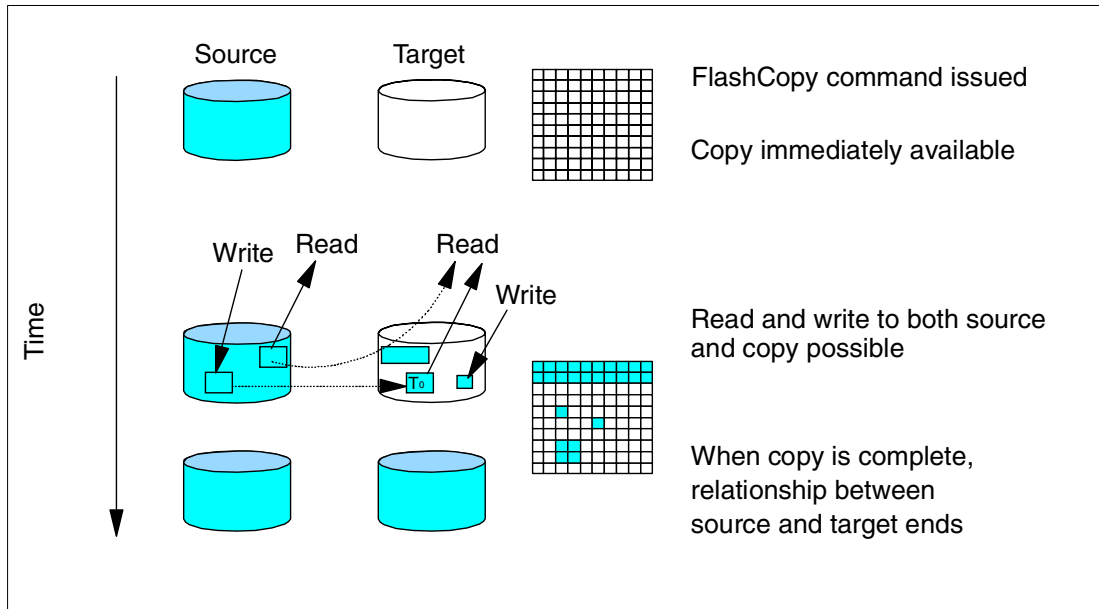


Figure 7-6 FlashCopy point-in-time copy

7.4.1 Overview

FlashCopy provides a point-in-time copy of an ESS logical volume. The point-in-time copy function gives you an immediate copy, or “view”, of what the original data looked like at a specific point in time. This is known as the *T0 (time-zero)* copy (see Figure 7-6).

When FlashCopy is invoked, the command returns to the operating system as soon as the FlashCopy pair has been established and the necessary control bitmaps have been created. This process takes only a short moment to complete. Thereafter, you have access to a *T0* copy of the source volume. As soon as the pair has been established, you can read and write to both the source and the target logical volumes.

The point-in-time copy created by FlashCopy is typically used where you need a copy of production data to be produced with minimal application downtime. It can be used for online backup, testing of new applications, or for creating a database for data-mining purposes. The copy looks exactly like the original source volume and is an immediately available binary copy. See Figure 7-6 for an illustration of FlashCopy concepts.

FlashCopy is possible only between logical volumes in the same logical subsystem (LSS). The source and target volumes can be on the same or on different arrays, but only if they are part of the same LSS.

A source volume and the target can be involved in only one FlashCopy relationship at a time. When you set up the copy, a relationship is established between the source and the target volume and a bitmap of the source volume is created. Once this relationship is established and the bitmap created, the target volume copy can be accessed as though all the data had been physically copied. While a relationship between source and target volumes exists, a background task copies the tracks from the source to the target. The relationship ends when the physical background copy task has completed.

You can suppress the background copy task using the **Do not perform background copy (NOCOPY)** option. This may be useful if you need the copy only for a short time, such as making a backup to tape. If you start a FlashCopy with the **Do not perform background copy** option, you must withdraw the pair (a function you can select) to end the relationship

between source and target. Note that you still need a target volume of at least the same size as the source volume.

You cannot create a FlashCopy on one type of operating system and make it available to a different type of operating system. You can make the target available to another host running the same type of operating system.

7.4.2 Consistency

At the time when FlashCopy is started, the target volume hasn't had any data copied yet. If `BACKGROUND(COPY)` is specified, the background copy task copies updated data from the source to the target. The FlashCopy bitmap keeps track of which data has been copied from source to target. If an application wants to read some data from the target that has not yet been copied to the target, the data is read from the source; otherwise, the read is satisfied from the target volume (See Figure 7-6 on page 201). When the bitmap is updated for a particular piece of data, it means that source data has been copied to the target and updated on the source. Further updates to the same area are ignored by FlashCopy. This is the essence of the T0 point-in-time copy mechanism.

When an application updates a track on the source that has not yet been copied, the track is copied to the target volume (See Figure 7-6 on page 201). Reads that are subsequently directed to this track on the target volume are now satisfied from the target volume instead of the source volume. After some time, all tracks will have been copied to the target volume, and the FlashCopy relationship will end.

Please note FlashCopy can operate at extent level in CKD environments.

Applications do not have to be stopped for FlashCopy. However, you have to manage the data consistency between different volumes. One example of things to consider is that only data on the physical disk is copied, while data in the buffers in the application server won't be copied. So either applications have to be frozen consistently, or you have to invoke built functions that maintain that consistency.

7.4.3 FlashCopy management on the zSeries

With the z/OS and OS/390 operating systems, FlashCopy can be invoked and managed by four different methods:

- ▶ Using the `DFSMSdss` utility, which invokes the FlashCopy function via the `ADRDSSU` program, when the `COPY FULL` command is used and the volumes are in the same LSS.
- ▶ TSO/E commands that are unique to FlashCopy. The three commands are:
 - FCESTABL** Used to establish a FlashCopy relationship
 - FCQUERY** Used to query the status of a device
 - FCWITHDR** Used to withdraw a FlashCopy relationship
- ▶ Using the `DFSMSdfp` Advanced Services `ANTRQST` macro, which interfaces with the System Data Mover API.
- ▶ ESS Copy Services Web user interface (WUI). The ESS provides a Web browser interface that can be used to control the FlashCopy functions. To use this interface, it must first be enabled for the CKD volumes by the IBM SSR. See "ESS Copy Services Web user interface" on page 196.

In VM environments, FlashCopy can be managed by using the ESS Copy Services Web user interface (WUI). FlashCopy is also supported for guest use (`DFSMSdss` utility of z/OS) for dedicated (attached) volumes or for full-pack minidisks. z/VM also supports a native CP

user's ability to initiate a FlashCopy function. The syntax and detailed explanation of the CP FlashCopy command can be found in the publication *z/VM CP Command and Utility Reference*, SC24-6008.

VSE/ESA provides support for the FlashCopy function by means of the IXFP SNAP command. Also the ESS Copy Services Web user interface (WUI) can be used; it previously has to be enabled by the IBM SSR for use with the CKD volumes.

You may refer to the redbook *Implementing ESS Copy Services on S/390*, SG24-5680 for additional information on invocation and management of FlashCopy.

7.4.4 FlashCopy management on the open systems

FlashCopy can be managed in the open systems environments by means of the ESS Copy Services Web user interface (WUI) and by means of the ESS Copy Services command-line interface (CLI).

When using the Web browser interface that the ESS Copy Services provides, the FlashCopy pairs will be established using the Volumes window, or using the Task window that the Web browser presents (see "ESS Copy Services Web user interface" on page 196).

The Java-based ESS Copy Services command-line interface (CLI) allows administrators to execute Java-based copy services commands from a command line. An example would be to automate tasks, such as FlashCopy, by invoking the ESS Copy Services commands within customized scripts (refer to 7.2.2, "ESS Copy Services command-line interface (CLI)" on page 198 for additional information).

7.5 Peer-to-Peer Remote Copy (PPRC)

Peer-to-Peer Remote Copy (PPRC) is a proven synchronous remote data mirroring technology utilized for many years. It is used primarily as part of the business continuance solution for protecting the organization's data against disk subsystem loss or, in the worst case, complete site failure. But it is also used for remote migration of data and of application workloads, and offsite backups.

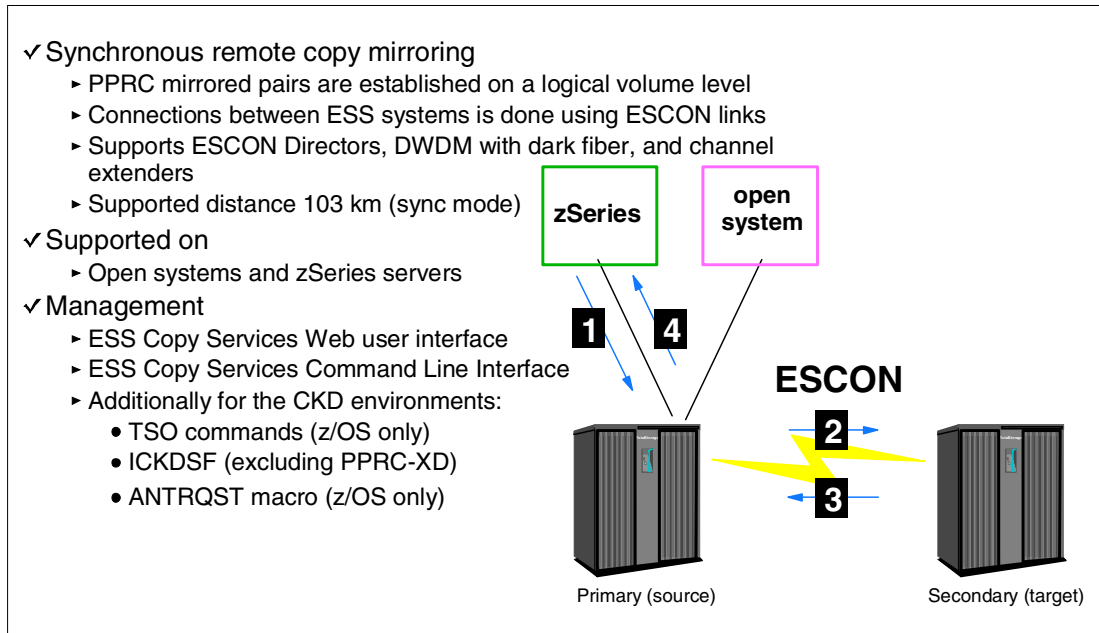


Figure 7-7 Synchronous volume copy PPRC

7.5.1 PPRC overview

PPRC is a real-time remote copy technique that synchronously mirrors a *primary* set of volumes (that are being updated by applications) onto a *secondary* set of volumes (see Figure 7-7). Typically the secondary volumes will be on a different ESS located at a remote location (the recovery site) some distance away from the application site. Mirroring is done at a logical volume level.

PPRC is a hardware solution, thus it is application independent and needs to be available on both the local and the remote ESS. Because the copy function occurs at the storage subsystem level, the application does not need to know of its existence.

PPRC guarantees that the secondary copy is up-to-date by ensuring that the primary volume update will be successfully completed only when the primary ESS receives acknowledgment that the secondary copy has been successfully updated.

The sequence when updating records is (refer to Figure 7-7):

1. Write to primary volume (to primary ESS cache and NVS). The application writes data to a primary volume on an ESS at the application site, and cache hit occurs.
2. Write to secondary (to secondary ESS cache and NVS). The application site ESS then initiates an I/O channel program to the recovery site ESS, to transfer the updated data to the recovery site cache and NVS.
3. Signal write complete on the secondary. The recovery site ESS signals write complete to the application site ESS when the updated data is in its cache and NVS.
4. Post I/O complete. When the application site ESS (primary ESS) receives the write complete from the recovery site ESS, it returns I/O complete status to the application system.

Destage from cache to the back-end disk drives on both the application ESS and the recovery site ESS is performed asynchronously. The synchronous technique of PPRC ensures that

application-dependent writes will be applied in the same sequence upon the secondary volumes, thus providing application consistency at every moment.

When in the process an error occurs that prevents the secondary copy from being updated, PPRC automatically suspends the mirroring function and applies the critical attribute behavior. Also, if the logical paths have been defined with the consistency group option enabled, then a long busy condition may be instigated that will allow for automation routines to take appropriate actions. CRIT and consistency grouping are described later in “CRIT attribute and consistency groups” on page 206.

7.5.2 PPRC volume states

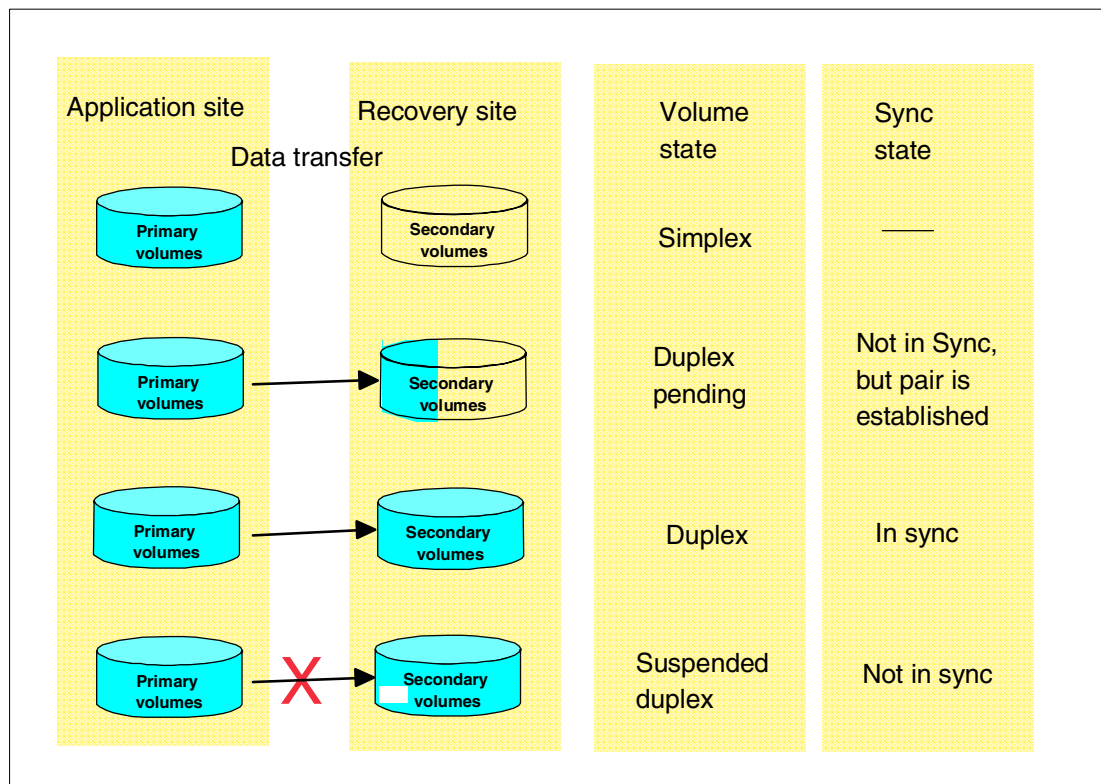


Figure 7-8 PPRC volume states - synchronous mode

From a PPRC perspective, volumes within the ESS can be in one of the following states (refer to Figure 7-8):

► **Simplex**

The simplex state is the initial state of the volumes before they are used in any PPRC relationship, or after the PPRC relationship has been withdrawn. Both volumes are accessible only when in simplex state; otherwise, just the primary is accessible.

► **Duplex pending**

Volumes are in duplex pending state after the PPRC copy relationship has been initiated, but the source and target volume are still out of synchronization (sync). In that case, data still needs to be copied from the source to the target volume of a PPRC pair. This situation occurs either after an initial PPRC relationship is just established, or after a suspended PPRC volume pair is just re-established. While in the duplex pending state, tracks are being copied in a very throughput-efficient process, from the primary volume to the

secondary volume. The PPRC secondary volume is not accessible when the pair is in duplex pending state. This state is a transitional state, until the duplex state is reached.

► **Duplex**

This is the state of a volume pair that is in sync (full synchronous); that is, both source and target volumes contain exactly the same data. Sometimes this state is also referred to as the Full Copy mode. The PPRC secondary volume is not accessible when the pair is in duplex state.

► **Suspended**

In this state of the PPRC pair, the writes to the primary volume are not mirrored onto the secondary volume. The secondary volume becomes out of synchronization. During this time PPRC keeps a bitmap record of the changed tracks in the primary volume. Later, when the volumes are re-synchronized, only the tracks that were updated will be copied.

A PPRC volume pair will go to suspended state, for instance, when the primary ESS cannot complete a write operation to the recovery site ESS. Also the operator can suspend pairs by command.

► **Duplex pending-XD**

This state is found when a volume pair is established in a nonsynchronous PPRC-XD relationship. This PPRC volume state is explained in 7.7.1, “PPRC-XD operation” on page 213.

FlashCopy of PPRC secondary

A PPRC secondary volume can be a FlashCopy primary volume. This allows you, for example, to produce a consistent point-in-time tertiary copy of the mirror volume while it is suspended.

CRIT attribute and consistency groups

The critical attribute of a pair is set up when the pair is established or resynchronized. This parameter defines the behavior of PPRC in case of an error. There are two alternatives for PPRC when an error occurs:

- Either suspend the pair and do not accept any further writes on the primary address (CRIT=YES)
- Or suspend the pair and accept further writes to the primary (CRIT=NO), even when the updates cannot be delivered to the secondary

There are two global strategies for the process triggered by CRIT=YES, which can be parameterized on an ESS:

- Either CRIT=YES - Paths (Light version), which suspends the pair, and does not accept any further writes to primary if the logical control units, primary and secondary, can no longer communicate. Otherwise, it has a behavior similar to CRIT= NO. The assumption is that the problem is at the device level and not a disaster that has affected the whole storage subsystem.
- Or CRIT=YES - All (Heavy version), which suspends the pair and does not accept any further writes to the primary volume if data cannot be sent to the secondary volume.

The CRIT = YES - All alternative assures that primary and the secondary will always have the identical data, but at the potential cost of a production outage. However, this does not prevent a database management system, for example DB2, to flag into its log files (which are mirrored) the need to do a recovery on a table space that has met a primary physical I/O error when the secondary is still valid. After a rolling disaster, this mechanism can make the

recovery task very difficult when you have many table spaces with many indexes. This is what is called the *dying scenario*.

When a logical path is defined between a primary and a secondary LSS, the consistency group attribute can be enabled for that path. Consistency grouping provides the ability to temporarily queue the write operations — to all PPRC volumes — on a single LSS pairing when an error occurs. If a volume pair error occurs that prevents PPRC from updating the secondary volume, the pair is suspended, and the LSS will enter the extended long busy (ELB) condition when a new write operation is tried. During this short ELB interval, the write operations to the primary LSS volumes are queued. At the same time, notification alerts are sent to the host system. Using automation triggered from the alert messages, during this ELB interval a set of related LSSs can be frozen, so consistency can be maintained across the volumes that make up the application set of volumes.

More information on the CRIT attribute and on consistency groups, as well as how they work together, can be found in the redbooks *Implementing ESS Copy Services on S/390*, SG24-5680 and *Implementing ESS Copy Services on UNIX and Windows NT/2000*, SG24-5757.

7.5.3 PPRC with static volumes

A *static volume* is a volume that is not receiving any write updates. Synchronous PPRC can be used over long distances (for example, to do remote data migration) if used with static primary volumes.

This particular implementation allows the use of synchronous PPRC over long distances, beyond the supported 103 km, while taking advantage of the excellent throughput of PPRC when it does the initial copy or the re-synchronization of volume pairs (duplex pending state).

This powerful way of PPRC data copying, combined with the latest microcode improvements in the supported channel extenders, makes the static volumes implementation ideal for data migration, data copying, or remote backup over long distances.

Additional information on this particular implementation can be found in the redbook *IBM TotalStorage Enterprise Storage Server PPRC Extended Distance*, SG24-6568.

7.5.4 PPRC management on the zSeries

In z/OS and OS/390 systems, PPRC can be managed using TSO commands as well as the ESS Copy Services Web user interface (WUI). Also, the ICKDSF utility (except for PPRC-XD) and the ANTRQST macro can be used for PPRC requests in the z/OS and OS/390 systems.

For z/VM and VSE/ESA systems, the management of PPRC is done by means of the ESS Copy Services Web user interface (WUI) and the ICKDSF utility. ICKDSF is also necessary for stand-alone situations. ICKDSF cannot be used to manage the PPRC-XD (nonsynchronous PPRC Extended Distance).

For all zSeries operating systems (z/OS, z/VM and VSE/ESA and TPF), the ESS Copy Services Web user interface (WUI) can be used to control the PPRC operation. The IBM SSR has to previously enable the ESS Copy Services Web user interface for CKD volumes.

TSO commands

Specific TSO/E commands can be used in z/OS and OS/390 to control PPRC:

- ▶ CESTPATH — used to establish logical paths (over the ESCON links) between a primary site (source) logical subsystem (LSS) and a recovery site (target) LSS. Each CESTPATH

command can establish up to eight paths from one primary site LSS to a single recovery site LSS. Each primary LSS can connect to up to four secondary LSSs. Each secondary LSS can connect to any number of primary LSSs as long as there are enough ESCON connections.

- ▶ **CESTPAIR** — used to specify PPRC primary and secondary volumes; thus a PPRC relationship is established. The primary and secondary volumes must have the same number of tracks on each cylinder and the same number of bytes on each track. The secondary device must have the same number or a greater number of cylinders as compared to the primary device.

This command has parameters that allow to indicate whether the operation is an initial establish of volumes that were in simplex state, or if it is a re-synchronization of a suspended pair of volumes.

- ▶ **CSUSPEND** — Used to suspend PPRC operations between a volume pair. PPRC stops transferring data to the secondary volume and keeps a bitmap to track the updates done on the primary volume (this bitmap will be used later when the pair is re-established, to copy just the updated tracks).
- ▶ **CQUERY** — Used to query the status of one volume of a PPRC pair, or the status of all paths associated with an LSS.
- ▶ **CGROUP** — Used to control operations for all PPRC volume pairs on a single primary and secondary LSS pairing.
- ▶ **CRECOVER** — Used to allow the recovery system to gain control of a logical volume on its ESS. This command is issued from the recovery system.
- ▶ **CDELPAIR** — Used to specify the primary and secondary volumes to remove from PPRC pairing.
- ▶ **CDELPATH** — Used to delete all established ESCON paths between a primary site (source) LSS and a recovery site (target) LSS. Only active paths to the specified recovery site LSS are affected; all other paths to other LSSs are unaffected.

ANTRQST macro

For the OS/390 and z/OS systems, the DFSMSdftp Advanced Services ANTRQST macro that calls the System Data Mover API can be used. The new ANTRQST macro supports the PPRC requests **CESTPATH RESETHP(YES|NO)** parameter, and the **CESTPAIR ONLINSEC (YES|NO)** parameter.

ICKDSF

ICKDSF supports PPRC operations in the z/OS, z/VM, VSE/ESA and stand-alone environments. The PPRC functions are supported through the **PPRCOPY** command. This command uses the **DELPAIR**, **ESTPAIR**, **DELPATH**, **ESTPATH**, **SUSPEND**, **RECOVER**, and **QUERY** parameters. The **ESTPAIR** command does not support the **XD** option needed to establish pairs in a nonsynchronous PPRC-XD relationship, so PPRC Extended Distance cannot be managed using ICKDSF.

ICKDSF does not support consistency grouping. Therefore, you cannot request any long busy states on failing devices or freeze control unit pairings, and thus it cannot have the **CGROUP** command or parameter.

ICKDSF provides a subset of the PPRC commands for stand-alone use when no operating system is available. This subset can be used to issue **PPRCOPY RECOVER** commands at a secondary site and gain access to the recovery site volumes.

However, because ICKDSF contains a subset of the PPRC commands and functions, then for the z/VM and the VSE/ESA environments using ICKDSF, additional procedures must be

established to control recovery in the event of a disaster as compared to TSO-based implementations.

ESS Copy Services Web user interface

The ESS Copy Services provides a Web browser interface that can be used to control the PPRC functions on all the zSeries based operating systems (z/OS, OS/390, VM/ESA, z/VM, VSE/ESA, TPF). See 7.2.1, “ESS Copy Services Web user interface” on page 196.

For a complete description of the specific windows and the use of the ESS Copy Services Web user interface, refer to the publication *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448. To find this publication go to:

<http://ssddom02.storage.ibm.com/techsup/webnav.nsf/support/2105>

and click the **Documentation** link.

7.5.5 PPRC management on the open systems

ESS Copy Services provides two ways of management for PPRC in the open systems:

- ▶ ESS Copy Services Web user interface (WUI).

ESS Copy Services provides a Web browser interface that can be used to control the PPRC functions in the open systems environments. The WUI is discussed in 7.2.1, “ESS Copy Services Web user interface” on page 196.

When using the WUI, PPRC pairs can be established in three different ways:

- From the Volumes window (based on volumes)
- From the Logical subsystems window (based on entire logical subsystems)
- From the Task window (once a task for PPRC is created)

Before establishing the pairs, the paths must be defined and made available. This is done using the Paths window provided by the Web browser interface.

- ▶ The Java-based ESS Copy Services command-line interface (CLI).

The CLI interface allows administrators to start predefined tasks, which contain copy services commands, from a command line. This command-line interface is available for selected operating systems. The CLI is discussed in 7.2.2, “ESS Copy Services command-line interface (CLI)” on page 198.

For detailed information and how to use, the following publications should be referenced: *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448 and *IBM TotalStorage Enterprise Storage Server Copy Services Command-Line Interface Reference*, SC26-7449.

7.6 PPRC implementation on the ESS

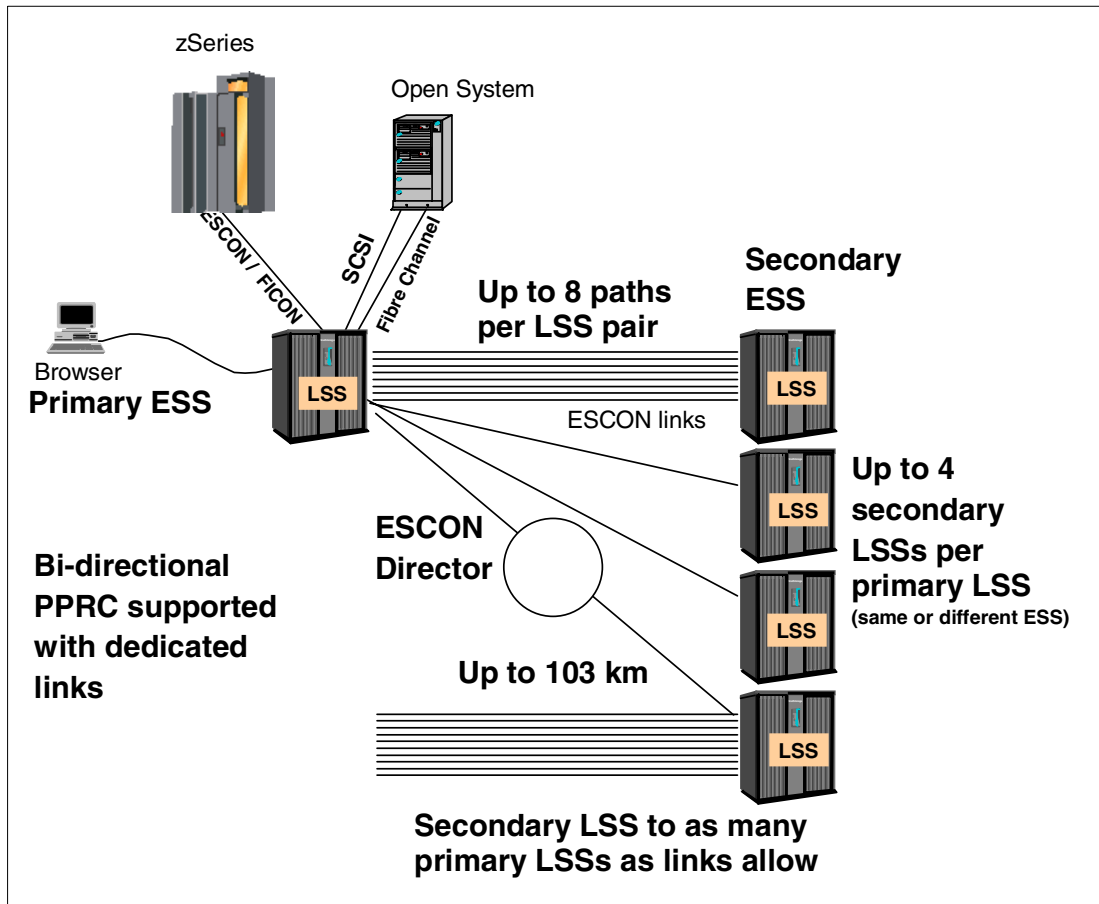


Figure 7-9 PPRC configuration options

As with other PPRC implementations, you can establish PPRC pairs only between storage control units of the same type, which means that primary and secondary must both be ESSs with the optional PPRC function enabled.

ESCON links

ESCON links between ESS subsystems are required. The local ESS is usually called primary if it contains at least one PPRC source volume, while the remote ESS is called secondary if it contains at least one PPRC target volume. An ESS can act as primary and secondary at the same time if it has PPRC source and target volumes. This mode of operation is called bi-directional PPRC.

The PPRC logical paths are established between the LSSs of the ESSs. A primary LSS can be connected to up to four secondary LSSs (see Figure 7-9), from the same or different ESSs. A secondary LSS can be connected to as many primary LSSs as ESCON links are available.

PPRC links are unidirectional. This means that a physical ESCON link can be used to transmit data from the primary ESS to the secondary ESS. If you want to set up a bi-directional PPRC configuration with source and target volumes on each ESS, you need ESCON PPRC links in each direction (see Figure 7-10 on page 211). The number of links depends on the write activity to the primary volumes.

Primary PPRC ESCON ports are dedicated for PPRC use. An ESCON port is operating in *channel mode* when it is used on the primary ESS for PPRC I/O to the secondary ESS.

A port operates in *control unit mode* when it is talking to a host. In this mode, a secondary ESCON port can also receive data from the primary control unit when the ESS port is connected to an ESCON director. So, the ESCON port on the secondary ESS does not need to be dedicated for PPRC use. The switching between control unit mode and channel mode is dynamic.

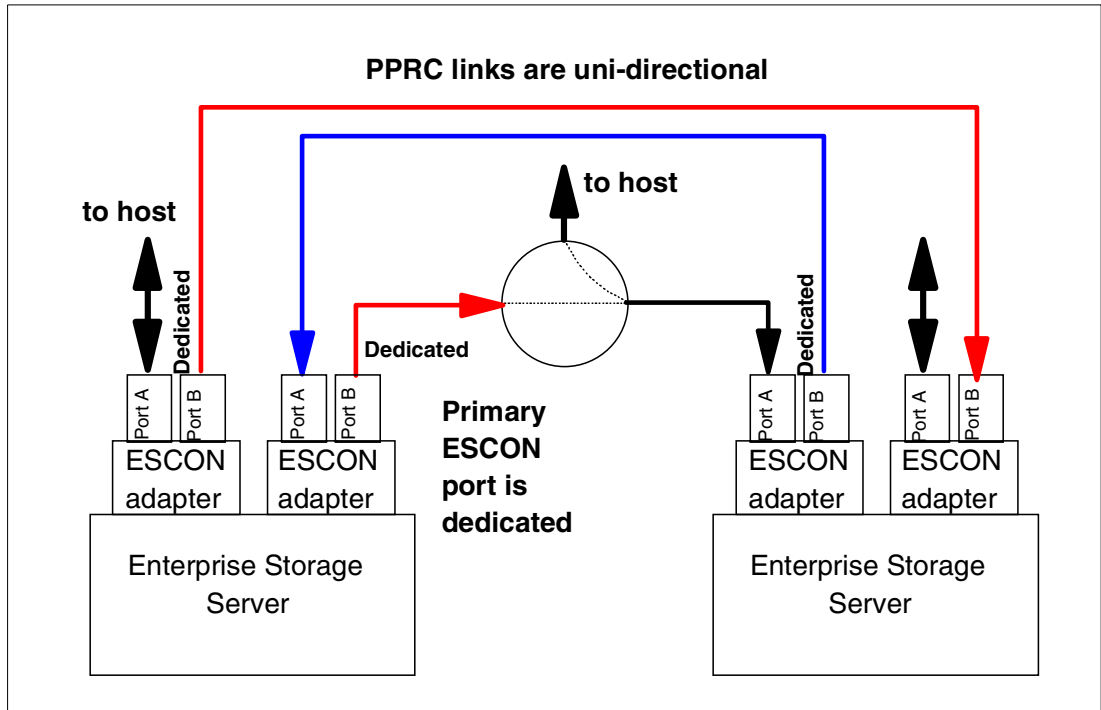


Figure 7-10 PPRC links

If there is any logical path defined over an ESCON port to a zSeries host, you cannot switch this port to channel mode for PPRC to use as primary port. You must first configure all logical paths from the zSeries host to that port offline. Now you can define a PPRC logical path over this ESCON port from the primary to the secondary ESS. When you establish the logical path, the port will automatically switch to channel mode.

PPRC logical paths

Before PPRC pairs can be established, logical paths must be defined between the logical control unit images. The ESS supports up to 16 CKD logical control unit images and up to 16 FB controller images. An ESCON adapter supports up to 64 logical paths. A pair of LSSs can be connected with up to eight logical paths. You establish logical paths between control unit images of the same type over physical ESCON links (see Figure 7-11 on page 212).

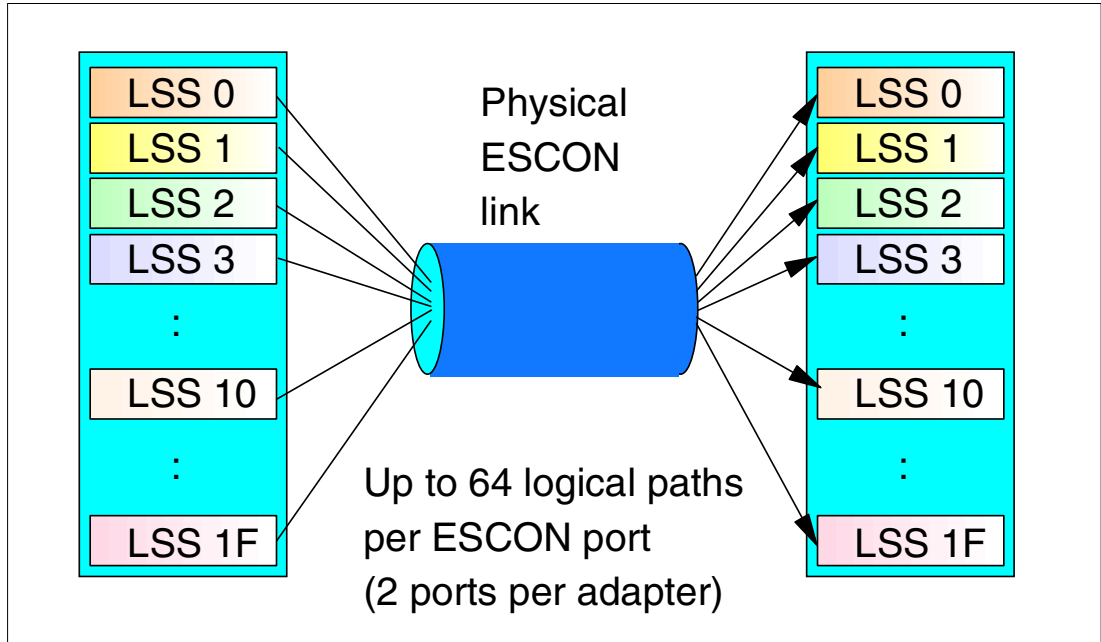


Figure 7-11 PPRC logical paths

7.7 PPRC Extended Distance (PPRC-XD)

Peer-to-Peer Remote Copy Extended Distance (PPRC-XD) brings new flexibility to the IBM TotalStorage Enterprise Storage Server and PPRC. PPRC-XD is a nonsynchronous long-distance copy option for both open systems and zSeries servers (see Figure 7-12 on page 213). PPRC-XD can operate at very long distances, even continental distances, well beyond the 103 km (maximum supported distance for synchronous PPRC) with minimal impact on the applications. Distance is limited only by the network and channel extenders technology capabilities.

This section presents an overview of the characteristics of PPRC Extended Distance, for detailed information and how to use it, refer to the redbook *IBM TotalStorage Enterprise Storage Server PPRC Extended Distance*, SG24-6568.

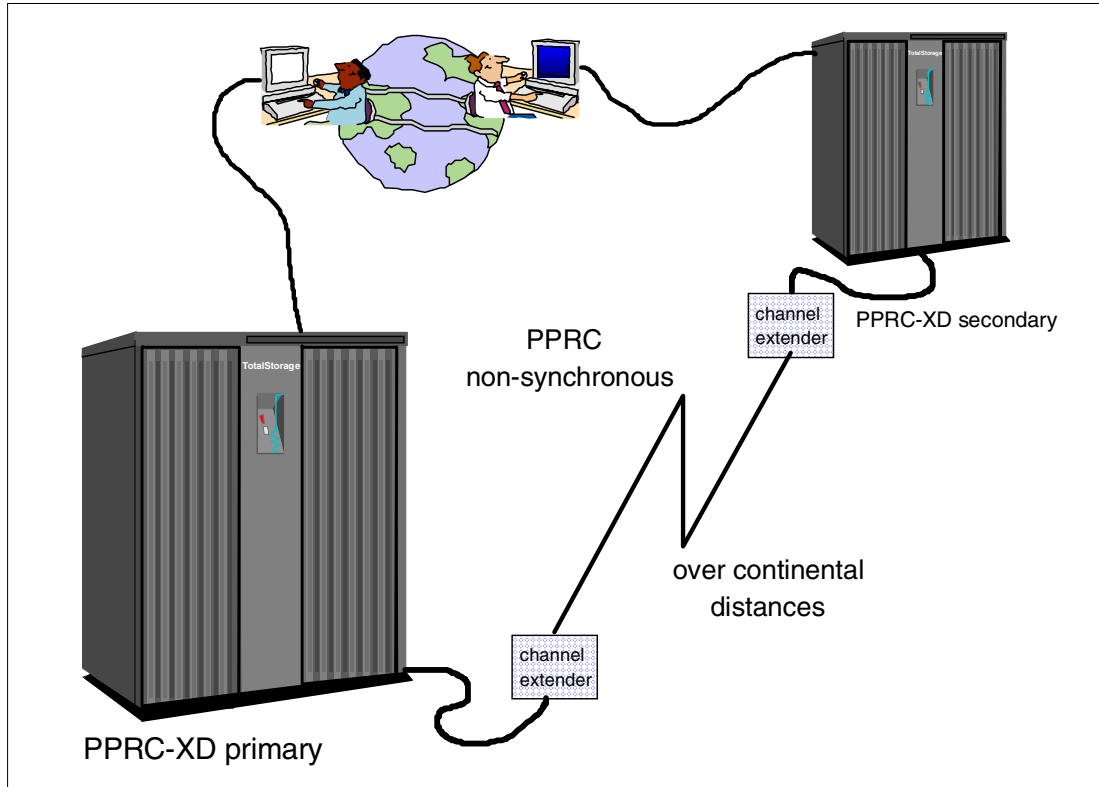


Figure 7-12 PPRC Extended Distance (PPRC-XD)

The nonsynchronous operation of PPRC-XD, together with its powerful throughput (copying sets of track updates only) and the supported channel extenders improvements, make PPRC-XD an excellent copy solution at very long distances and with minimal application performance impact. PPRC Extended Distance is an excellent solution for:

- ▶ Data copy
- ▶ Data migration
- ▶ Offsite backup
- ▶ Transmission of data base logs
- ▶ Application recovery solutions based on periodic PiT (point-in-time) copies of data

7.7.1 PPRC-XD operation

In PPRC-XD, the primary volumes updates are mirrored to the secondary volumes in a nonsynchronous operation while the application is running. Due to the nonsynchronous transmission of the updates, the write operations at the primary site (thus the application's write response time) are not affected by any transmission delay independently of the distance.

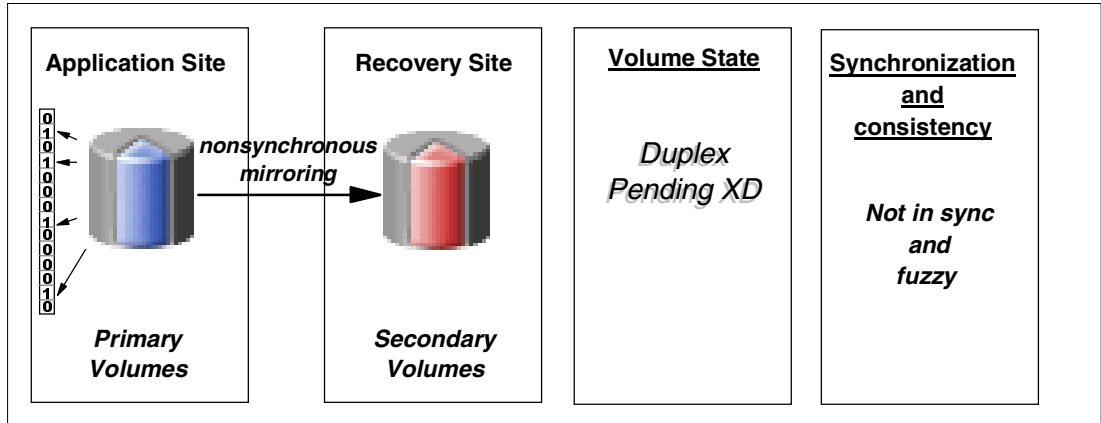


Figure 7-13 Duplex-pending XD volume state

With PPRC-XD, there is an addition to the PPRC traditional volume states (PPRC volume states are illustrated in Figure 7-8 on page 205). This is the duplex-pending XD state (see Figure 7-13). While in this duplex-pending XD-state, PPRC is doing nonsynchronous mirroring of primary volumes updates to the secondary site. PPRC-XD will periodically cycle through the bitmap of each volume for updated tracks and place them in a batch for copying to the secondary. This is a throughput-oriented, very efficient method of nonsynchronous mirroring.

The pair will stay in this status until either a command is issued to go into synchronous mode (duplex state), or a command is issued to suspend the pair (suspended state) or delete the PPRC-XD pair (simplex state). Figure 7-14 shows the basics of PPRC-XD.

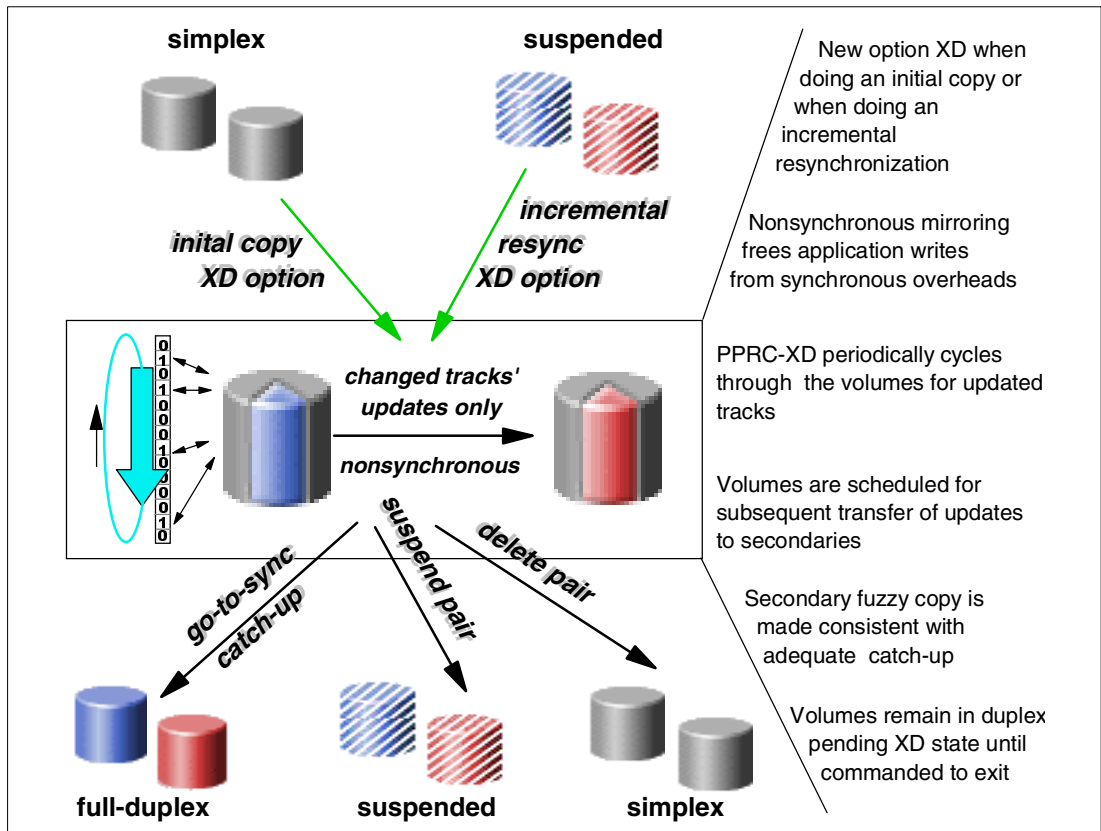


Figure 7-14 PPRC-XD Basic operation

Catch-up operation (go-to-sync)

PPRC-XD *catch-up* is the name of the transition that occurs to a PPRC-XD pair when it goes from its normal out-of-sync condition until it reaches a full synchronous condition. At the end of this transition, primary and secondary volumes become fully synchronized.

The catch-up transition can be accomplished by commanding PPRC to *go-to-SYNC*, so the volume pair leaves the duplex-pending XD state and reaches the duplex state. From this moment on, if the pairs were not immediately suspended, primary write updates would be synchronously transmitted to the recovery site.

Also, the catch-up transition can be accomplished by temporarily quiescing the application writes to the primary volumes and waiting for PPRC-XD to finish the synchronization.

When triggering a catch-up by commanding *go-to-SYNC*, the **rsQueryComplete** command (for open systems that support the CLI) and the **IEA494** system messages (for z/OS and OS/390 systems) allow you to detect the moment when the volume pairs reach the duplex state, that is when the catch-up is completed.

When doing the catch-up transition from duplex-pending XD to duplex by commanding PPRC *go-to-SYNC*, you will not want any synchronous copy operations to occur if the volumes being mirrored are separated by long distances, beyond the 103 km. For this, there is a new option when the copy options is selected that allows you to ask PPRC to suspend the pair as soon as it is established (this option is available only from the ESS Copy Services Web user interface, not with the TSO commands).

Note: The commanded catch-up transition (*go-to-SYNC*) should be triggered when the application write activity is low or preferably none. Also note that in the *go-to-SYNC* operation, PPRC does an incremental copy of the changed tracks' updates. This is a very efficient synchronization technique that minimizes the time needed for the catch-up transition.

The *go-to-SYNC* operation, and how it is controlled, is presented in detail in the redbook *IBM TotalStorage Enterprise Storage Server PPRC Extended Distance*, SG24-6568.

7.7.2 Data consistency

While in PPRC-XD state, the volume pairs remain in duplex-pending XD state. When the application is doing writes on the primary volumes, the updates to secondary volumes are nonsynchronous. This means secondary volumes are keeping a *fuzzy* copy of the data. Application-dependent writes are not assured to be applied in the same sequence as written on the primary.

Because of the nonsynchronous characteristics of PPRC-XD, at any time there will be a certain amount of application-updated data that will not be reflected at the secondary volumes. This data corresponds to the tracks that were updated since the last volume bitmap scan was done. These are the out-of-sync tracks, which can be checked by **CQUERY** (TSO/E), **rsQuery** (CLI), or on the Information window (Web interface).

The catch-up transition will bring the pairs back to synchronous, either by setting them to duplex or by stopping write updates to the primary. This will synchronize the pairs in a minimum interval of time. When reaching the duplex state, the pairs can be temporarily suspended to flash the secondary before resuming the PPRC-XD relation between the pairs. This FlashCopy will be a tertiary consistent point-in-time copy. Or, if it is a database application and the archive logs are being mirrored, then once the pairs catch up and are suspended, then the logs can be applied on the shadow database.

7.7.3 Automation

As already mentioned, to build consistency at the recovery site, you may do short planned outages of the application in order to quiesce the write activity upon the primary volumes. These data consistency windows involve several operations on the selected set of volumes. To make this procedure more efficient, we recommend the implementation of automation routines.

When implementing automation scripts in open servers with CLI support, you will find the `rsQueryComplete` command very useful, because it signals the completion of the task.

When automating in z/OS and OS/390 environments, you will find the state change messages that the system issues very useful when PPRC relationships transition from one state to a different one. IEA494I and IEA494E messages can be detected by automation routines and used to suspend secondary PPRC operations, using the `CGROUP` command with the `FREEZE` parameter.

7.7.4 PPRC-XD for initial establish

If a large number of volumes needs to be established in synchronous mode (full duplex), when this operation is triggered the volumes will not be reaching the duplex state all at the same time. While in this transition, at some moment volumes that already reached the full duplex state (thus receiving synchronous updates) will coexist with volumes still in the duplex pending state, receiving track updates. This coexistence may adversely affect the application performance. If this is the case for a particular environment at initial establish, then it can be worth it to evaluate the use of PPRC-XD for the initial establish of the whole set of volumes:

- ▶ The volumes are initially established in PPRC-XD mode
- ▶ The establish is monitored for the pairs when reaching a 95-98% in-sync
- ▶ Then the go-to-SYNC command can be issued

This implementation can lessen the impact of the establish process on the application performance.

7.7.5 Implementing and managing PPRC Extended Distance

PPRC-XD is part of PPRC, and so the ESS Copy Services setup for starting to use it is similar to PPRC. Nonetheless because of its different way of operating, its long distance, and its nonsynchronous characteristics, different considerations apply at the time of planning for PPRC-XD implementation and uses. These specific implementation and uses considerations are discussed in detail in the redbook *IBM TotalStorage Enterprise Storage Server PPRC Extended Distance*, SG24-6568.

7.8 PPRC connectivity

This section reviews the connectivity and distance considerations for PPRC implementations. These considerations are summarized in Figure 7-15 on page 217.

- ▶ PPRC can use the following connection capabilities:
 - ESCON Directors
 - Channel Extenders over Wide Area Network (WAN) lines
 - Dense Wave Division Multiplexors (DWDM) over dark fibers
- ▶ Synchronous (SYNC) maximum supported distance is 103 km
 - For slighter longer distances an RPQ must be submitted
- ▶ Non-synchronous (XD) can be used over continental distances
 - Only limited by the network and channel extender technology capabilities
- ▶ The connectivity infrastructure is transparent to PPRC
- ▶ Evaluation, qualification, approval and support of PPRC configurations using channel extender products is the sole responsibility of the channel extender vendor
- ▶ The channel extender vendors and DWDM vendors should be consulted regarding prerequisites when using their products
- ▶ A complete and current list of PPRC supported environments, configurations, networks, and products is available at:

<http://www.ibm.com/storage/hardsoft/products/ess/supserver.htm>

Figure 7-15 Connectivity - distance and support considerations

PPRC connectivity capabilities

PPRC can be used with following connectivity technologies:

- ▶ ESCON Directors
- ▶ Channel Extenders over Wide Area Network (WAN) lines
- ▶ Dense Wave Division Multiplexors (DWDM) on dark fiber

7.8.1 PPRC supported distances

The supported distances vary according to the method of PPRC mirroring.

PPRC synchronous transmission

With PPRC synchronous, the supported distance is 103 km. For slightly longer distances beyond 103 km, IBM offers an RPQ (Request for Price Quotation).

PPRC-XD nonsynchronous transmission

With PPRC nonsynchronous, the distance can be a continental distance. This is limited only by channel extender and network technology capabilities. Channel extenders are needed to implement distances beyond the 103 km.

7.8.2 PPRC channel extender support

Channel extender vendors connect PPRC environments via a variety of wide area network (WAN) connections, including Fibre Channel, Ethernet/IP, ATM-OC3, and T1/T3.

When using channel extender products with PPRC, the channel extender vendor will determine the maximum supported distance between the primary and secondary site. The channel extender vendor should be contacted for their distance capability, line quality requirements, and WAN attachment capabilities.

For a complete list of PPRC-supported environments, configurations, networks, and products, refer to:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

The channel extender vendor should be contacted regarding hardware and software prerequisites when using their products in a PPRC configuration. Evaluation, qualification, approval and support of PPRC configurations using channel extender products is the sole responsibility of the channel extender vendor.

7.8.3 PPRC Dense Wave Division Multiplexor (DWDM) support

Wave Division Multiplexing (WDM) and Dense Wave Division Multiplexing (DWDM) is the basic technology of fibre optical networking. It is a technique for carrying many separate and independent optical channels on a single dark fibre.

A simple way to envision DWDM is to consider that at the primary end, multiple fiber optic input channels are bunched together into a single fiber optic cable. Each channel is encoded as a different wavelength, imaging each channel with a different color. At the receiving end, the DWDM fans out the different optical channels. DWDM provides the full bandwidth capability of the individual channel.

For a complete list of PPRC supported environments, configurations, networks and products refer to:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

The DWDM vendor should be contacted regarding hardware and software prerequisites when using their products in an ESS PPRC configuration.

7.9 Concurrent Copy

Concurrent Copy is a copy function available for the z/OS and OS/390 operating systems. It involves a System Data Mover (SDM) only found in OS/390 and z/OS. This section presents an overview of the Concurrent Copy function. Detailed information can be found in the publication *z/OS DFSMS Advanced Copy Services*, SC35-0428.

Concurrent Copy allows you to generate a copy or a dump of data while applications are updating that data. It can be invoked with the DFSMSDss utility COPY and DUMP commands. Concurrent Copy works not only on a full-volume basis, but also at a data set level. Also the target is not restricted only to DASD volumes in the same ESS. For Concurrent Copy, the target can also be a tape cartridge or a DASD volume on another ESS (see Figure 7-16 on page 219).

The System Data Mover (SDM), a DFSMS/MVS component, reads the data from the source (volume or data set) and copies it to the target.

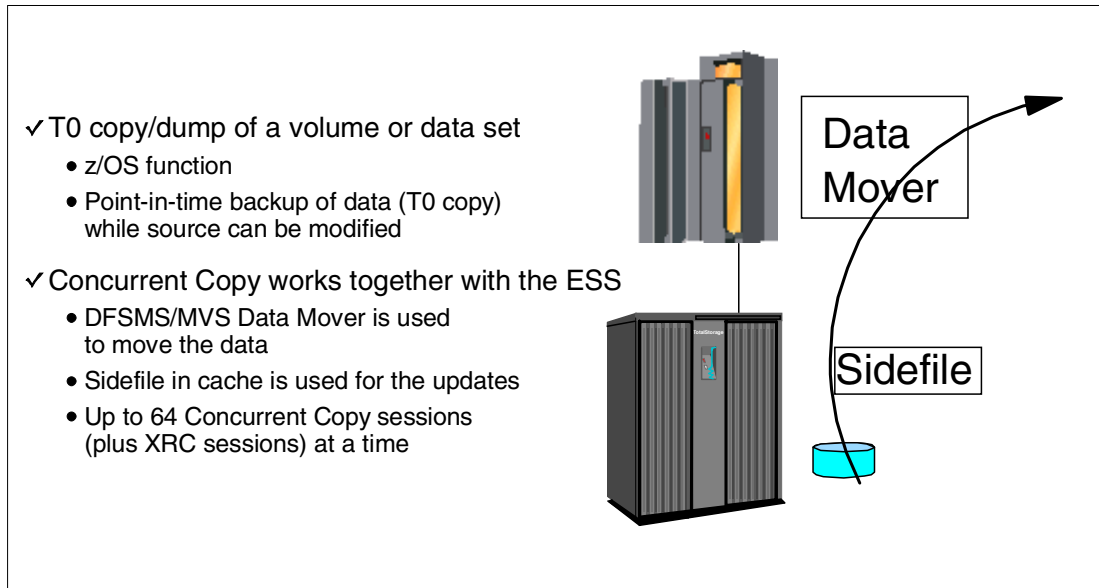


Figure 7-16 Concurrent Copy

Concurrent Copy process

For the copy process, we must distinguish between the logical completion of the copy and the physical completion. The copy process is logically complete when the System Data Mover has figured out what to copy. This is a very short process. After the logical completion, updates to the source are allowed while the System Data Mover, in cooperation with the IBM TotalStorage Enterprise Storage Server, ensures that the copy reflects the state of the data when the COPY command was issued. When an update to the source is to be performed and this data has not yet been copied to the target, the original data is first copied to a sidefile in cache before the source is updated.

Concurrent Copy on the ESS

Concurrent Copy on the ESS works the same way as on the previous IBM 3990 Model 6 and the IBM 9390 models 1 and 2 storage controllers. Concurrent Copy is initiated using the CONCURRENT keyword in DFSMSdss or in applications that internally call DFSMSdss as the copy program, for example, DB2's COPY utility.

The System Data Mover establishes a session with the ESS. There can be up to 64 sessions active at a time (including sessions for Extended Remote Copy XRC copy function).

If you used Concurrent Copy on an IBM 3990 or 9390, or if you used Virtual Concurrent Copy on an IBM RVA, no changes are required when migrating to an ESS.

Concurrent Copy and FlashCopy

If DFSMSdss is instructed to do a Concurrent Copy by specifying the CONCURRENT keyword, and the copy is for a full volume with the target within the same logical storage subsystem (LSS) of the ESS, then DFSMSdss will choose the fastest copy process and start a FlashCopy copy process instead of Concurrent Copy.

7.10 Extended Remote Copy (XRC)

Extended Remote Copy (XRC) is a copy function available for the z/OS and OS/390 operating systems. It involves a System Data Mover (SDM) that is found only in OS/390 and z/OS. This section presents an overview of the Extended Remote Copy function. Detailed discussion and how-to-use information can be found in the publication *z/OS DFSMS Advanced Copy Services, SC35-0428*.

XRC maintains a copy of the data asynchronously, at a remote location over unlimited distances. It is a combined hardware and software solution that offers data integrity and data availability that can be used as part of business continuance solutions implementations, for workload movement and for data migration. XRC is an optional feature on the ESS (refer to Appendix A, “Feature codes” on page 263).

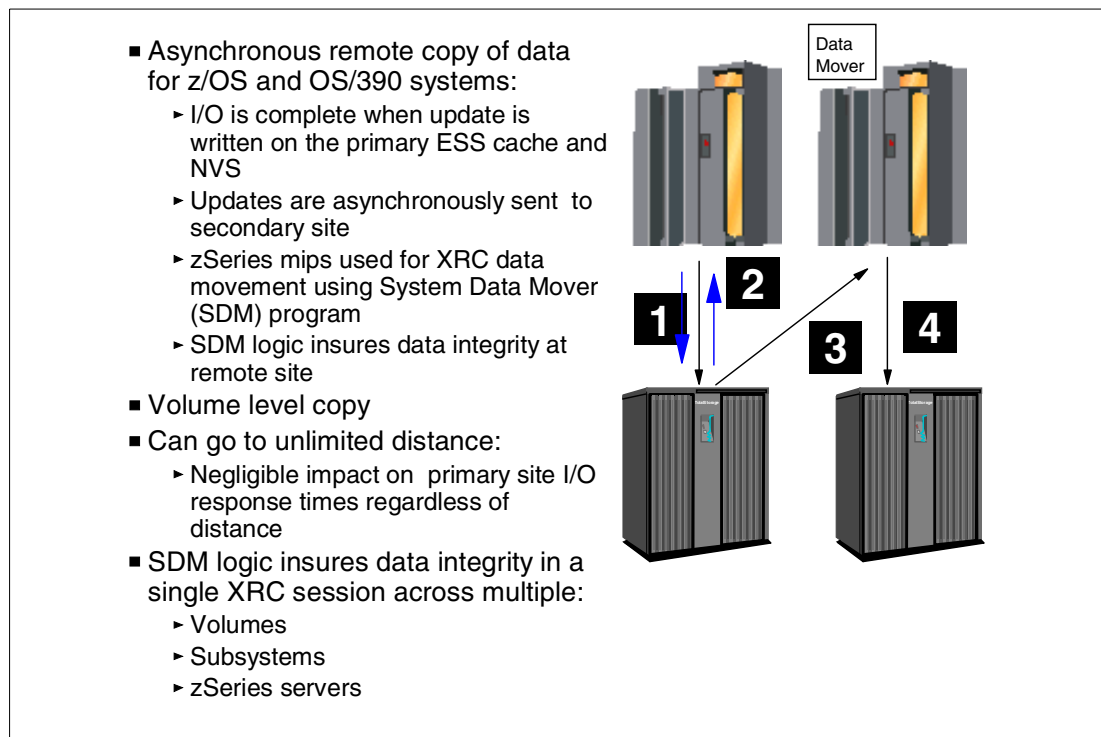


Figure 7-17 Extended Remote Copy

7.10.1 Overview

Extended Remote Copy (XRC) is an asynchronous remote mirroring solution. It uses the System Data Mover (SDM) of the DFSMS/MVS program, and hence works only in the z/OS and OS/390 operating systems.

Applications accessing the same source volumes need to have the same internal clock time, which is provided by a sysplex timer, within the sysplex. Each write I/O on an XRC mirrored volume gets a time stamp.

Applications doing write I/Os to primary (source) volumes — (1) in Figure 7-17 — get a Device End status (write I/O complete) as soon as the data has been secured in cache and NVS of the primary ESS (2). The System Data Mover that may be running at the recovery site host reads out the updates to the XRC source volumes from the cache (3) and sends them to the secondary volume on a remote storage control (4).

The System Data Mover needs to have access to all primary storage control units with XRC volumes the Data Mover has to handle, as well as to the target storage control units. In this way, the Data Mover has the higher authority to all control units involved in the remote mirroring process and can assure data integrity across several primary storage control units. The I/O replication in the right sequence to the target volumes is guaranteed by the System Data Mover.

XRC is designed as a solution that offers the highest levels of data integrity and data availability in a disaster recovery, workload movement, and/or device migration environment. It provides real-time data shadowing over extended distances. With its single command recovery mechanism furnishing both a fast and effective recovery, XRC provides a complete disaster recovery solution.

7.10.2 Invocation and management of XRC

XRC can be controlled by using TSO/E commands, or through the DFSMSdfp Advanced Services ANTRQST Macro, which calls the System Data Mover API.

Managing a large set of mirrored volumes over long distance requires automation for monitoring and decision making. The GDPS/XRC automation package, developed from customer requirements, offers a standard solution to that demand.

User information for managing XRC can be found in the publication *z/OS DFSMS Advanced Copy Services*, SC35-0428.

7.10.3 XRC implementation on the ESS

The implementation of XRC on the ESS is compatible with XRC's implementation on the previous IBM 3990 Model 6 and 9390 Models 1 and 2.

- ESS's XRC implementation is compatible with the previous implementation on IBM 3990-6s and 9390-1 / -2
- Support of XRC version 2 functions
 - Planned outage support
 - Multiple reader support (max 64 /LSS)
 - Dynamic balancing of application write bandwidth vs SDM read performance
 - Floating utility device
 - Or use 1-cylinder utility device
- Support of XRC version 3 functions
- ESS also supports
 - Geographically Dispersed Parallel System facility for XRC
 - Remote Copy Management facility for XRC

Figure 7-18 XRC implementation

The ESS supports all the XRC Version 2 enhancements (refer to Figure 7-18 on page 221):

- ▶ Planned outage support is the capability to suspend for a while the SDM function (to do host maintenance, for example) and later to re-synchronize the suspended volumes with the updates without a full copy of the mirrored volumes.
- ▶ The System Data Mover supports up to 64 readers per logical subsystem.
- ▶ Dynamic balancing of application write bandwidth with SDM read performance.
- ▶ Floating utility device: this facility is the default on ESS.
- ▶ Use of 1-cylinder utility device (ESS's custom volume capability).

- ✓ ESS supports XRCVersion 3 functions
 - ▶ Unplanned outage support
 - ▶ Use of new performance enhanced CCWs
 - ▶ Coupled XRC (CXRC)
- ✓ XRC unplanned outage (suspend/resume) support
 - ▶ New level of suspend/resume support unique to ESS
 - ▶ If the XRC session is suspended for any reason, ESS XRC will track changes to the primary database in hardware.
 - ▶ Unlimited ESS XRC suspend duration and no application impact
 - ESS XRC starts and maintains hardware-level bitmap during suspend
 - ▶ Upon XRC resynch, ESS XRC transmits only incremental changed data
 - ▶ Non-ESS storage controllers use cache memory to hold updates during suspend
 - Limitation on suspend time and possible application impact

Figure 7-19 XRC unplanned outage support

The ESS provides support for XRC Version 3 functions (summarized in Figure 7-19).

Unplanned outage support

On an IBM 3990 Model 6, XRC pairs could be suspended only for a short time or when the System Data Mover was still active. This was because the bitmap of changed cylinders was maintained by the System Data Mover in the software. This software implementation allowed a re synchronization of pairs only during a planned outage of the System Data Mover or the secondary subsystem.

The IBM TotalStorage Enterprise Storage Server starts and maintains a bitmap of changed tracks in the hardware, in non-volatile storage (NVS), when the connection to the System Data Mover is lost or an XRC pair is suspended by command.

The bitmap is used in the re-synchronization process when you issue the XADD SUSPENDED command to re synchronize all suspended XRC pairs. Copying only the changed tracks is much faster compared to a full copy of all data. With ESS's XRC support, a re-synchronization is now possible for a planned as well as an unplanned outage of one of the components needed for XRC to operate.

New performance enhanced CCWs

The IBM TotalStorage Enterprise Storage Server supports performance-enhanced channel command words (CCWs) that allow reading or writing more data with fewer CCWs and thus reducing the overhead of previous CCW chains.

The System Data Mover will take advantage of these performance-enhanced CCWs for XRC operations on an ESS.

7.10.4 Coupled Extended Remote Copy (CXRC)

Coupled Extended Remote Copy (CXRC) expands the capability of XRC so that very large installations that have configurations consisting of thousands of primary volumes can be assured that all their volumes can be recovered to a consistent point in time (see Figure 7-20)

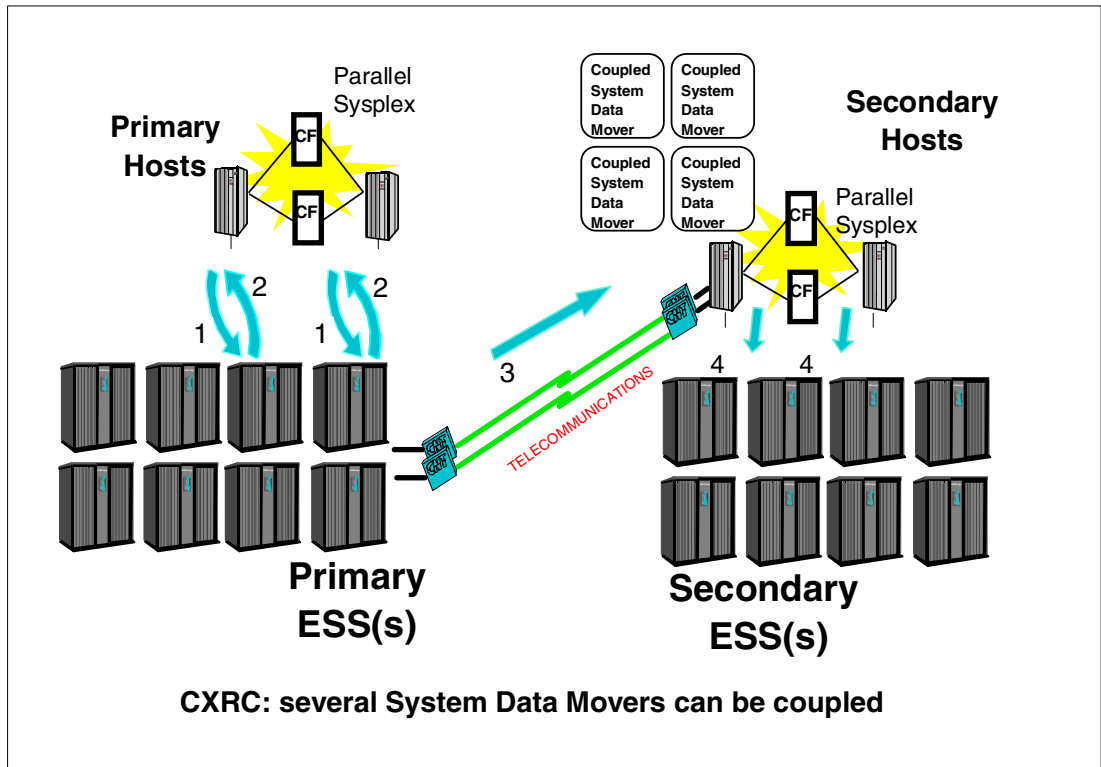


Figure 7-20 CXRC performance and scalability

In a disaster recovery situation, for the recovered data of a particular application to be usable the data must be recovered to a consistent point in time. CXRC provides this capability by allowing multiple XRC sessions to be coupled with a coordination of consistency between them. This way, all the volumes in all the coupled sessions can be recovered to a consistent point in time.

Migration for current users of XRC

Existing XRC users will be able to continue to use XRC in the same manner. The new CXRC support will not directly affect existing XRC sessions. The new support is provided by new commands and keywords and enhancements to existing commands. When the new support is not in use, all existing commands and keywords continue to work as currently defined. Once the support has been installed, users can choose to start one or more additional XRC sessions and couple them into a master session.

7.10.5 XRC FICON support

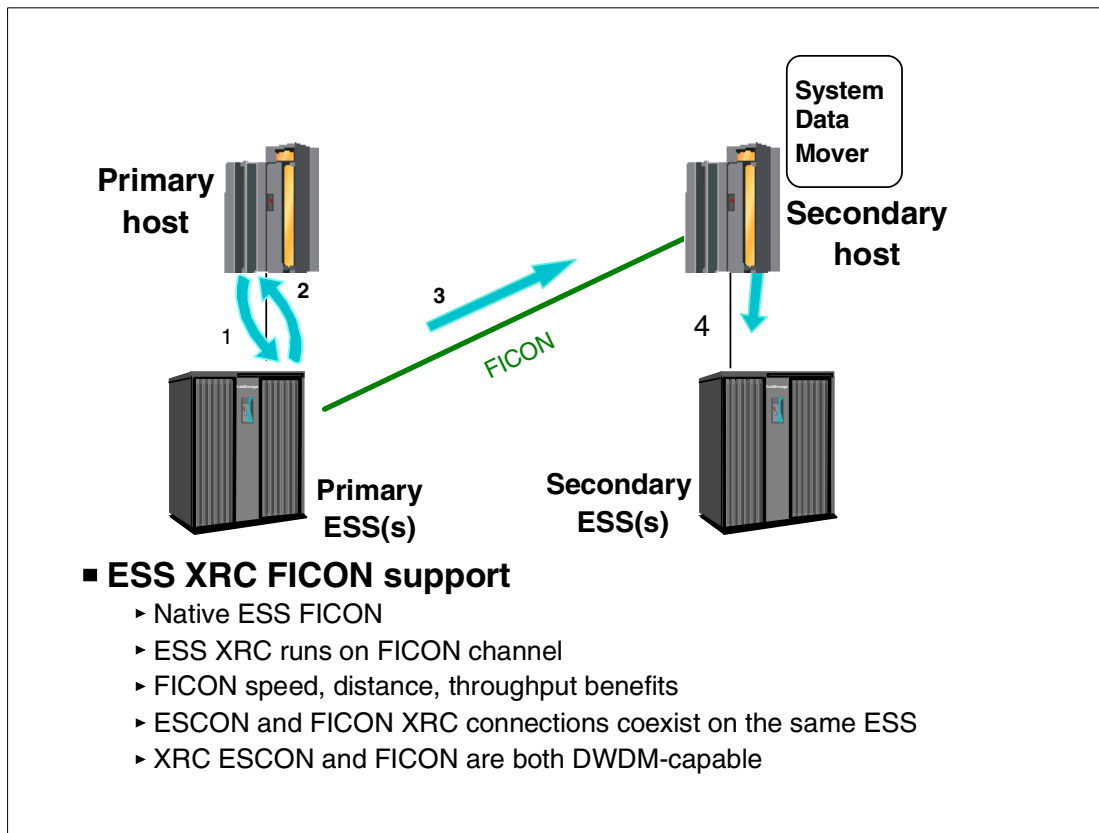


Figure 7-21 XRC FICON support

The SDM benefits greatly from the higher bandwidth FICON channels provide to read the updates from the primary ESS. The transfer bandwidth for a single thread read for XRC is at least five times better than with ESCON due to the larger block sizes of data transfers.

The improved bandwidth of FICON together with the longer unrepeated distance support (up to 100 km) results in a considerable improvement over ESCON channels, when using direct channel attachment. These capabilities position XRC as a disaster recovery solution in the range of metropolitan areas. With ESCON, the recommendation was channel extenders through telecom lines beyond 25 km. With FICON, there is no need to use channel extenders until the distance exceeds 100 km.

7.11 ESS Copy Services for iSeries

The iSeries servers take advantage of the IBM TotalStorage Enterprise Storage Server advanced copy functions FlashCopy and PPRC. FlashCopy and PPRC are supported with OS/400 V4R5 or later. Because of the particular way in which the storage is managed by the iSeries servers, some considerations apply.

7.11.1 The Load Source Unit (LSU)

The Load Source Unit (LSU) is a special DASD in the iSeries. This is the device that is used to IPL the system (among other things). It is similar to a boot drive. All other user data can be

located on external DASD units, but the LSU must be an internal drive. This is because the system cannot be IPLed from an I/O adapter (IOA) supporting external drives.

Due to the nature of iSeries Single Level Storage, it is necessary to consider the LSU as a special case. On other open system platforms, such as UNIX and Windows NT, each volume can be identified with its contents. The iSeries is different, since all storage is considered as a single large address space. The LSU is within this address space.

Therefore, to use facilities such as Peer-to-Peer Remote Copy (PPRC) or FlashCopy to do a hardware copy of the volumes attached to the iSeries, then the LSU from the internal drive must be mirrored into the ESS, to ensure the whole single level storage is copied.

7.11.2 Mirrored internal DASD support

Support has been added to the iSeries to allow internal DASD, such as the LSU, to be mirrored to external DASD. This requires that the external DASD reports as unprotected, even though in practice, it may actually be protected in a RAID 5 rank within the ESS.

Before implementing remote load source mirroring, the latest maintenance required to support this function must be verified. The maintenance APARs, PTFs, and PSP buckets can be found at:

<http://www.as400service.ibm.com>.

7.11.3 LSU mirroring

The primary reason for using remote load source mirroring is to get a copy of the LSU into the ESS, so that the entire DASD space in single-level storage can be duplicated by the hardware facilities such as PPRC and FlashCopy.

When using remote load source mirroring, normal OS/400 rules for mirroring apply. Both the source and target disk drives must be the same size, although they can be of different drive types and speeds. It is simply capacity that must match.

To allow the LSU to be mirrored, the target disk must be unprotected, because OS/400 does not allow mirroring any disk to another disk that is already protected. This must be done when first defining the LUN in the ESS. Once the LUN has been specified, the designated protection cannot be changed. Normally only the LUN for the LSU will be defined as unprotected. All other LUNs will be defined as protected, reflecting their true status to OS/400. To do this, Unprotected must be selected in the Add Volumes window of the ESS Specialist when doing the logical configuration.

Detailed information can be found in the redbook *iSeries in Storage Area Networks A Guide to Implementing FC Disk and Tape with iSeries*, SG24-6220.

7.11.4 FlashCopy

There is no native iSeries command interface to initiate FlashCopy service on the iSeries. This copy function is managed using the ESS Specialist.

Note: The iSeries backup and recovery offers a very similar function to FlashCopy. This function is called Save-while-active and also provides a T0 point-in-time copy of the data.

7.11.5 PPRC

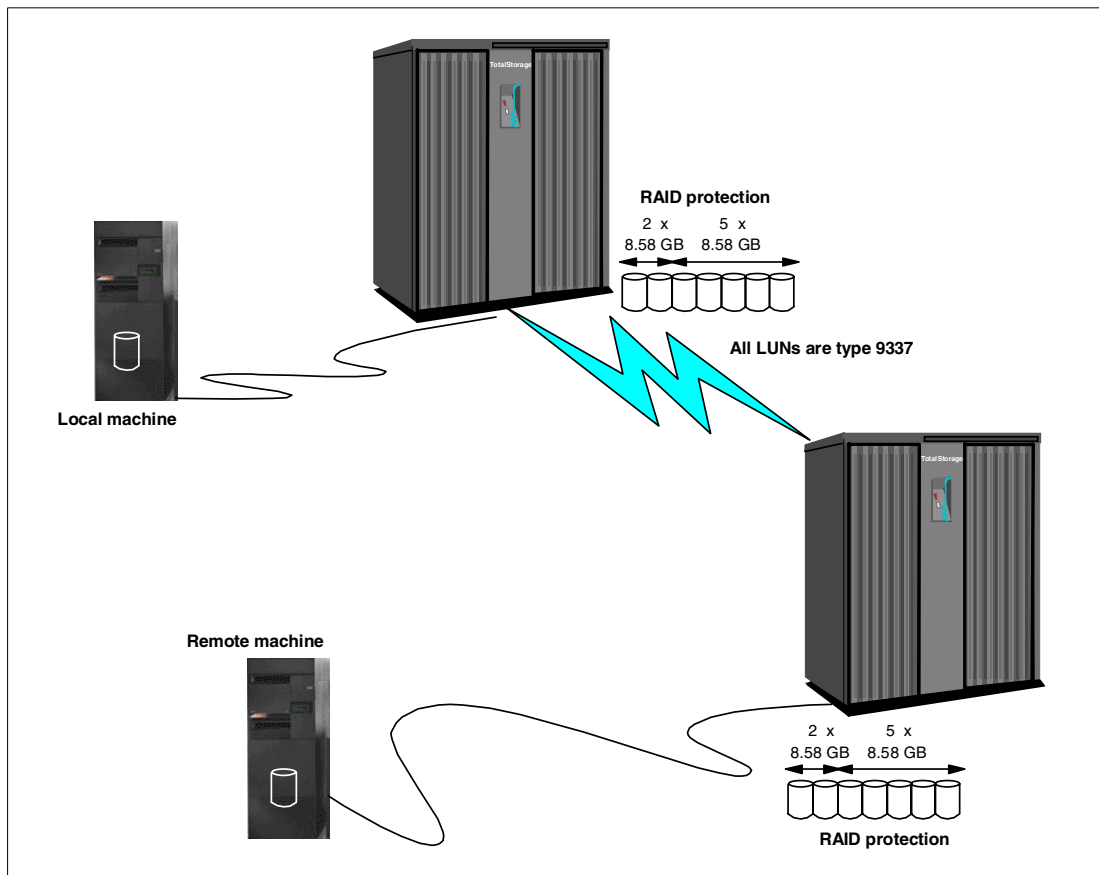


Figure 7-22 iSeries and ESS PPRC implementation

Figure 7-22 shows an example of iSeries and ESS in a PPRC implementation when the attachment between the iSeries and the ESS is SCSI. Because the attachment between the iSeries and the ESS is a SCSI attachment, then all the DASDs are 9337s.

In this implementation, the local iSeries has only one internal volume, and this is the load source unit (LSU). Within the local ESS there is the remote load source mirrored pair and the rest of the local iSeries LUNs.

On the remote site there is another ESS. This remote ESS contains the PPRC copy of the remote load source unit and also the PPRC copies of all the other LUNs from the local iSeries.

In the event of a disaster at the production site (the local site in Figure 7-22), then the backup iSeries at the remote site recovers to ESS LUNs at that same remote site. This recovery process includes an abnormal IPL on the iSeries. This IPL could be many hours. This time can be reduced by implementing OS/400 availability functions to protect applications and databases, for example journaling, commitment control, and system managed access paths (SMAP).

Note: In a configuration with remote load source mirroring, the system can only IPL from, or perform a main storage dump to, an internal LSU.

For more information on the implementation of the ESS advanced copy functions in an iSeries environment, refer to *iSeries in Storage Area Networks A Guide to Implementing FC Disk and Tape with iSeries*, SG24-6220.

7.12 Geographically Dispersed Parallel Sysplex (GDPS)

How would a shutdown of your system affect your business? Do you put off system maintenance and upgrades to avoid system downtime? What about a site disaster? Is your business-critical processing and data protected from a site disaster? In today's highly competitive e-business world, outages will have a devastating impact on a business; they could mean its demise. Many companies have inadequate business continuance plans developed on the premise that back office and manual processes will keep the business running until computer systems are available. Characteristics of these recovery models allow critical applications to recover within 24-48 hours, with data loss potentially exceeding 24 hours, and full business recovery taking days or weeks. As companies transform their business to compete in the e-marketplace, business continuance strategies and availability requirements must be reevaluated to ensure that they are based on today's business requirements.

In e-business, two of the most stringent demands are continuous availability and near transparent disaster recovery (D/R). Systems that deliver continuous availability combine the characteristics of high availability and continuous operations to always deliver high levels of service (24x7x365). High availability is the attribute of a system to provide service at agreed-upon levels and mask unplanned outages from end users. Continuous operation, on the other hand, is the attribute of a system to continuously operate and mask planned outages from end users. To attain the highest levels of continuous availability and near-transparent D/R, the solution must be based on geographical clusters and data mirroring. These technologies are the backbone of the Geographically Dispersed Parallel Sysplex (GDPS) solution.

GDPS complements a multi-site Parallel Sysplex by providing a single, automated solution to dynamically manage storage subsystem mirroring, processors, and network resources to allow a business to attain continuous availability and near-transparent business continuity (disaster recovery) without data loss. GDPS is designed to minimize and potentially eliminate the impact of any failure, including disasters or a planned site outage. It provides the ability to perform a controlled site switch for both planned and unplanned site outages, with no data loss, maintaining full data integrity across multiple volumes and storage subsystems and the ability to perform a normal Data Base Management System (DBMS) restart - not DBMS recovery - at the other site. GDPS is application independent and, therefore, covers the customer's complete application environment.

GDPS is enabled by means of key IBM technologies:

- ▶ Parallel Sysplex
- ▶ Systems Automation for OS/390
- ▶ IBM TotalStorage Enterprise Storage Server
- ▶ PPRC (Peer-to-Peer Remote Copy)
- ▶ PPRC XD (Peer-to-Peer Remote Copy Extended Distance)
- ▶ XRC (Extended Remote Copy)



Support information

The IBM TotalStorage Enterprise Storage Server Model 800 provides advanced functions for both zSeries and open systems.

This chapter gives complementary information about the support and requirements of the ESS and its features for the different environments, but most importantly it gives recommendations on how to find the detailed and current support information that will be needed at the time of installation and implementation of the ESS. ESS Specialist and IBM TotalStorage Expert (ESS Expert feature) requirements are also covered.

Important: Because the information presented in this chapter changes frequently, you are strongly advised to always refer to the online resources listed in this chapter for current information.

8.1 Key requirements information

The ESS ships with IBM Licensed Internal Code (LIC) that is licensed for use by a customer on a specific machine, designated by serial number, under the terms and conditions of the IBM Customer Agreement or the IBM Agreement for Licensed Internal Code.

The IBM TotalStorage Enterprise Storage Server Model 800 requires ESS LIC Level 2.0.0 or later. All the features and functions described so far are available for the ESS Model 800 either as a standard or as an optional feature, but in either case not all are supported in all the environments.

Current information on supported environments and minimum operating system levels is available at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

Current information on zSeries, S/390, and open system servers, operating systems, host adapters and connectivity products supported by the ESS is also available at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

The information available at this site should be referenced for the latest support data.

8.2 zSeries environment support

The zSeries sections of this chapter present a brief reminder of the ESS features that are relevant to the CKD server environment, and then proceed to show by operating system which of those features are supported. Some sections may also include extra operating system-specific information.

The IBM TotalStorage Enterprise Storage Server Model 800 supports the following zSeries and S/390 operating systems at the current versions and releases:

Table 8-1 Supported zSeries and S/390 operating systems

Operating System	ESCON Support?	FICON Support?
z/OS	YES	YES
z/OS.e	YES	YES
OS/390	YES	YES
z/VM	YES	YES
VM/ESA	YES	YES
Transaction Processing Facility (TPF)	YES	YES
Linux for S/390	YES	NO (*)
(*) Note: FCP is supported.		

PSP buckets

The *Preventive Service Planning* (PSP) buckets contain operating system support and planning information that includes *application program analysis reports* (APARs) and *program temporary fixes* (PTFs). They are available from your IBM Service Representative or by contacting the IBM Software Support Center. For the ESS ask for the 2105DEVICE bucket.

Important: For the latest supported server information, see the ESS with IBM zSeries and S/390 Systems Support Summary link from the Web page:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver>

8.2.1 Multiple Allegiance and I/O Priority Queuing

Multiple Allegiance and *I/O Priority Queuing* are standard ESS hardware functions available for the shared environments. See 6.3, “Multiple Allegiance” on page 179 and 6.4, “I/O Priority Queuing” on page 181 for functions details.

8.2.2 Parallel Access Volumes

Parallel Access Volumes (PAV) enable a single attaching z/OS or OS/390 server image to simultaneously process multiple I/O operations to the same logical volume. This is achieved by defining multiple addresses for the same volume. Defining device aliases can significantly decrease device queue delays, which means improved throughput and I/O response times. PAV is an optional feature of the ESS Model 800. See 6.2, “Parallel Access Volume” on page 166 for a detailed description.

8.2.3 PPRC

Peer-to-Peer Remote Copy (PPRC) is a hardware solution that enables the synchronous shadowing of data from one site to a second site. It is an optional feature of the ESS Model 800. See 7.5, “Peer-to-Peer Remote Copy (PPRC)” on page 203 for a detailed description.

If you are planning to implement ESS Copy Services in a zSeries environment, we recommend you refer to the redbook *Implementing ESS Copy Services on S/390*, SG24-5680.

8.2.4 PPRC-XD

PPRC Extended Distance (PPRC-XD) is a nonsynchronous long-distance remote copy function that works over much greater distances than synchronous PPRC. It is supplied as part of the optional PPRC feature of the ESS Model 800. See 7.7, “PPRC Extended Distance (PPRC-XD)” on page 212 for a detailed description.

If you are planning to utilize PPRC-XD, we recommend you see *IBM TotalStorage Enterprise Storage Server: PPRC Extended Distance*, SG24-6568.

8.2.5 FlashCopy

FlashCopy is a function that makes a point-in-time (T0) copy of data. It is an optional feature of the ESS Model 800. See 7.4, “FlashCopy” on page 200 for a detailed description.

8.2.6 FICON support

FICON channels provide enhanced performance for the execution of channel programs allowing the use of CCW and data pre-fetching and pipelining. The FICON channel protocols are fully compatible with existing channel programs and access methods. IBM software has been changed to exploit this function, but an installation should review the readiness of its non-IBM software. See 2.14, “FICON host adapters” on page 41 for more detailed information.

When planning for FICON attachment, refer to the redbook *FICON Native Implementation and Reference Guide*, SG24-6266.

Important: When planning to install the ESS with FICON attachment, the latest processor PSP buckets should always be reviewed prior to installation. These are listed in Table 8-2.

Table 8-2 Processor PSP buckets for FICON support

Processor	PSP upgrade
zSeries 900	2064DEVICE
9672 G6/G6	9672DEVICE
FICON	FICON

8.2.7 Control-Unit-Initiated Reconfiguration

In large configurations, quiescing channel paths in preparation for upgrades or service actions is a complex, time-consuming and potentially error-prone process. *Control-Unit-Initiated Reconfiguration* (CUIR) automates the process of quiescing and re-enabling channel paths, reducing the time required for service actions and reducing requirements on the operations staff, hence reducing the possibility for human error. See 3.3.3, “CUIR” on page 54 for detailed information.

Important: CUIR support may also require host processor channel microcode updates. Please check PSP bucket information for required PTFs.

8.2.8 Large Volume Support

Large Volume Support (LVS) increases the size of the largest CKD volume from 10,017 to 32,760 cylinders (approximately 27.8 GB).

Tip: Install Large Volume Support on your operating system *before* starting with the definition of large volumes on the ESS.

Important: Check with your OEM software product vendors for changes to their products which may be required, if you are going to use large volumes.

8.3 z/OS support

Because the information presented in this chapter changes frequently, the latest z/OS operating system support information should be retrieved from:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

and from the respective PSP buckets.

For information on Z/OS refer to <http://www.ibm.com/os390>.

Tip: The 2105DEVICE PSP bucket subset name for z/OS is 2105MVS/ESA.

Multiple Allegiance and I/O Priority Queuing

Current z/OS and OS/390 systems exploit these features. I/O Priority Queuing requires *Workload Manager* (WLM) running in Goal mode. For details on Workload Manager, visit:

<http://www.ibm.com/s390/wlm>

Parallel Access Volumes

Current z/OS and OS/390 systems exploit this feature. If WLM is running in Goal mode, then dynamic PAVs are supported; otherwise static PAVs are supported.

PPRC

Current systems support this feature via the powerful TSO commands, the ICKDSF utility, or the ESS Copy Services Web user interface (WUI). Also the DFSMSdfp Advanced Services ANTRQST macro can invoke this function. When planning for PPRC implementation, the redbook *Implementing ESS Copy Services on S/390*, SG24-5680, should be referenced.

PPRC-XD

Current systems support for PPRC-XD is via TSO commands in addition to the ESS Copy Services Web user interface (WUI). When planning for PPRC-XD implementation, the redbook *IBM TotalStorage Enterprise Storage Server PPRC Extended Distance*, SG24-6568 should be referenced.

FlashCopy

Current z/OS and OS/390 systems support this feature. FlashCopy can be controlled using TSO commands or the DFSMSdss utility, as well as from the ESS Copy Services WUI.

Tip: If you are planning to implement ESS Copy Services on a zSeries environment, we recommend you visit the DFSMS SDM Copy Services home page at:

<http://www.storage.ibm.com/software/sms/sdm>

FICON

Current z/OS and OS/390 systems support this attachment. Table 8-3 shows the relevant PSP buckets to check.

Table 8-3 Processor and Channel PSP buckets for z/OS and OS/390

	PSP upgrade	PSP subset
Processor	2064DEVICE	2064/ZOS
	2064DEVICE	2064/OS390
	9672DEVICE	9672OS3690G5+
Channel	FICON	FICON/ZOS
	FICON	FICON/MVSESA

CUIR

z/OS and OS/390 support CUIR. Consult the latest PSP bucket information for software level and PTFs requirements.

Large Volume Support

Large Volume Support (LVS) is available on z/OS and OS/390 operating systems, and the ICKDSF and DFSORT utilities. Consult the respective PSP buckets for information the latest software level support and PTFs.

Large Volume Support needs to be installed on all systems in a sysplex prior to sharing data sets on large volumes. Shared system/application data sets cannot be placed on large volumes until all system images in a sysplex have Large Volume Support installed.

For some levels of DFSMS/MVS, a coexistence PTF is provided that will allow these system levels to coexist in the same sysplex with LVS systems. You must install this PTF in order to prevent unpredictable results that may arise from systems without Large Volume Support accessing volumes that have more than 10,017 cylinders. The coexistence PTF will:

- ▶ Prevent a device with more than 10,017 cylinders from being varied online to the system
- ▶ Prevent a device from coming online during an IPL if it is configured with more than 10,017 cylinders

Coexistence PTFs will also be required by DFSMSHsm for some previous releases of OS/390. Consult the latest PSP bucket information for software level support and PTFs requirements.

8.3.1 Other related support products

Several components of z/OS software have been changed to support the ESS.

ICKDSF

The formatting of CKD volumes is performed when you set up the ESS after defining the CKD volumes through the ESS Specialist.

To use the volumes in z/OS, only a `minimal init` by ICKDSF is required to write a label and VTOC index.

The following ICKDSF functions are not supported and not required on an ESS:

- ▶ ANALYZE with DRIVETEST
- ▶ INSTALL
- ▶ REVAL
- ▶ RESETICD

Access Method Services

The AMS LISTDATA command provides *Rank Counter* reports. This is how you can get information on the activities of a RAID rank. While z/OS performance monitoring software only provides a view of the logical volumes, this rank information shows the activity of the physical drives.

EREP

EREP provides problem incidents reports and uses the device type 2105 for the ESS.

Media Manager and AOM

Both components take advantage of the performance-enhanced CCWs on ESS, and they limit extent access to a minimum to increase I/O parallelism.

DFSMSdss and DASD ERP

Both of these components also make use of the performance-enhanced CCWs for their operations.

8.4 z/VM support

z/VM and VM/ESA support the IBM TotalStorage Enterprise Storage Server Model 800 natively and for guests. It also allows z/OS and TPF guests to use their exploitation support.

Because the information presented in this chapter changes frequently, the latest z/VM and VM/ESA operating system support information should be consulted at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

and from the respective PSP buckets.

For information on z/VM refer to:

<http://www.ibm.com/vm>

Tip: The 2105DEVICE PSP bucket subset name for z/VM is 2105VM/ESA.

Multiple Allegiance and I/O Priority Queuing

VM systems can take advantage of Multiple Allegiance in a shared environment. The priority byte, however, is not set by VM and therefore I/O Priority Queuing is not exploited.

Parallel Access Volumes

The PAV function is supported on VM systems for guest use via dedicated (otherwise known as attached) DASD. The VM support includes:

- ▶ The CP QUERY PAV command, which displays information about the Parallel Access Volume devices on the system.
- ▶ Enhancements to the CP QUERY DASD DETAILS command to display additional information if the queried device is a Parallel Access Volume.
- ▶ A CP Monitor Record, which has been added to Domain 6 (I/O) to record state change interrupts that indicate a change in the Parallel Access Volumes information: Record 20 – MRIODSTC – State change.

PPRC

PPRC support is provided for VM systems by the use of the ESS Copy Services Web user interface (WUI), or the more limited ICKDSF utility.

PPRC-XD

PPRC-XD support is provided for VM systems by the ESS Copy Services Web user interface (WUI).

FlashCopy

z/VM allows a native CP user to initiate a FlashCopy function of a source device to a target device on an ESS via the CP FLASHCOPY command. The ESS Copy Services Web user interface can also be used.

FICON

z/VM contains support for FICON. The most complete and up-to-date list of required maintenance (APARs and PTFs) will be available in the PSP buckets shown in Table 8-4.

Table 8-4 Processor PSP buckets for z/VM and VM/ESA

PSP upgrade	PSP subset
2064DEVICE	2064/ZVM
2064DEVICE	2064VM/ESA
9672DEVICE	9672VM/ESA

CUIR

z/VM supports CUIR. Consult the latest PSP bucket information for software level and PTFs requirements.

Large Volume Support

z/VM has Large Volume Support. z/VM supports these large volumes as native devices as well as for guests. Consult the latest PSP bucket information for software level and PTFs requirements.

8.4.1 Guest support

z/VM does not recognize the ESS by its device type 2105, but sees it as an IBM 3990 Model 6. However, when the operating system senses the control unit, the returned function bits can be interpreted by guest systems to see what functions are supported on this control unit.

z/VM will allow guest systems to use the performance-enhanced CCWs, PAVs, and advanced copy functions. FlashCopy is supported for guest use for dedicated (attached) volumes or for full-pack minidisks.

8.5 VSE/ESA support

VSE/ESA sees the IBM TotalStorage Enterprise Storage Server as an IBM 3990 Model 6 storage control.

Because the information presented in this chapter changes frequently, the latest VSE/ESA operating system support information should be consulted at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

and from the respective PSP buckets.

For information on VSE/ESA refer to:

<http://www.ibm.com/vse>.

Tip: The 2105DEVICE PSP bucket subset name for VSE is 2105VSE/ESA.

Multiple Allegiance and I/O Priority Queuing

VSE systems can take advantage of Multiple Allegiance in a shared environment. The priority byte, however, is not set by VSE and therefore I/O Priority Queuing is not exploited.

Parallel Access Volumes

PAVs are not supported.

PPRC

PPRC support is provided for VSE/ESA by the use of the ESS Copy Services Web user interface (WUI), or the more limited ICKDSF utility.

PPRC-XD

PPRC-XD is supported via the ESS Copy Services Web user interface (WUI).

FlashCopy

VSE/ESA provides FlashCopy support with VSE/ESA 2.5 and later. FlashCopy can also be used via the ESS Copy Services Web user interface (WUI).

VSE/ESA uses the IXFP SNAP command to invoke FlashCopy. In order to prevent its casual use, you may use the VSE/ESA STACK command to hide the IXFP command in the following fashion:

```
STACK IXFP | * " IXFP " ...command reserved by Systems Programming
```

If a user were to issue the command `IXFP SNAP,120:123` after using the above STACK command, the result would be that the command would be treated as a comment command and logged on SYSLOG. Then you may have batch jobs for invoking FlashCopy and use the UNSTACK command at the beginning of the job step to allow the IXFP command to be used and then reissue the STACK command at the end of the step.

More information on the STACK command is available at:

<http://www-1.ibm.com/servers/eserver/zseries/os/vse/pdf/vseesht.pdf>

You arrive here from the System Hints and Tips from VSE entry on the VSE/ESA Web page at:

<http://www-1.ibm.com/servers/eserver/zseries/os/vse/library/library.htm>

FICON

VSE/ESA 2.6 contains support for FICON. The most complete and up-to-date list of required maintenance (APARs and PTFs) will be available in the PSP buckets listed in Table 8-5.

Table 8-5 Processor PSP buckets for VSE/ESA

PSP upgrade	PSP subset
2064DEVICE	2064VSE/ESA
9672DEVICE	9672VSE/ESA

CUIR

CUIR is not supported.

Large Volume Support

Large Volume Support has been previewed for VSE/ESA Version 2 Release 7. It is also planned to be made available for Version 2 Release 6 (please refer to “Statement of general direction” on page 28 for plans and previews announcements terms).

8.6 TPF support

The inherent performance of the IBM TotalStorage Enterprise Storage Server Model 800 makes it an ideal storage subsystem for a TPF environment.

Because the information presented in this chapter changes frequently, the latest TPF operating system support information should be consulted at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

and from the respective PSP buckets.

For information on TPF refer to:

<http://www.ibm.com/tpf>.

Tip: The 2105DEVICE PSP bucket subset name for TPF is 2105/TPF.

8.6.1 Control unit emulation mode

To use an ESS in a TPF system, at least one logical subsystem in the ESS has to be defined to operate in IBM 3990 Model 3 TPF control unit emulation mode. The volumes defined in this logical subsystem can be used by TPF.

8.6.2 Multi Path Locking Facility

The ESS supports the *Multi Path Locking Facility* as previously available on IBM 3990 control units for TPF environments.

8.6.3 TPF support levels

The ESS is supported by TPF Version 4 Release 1. With the appropriate software maintenance applied, the ESS function bits are interpreted and TPF will use the performance-enhanced CCWs. Consult the latest PSP bucket information for software level and PTFs requirements.

Multiple Allegiance and I/O Priority Queuing

Multiple Allegiance was a function already available for TPF environments on IBM 3990 systems as an RPQ. TPF benefits from the ESS's Multiple Allegiance and I/O Priority Queuing functions.

Parallel Access Volumes

PAVs are not supported.

PPRC

PPRC is supported via the ESS Copy Services Web user interface (WUI). Consult the latest PSP bucket information for software level and PTFs requirements.

PPRC-XD

PPRC-XD is supported via the ESS Copy Services Web user interface (WUI). Consult the latest PSP bucket information for software level and PTFs requirements.

FlashCopy

FlashCopy is supported via the IBM TotalStorage ESS Copy Services Web interface. Host PTFs may be required.

FICON

TPF contains support for FICON. Consult the latest PSP bucket information for software level and PTFs requirements.

CUIR

CUIR is not supported.

8.7 Linux

Linux can be run natively as a stand-alone or as a logical partition (LPAR), or under z/VM. Several Linux distributions are available for S/390 and zSeries.

Because the information presented in this chapter changes frequently, the latest LINUX for S/390 operating system support information should be consulted at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

and from the respective PSP buckets.

For information on LINUX see the Enterprise servers link at:

<http://www.ibm.com/linux>

Multiple Allegiance and I/O Priority Queuing

Linux systems can take advantage of Multiple Allegiance in a shared environment. The priority byte, however, is not set by Linux and therefore I/O Priority Queuing is not exploited.

Parallel Access Volumes

PAVs are not supported.

PPRC

PPRC is supported via the IBM TotalStorage ESS Copy Services Web interface.

PPRC-XD

PPRC-XD is supported via the IBM TotalStorage ESS Copy Services Web interface.

FlashCopy

FlashCopy is supported via the IBM TotalStorage ESS Copy Services Web interface.

FICON

FICON is not supported, but FCP attachment has been previewed for Linux on zSeries.

CUIR

CUIR is not supported.

8.8 Open systems environment support

The IBM TotalStorage Enterprise Storage Server Model 800 supports the majority of the open systems environments. New levels of operating systems, servers of different manufacturers, file systems, host adapters, and cluster software are constantly announced in the market. Consequently, storage solutions must be tested with these new environments in order to have the appropriate technical support. Because the information on supported servers changes frequently, you are strongly advised to always refer to the online resources listed in this section for current information.

Important: For the latest information on open servers supported, operating systems, adapters, and connectivity, click the **ESS Open Systems Support Summary** link from the Web page at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

Following are some tips for some of the fixed block attaching platforms.

IBM iSeries and AS/400

The latest maintenance information (APARs, PTFs and PSP buckets) for the iSeries servers can be found at:

<http://www.as400service.ibm.com>

The 2105DEVICE PSP bucket subset name for iSeries and AS/400 is 2105AS/400.

IBM NUMA-Q platforms

NUMA-Q servers only support ESS attachment via Fibre Channel, and not by SCSI adapters.

Sun

The Sun host file `sd.conf` may need to be edited to see all of the ESS LUNs. Without editing this file, LUN 0 can be seen, but not LUN 1 and above.

AIX and NT

The 2105DEVICE PSP bucket subset name for AIX and NT is 2105AIX/NT.

8.8.1 Installation scripts

Check the publication *IBM TotalStorage Enterprise Storage Server Host System Attachment Guide*, SC26-7446 for instructions on relevant installation scripts for your platform. For example, on AIX an installation script will update the ODM so that AIX will recognize the 2105.

Also the redbook *IBM TotalStorage Enterprise Storage Server: Implementing the ESS in Your Environment*, SG24-5420-01 should be referenced when planning your open server installation.

8.8.2 IBM Subsystem Device Driver

On selected platforms the *IBM Subsystem Device Driver* (SDD, formerly Data Path Optimizer) gives servers the ability to balance I/O across multiple paths and, in the process, improve subsystem performance. Not only does this provide better performance, it also improves availability. Should a path have a problem, the workload is automatically switched to an alternate path to the ESS from the server.

If you plan to use SDD, then you should refer to the IBM Subsystem Device Driver (SDD) link at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

Here you will find current information on supported functions and servers, and up-to-date software requirements information.

You can find the procedure for this support and additional software maintenance level information at:

<http://www.ibm.com/storage/support/techsup/swtechsup.nsf/support/sddupdates>

8.8.3 ESS Copy Services command-line interface (CLI)

The ESS Copy Services Web user interface (WUI) is available for all supported platforms. The ESS Copy Services WUI function is a Java/CORBA-based application that runs on the host server and requires a TCP/IP connection to each ESS under it.

ESS Copy Services also includes a command-line interface (CLI) feature. Using the CLI, users are able to communicate with the ESS Copy Services from the server command line. See 7.4.4, “FlashCopy management on the open systems” on page 203 and 7.5.5, “PPRC management on the open systems” on page 209 for information on how to invoke the ESS Copy Services.

If you are planning to use ESS Copy Services command-line interface, please refer to the Copy Services section of:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

8.8.4 Boot support

If you plan to use the ESS as a boot device, then please refer to the Boot Support section of:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

8.8.5 PPRC

Peer-to-Peer Remote Copy (PPRC) is a hardware solution that enables the synchronous shadowing of data from one site to a second site. It is an optional feature that can be ordered with the IBM TotalStorage Enterprise Storage Server Model 800. See 7.5, “Peer-to-Peer Remote Copy (PPRC)” on page 203 for more detailed information.

If you are planning to implement ESS Copy Services on any of the supported open environments, we recommend you read *Implementing ESS Copy Services on UNIX and Windows NT/2000*, SG24-5757, and *IBM @server iSeries in Storage Area Networks: A Guide to Implementing FC Disk and Tape with iSeries*, SG24-6220.

If you are planning to use PPRC, please refer to the Copy Services section of:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

8.8.6 PPRC-XD

PPRC Extended Distance (PPRC-XD) is a nonsynchronous long-distance copy that works over much greater distances than synchronous PPRC. It is supplied as part of the optional PPRC feature that can be ordered with the IBM TotalStorage Enterprise Storage Server

Model 800. See 7.7, “PPRC Extended Distance (PPRC-XD)” on page 212 for more detailed information.

If you are planning to utilize PPRC-XD, we recommend you refer to *IBM TotalStorage Enterprise Storage Server: PPRC Extended Distance*, SG24-6568.

If you are planning to use PPRC, please refer to the Copy Services section of:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

8.8.7 FlashCopy

FlashCopy is a function that makes a point-in-time (T0) copy of data. It is an optional feature that can be ordered with the IBM TotalStorage Enterprise Storage Server Model 800. See 7.4, “FlashCopy” on page 200 for more detailed information

If you are planning to use PPRC, please refer to the Copy Services section of:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

8.9 ESS Specialist

The IBM TotalStorage Enterprise Storage Server Specialist (ESS Specialist) is the Web user interface that is included as standard with the IBM TotalStorage Enterprise Storage Server. You use the ESS Specialist to view machine resources, problem status and machine configuration, and to modify the ESS configuration.

The ESS Specialist also provides the means for invoking ESS Copy Services to establish Peer-to-Peer Remote Copy and FlashCopy without having to involve the operating system running in the host server.

The ESS Specialist can be accessed from the browser provided with the ESS Master Console, which comes with the ESS.

Also using an Internet browser, such as Netscape Navigator or Microsoft Internet Explorer, the storage system administrator can access the ESS Specialist from a desktop or mobile computer as supported by the network.

Detailed information on supported Web browsers can be found in “Using a Web browser to access the ESS”, in *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448 and in the redbook *IBM TotalStorage Enterprise Storage Server: Implementing the ESS in Your Environment*, SG24-5420-01.

8.10 IBM TotalStorage Expert

The IBM TotalStorage Expert Version 2 is an optional chargeable product that gathers and presents information that can significantly help storage administrators manage one or more IBM TotalStorage Enterprise Storage Servers (ESS Expert feature) and Enterprise Tape Libraries (ETL Expert feature). Capabilities are provided for performance management, asset management, and capacity management. For more information on the ESS Expert, see 5.10, “IBM TotalStorage Expert” on page 160.

A Web browser such as Netscape Navigator 4.73 or Microsoft Internet Explorer 5.5 provides access to the Expert server, as shown in Figure 8-1 on page 243.

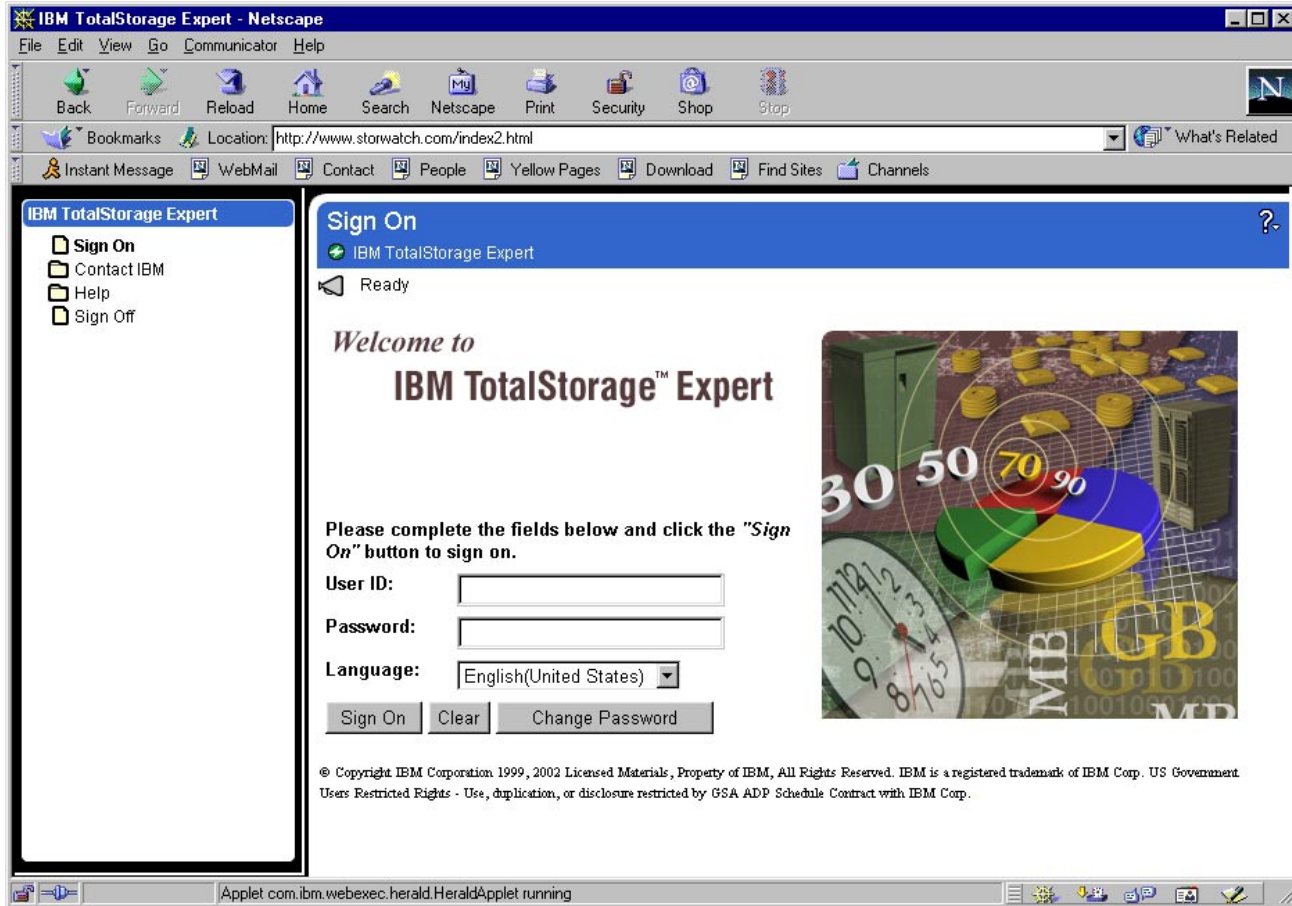


Figure 8-1 IBM TotalStorage Expert Welcome window

The TotalStorage Expert Version 2 can be installed on the following platforms:

- ▶ Windows 2000 Server or Advanced Server
- ▶ AIX 4.3.3 + fixes

Important: When the IBM TotalStorage Expert is installed, in addition to the Expert code, the installation process installs several other applications (such as DB2, WebSphere, and JDK). These programs cannot pre-exist on the target server because explicit releases are included and different levels cannot be present on the server. If any of these products are installed, then you must un-install them first.

For detailed up-to-date information, visit the IBM TotalStorage Expert home page at:

<http://www.storage.ibm.com/software/expert/index.html>

For a free demonstration of the IBM TotalStorage Expert, go to:

<http://www.storwatch.com/>

For installation and use of the IBM TotalStorage Expert, refer to *IBM StorWatch Expert Hands-On Usage Guide*, SG24-6102.



Installation planning and migration

The initial installation and implementation of the IBM TotalStorage Enterprise Storage Server, as well as the migration from other storage solutions, require planning.

This chapter presents an introductory overview of the necessary planning tasks that must be done for the initial installation of the IBM TotalStorage Enterprise Storage Server Model 800. For more a detailed discussion and information when planning an ESS installation, refer to the publication *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444.

9.1 Overview

When preparing for the installation of the IBM TotalStorage Enterprise Storage Server, for an efficient start the planning should begin some time well before the box arrives in the computer room. There are three major areas involved in installation planning:

1. Physical planning
 - Selection of appropriate ESS configuration (hardware components and features)
 - Site requirements, dimensions, weight
 - Site requirements power, additional equipment
2. Configuration planning
 - Attached hosts, CKD, FB or a mix of CKD and FB
 - Connection via ESCON, FICON, SCSI or Fibre Channel
 - Storage configuration, RAID 5 or RAID 10
3. Migration planning
 - Migrating data for FB data
 - Migrating data for CKD data

The installation planning of the ESS is further discussed in detail in the publication *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444.

9.2 Physical planning

This part of the planning addresses all the physical considerations, from the hardware configuration of the machine to be installed to the site preparation where it will be operated.

9.2.1 Hardware configuration

When configuring the IBM TotalStorage Enterprise Storage Server Model 800, several hardware components and optional features can be selected as illustrated in Figure 9-1 on page 247:

- ▶ Total capacity of the ESS
- ▶ Type (capacity/speed) of the disk drives
- ▶ Quantity and type of host adapters
- ▶ Cache size
- ▶ Advanced functions (PAV, PPRC, FlashCopy, XRC)
- ▶ Step Ahead capacity

These options make up the hardware configuration of the ESS that is going to be installed. They will be selected based on the requirements and characteristics of the workload that is going to be driven onto the ESS — current, and future workload projections that may be convenient to consider — as well as the number and type of servers to be connected and the attachments they will be using.

This hardware configuration planning will determine also if the Expansion Enclosure will be part of the configuration, thus impacting the space and power requirements that may be needed.

When determining the number and type of attachments, it will be a matter of path availability as well as performance — more paths, more parallelism for the applications' I/O operations. Also the cache size will be a matter of performance consideration. The IBM Storage Specialist using the Disk Magic modelling tool can assist in this activity.

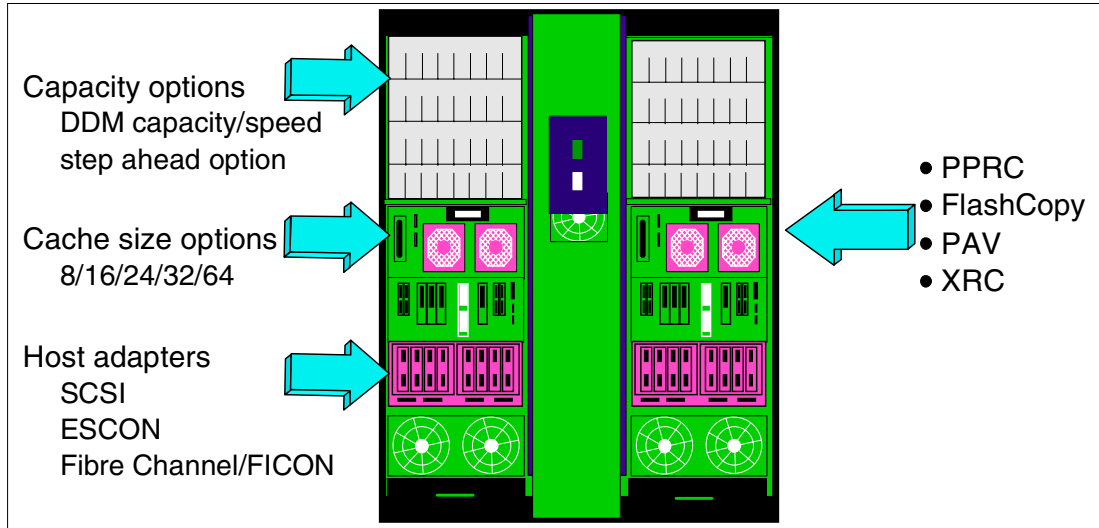


Figure 9-1 Several options - hardware components and advanced functions

9.2.2 Site requirements - dimensions and weight

The room where the ESS is going to be installed must meet some requirements. These site physical requirements must be checked well ahead of the arrival of the machine, in case some work needs to be done.

- ▶ The floor must tolerate the additional weight of the new ESS to be installed (also the future machine upgrades should be considered). The ESS can be installed on a non-raised floor, but it is recommended that a raised floor be used for increased air circulation and cooling, as well as for a neat and safe housing for the connecting cables (power, host attachment, and network).
- ▶ Enough clearance around the box must be left for cooling. Fans take in air through the front and exhaust it to the rear and top.
- ▶ Enough service and clearance space must be left. Space must be available for full opening of the ESS doors, front and back.

Information on weight and dimensions of the ESS can be obtained from several sources. One such source is the publication *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444.

9.2.3 Site requirements - power and others

Besides the room weight and space readiness, some other features are required for the ESS operation, including:

- ▶ Power cable connections.
The ESS requires two independent power supplies connected to separated power distributions.
- ▶ Power outlets for following components:
 - ESS Master Console
 - ESS Master Console monitor
 - Hub
 - Modem
- ▶ An analog telephone line for ESS remote support.

For detailed information and requirements, refer to *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444.

9.3 Configuration planning

Before the IBM SSR is able to physical install and configure the ESS, the configuration planning must be completed. The IBM SSR will need the following completed worksheets:

- ▶ Communication Resources worksheet
- ▶ Communication Resources worksheet for the ESS Master Console
- ▶ Communication Resources worksheet for ESS Copy Services (if required)
- ▶ Configuration worksheets

The first three work sheets, as well as instructions to complete them, can be found in the publication *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444. The configuration worksheets as well as instructions and examples on how to fill them in can be found in the publications *IBM TotalStorage Enterprise Storage Server Configuration Planner for Open-Systems Hosts*, SC26-7477 and *IBM TotalStorage Enterprise Storage Server Configuration Planner for S/390 and zSeries Hosts*, SC26-7476.

9.4 Installation and configuration

After thoughtful planning and checking of the site requirements, and after the needed actions to meet the full readiness conditions, then the actual installation and configuration of the machine takes place. Once the site meets the requirements, and having the configuration worksheets completed, then the installation of the ESS and its initial configuration is done by the IBM SSR based on the information provided by the user.

9.4.1 Physical installation

The physical installation of the ESS is done by the IBM SSR. The main activities involved during this phase of the installation are:

1. Unpack and locate the machine at the designated place. Complete the box setup: attach covers, doors and remaining shipped parts.
2. Connect power, do first power-on, and then the required checks.
3. Install and set up the ESS Master Console.
4. Complete machine checkout.
5. Set up machine for remote connection, network communication, and features according to information in the provided worksheets.

After these steps are completed by the IBM SSR, then the configuration takes place.

9.4.2 Configuration

Once the physical installation of the ESS is completed, it is now ready to be configured and attached to the hosts.

This section briefly describes the steps involved in the logical configuration process. The logical configuration process is discussed in detail in 4.16, "ESS logical configuration" on page 111.

To configure the ESS, the ESS Master Console can be used. It comes with a Web browser to interface with the ESS Specialist. If a different workstation is going to be used, then it must have one of the supported Web browsers, and it must be connected to the ESSnet so it has access to the ESS Specialist interface. The ESS TCP/IP host name of one of the ESS clusters must be used.

1. If one of the standard logical configurations is selected, then most of the steps of the logical configuration will be done by the IBM SSR using the Batch Configuration Tool — the SCSI and Fibre Channel ports will have to be configured later using the ESS Specialist.
2. If custom volumes have been planned, then these will be configured using the ESS Specialist.
3. If none of the standard logical configurations have been chosen, then the full logical configuration of the ESS is done using the ESS Specialist. This includes:
 - Formatting of CKD (zSeries) and FB (open systems) ranks
 - Definition of volume emulation type or LUN size
 - Rank type, either RAID 5 or RAID 10
 - Custom volumes
4. If using Parallel Access Volumes for zSeries volumes, then the alias addresses must be defined for the base volumes. Aliases are defined in the ESS as well as in the host HCD.
5. For zSeries systems, the ESS logical configuration must match that entered in HCD when creating the IODF — logical control units, base and alias addresses for PAV volumes.
6. With the host software prepared for the ESS attachment, the ESS is connected to the hosts.
7. Then some final tasks are needed to make the logical volumes usable for the host operating system:
 - The CKD volumes are formatted with the ICKDSF utility. Only a minimal INIT is required.
 - For UNIX systems, the ESS logical volumes (logical disks) can be added to logical volume groups and logical volumes in UNIX to create file systems on them.
 - In Windows NT, the volumes are assigned drive letters.

9.5 Migration

Migrating data to the IBM TotalStorage Enterprise Storage Server can be done using standard host utilities. Each operating system sees the logical volumes in an ESS as normal logical devices. For the CKD type of servers, this is an IBM 3390 or 3380. For the iSeries servers, these are 9337 or 2105. For UNIX systems, the logical disks are recognized as SCSI or Fibre Channel-attached drives (hdisk in AIX, for example), and Windows also sees them as SCSI or Fibre Channel-attached disks.

Data migration requires careful planning. The major steps are:

- ▶ Education (including documentation)
- ▶ Software readiness
- ▶ Configuration planning
- ▶ Data migration planning
- ▶ Hardware and software preparation
- ▶ Data migration

When planning for migration, additional information can be found in the publications *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444 and *IBM TotalStorage Enterprise Storage Server User's Guide*, SC26-7445.

Defining and connecting the system

Before migrating data to an ESS, the ESS must first be configured and connected to the host system(s). Configuration of the ESS is covered in Chapter 4, "Configuration" on page 93.

9.6 Data migration in z/OS environments

- Upgrade system software to support ESS
- Configure LCUs and define CKD volumes with ESS Specialist
 - Consider the use of custom volumes
 - Define *alias* addresses (optional)
- Make HCD definitions
- ICKDSF Minimal INIT (for *base* volumes only)
 - ESS does not emulate Alternate Tracks
 - Optional: ICKDSF REFORMAT REFVTOC
- Set Missing Interrupt time to 30 sec for *base* volumes
- Choose migration method and migrate volumes
 - IBM Migration Service with TDMF
 - COPY or DUMP/RESTORE
 - XRC/PPRC

Figure 9-2 Preparation for zSeries data migration

Figure 9-2 summarizes the main considerations and activities involved in migrating data to an ESS. These activities are discussed in the following section.

Software support

This activity was already done when initially planning for the ESS installation. Helpful recommendations and guidelines for doing this activity can be found in 8.2, "zSeries environment support" on page 230.

Configuring LCUs and CKD volumes

When configuring the CKD logical control units and volumes using the ESS Specialist, there are some considerations.

Custom volumes

The ESS allows the definition of custom volumes of any size from 1 to 32,760 cylinders. Having small logical volumes can drastically reduce contention for a volume, particularly when several data sets with high activity reside on the same volume. Each highly active data set can be placed on a separate *custom volume* without wasting a lot of space, and hence

avoid data set contention. Alternatively, large custom volumes are easier to administer, and when combined with PAV the performance impact for the potential increased IOSQ time will be neutralized.

Before migrating volumes one-to-one to the ESS, candidate data sets for custom volumes should be considered. Having identified such data sets, the size of the custom volumes can then be determined.

Aliases

If the ESS has the optional PAV feature installed, this gives the possibility of defining multiple aliases for each base volume. As already mentioned, this implementation allows you to use even the larger volume sizes, thus simplifying the administration while avoiding any performance penalty from IOSQ contention.

From the current RMF reports of the volumes to be migrated, the Avg IOSQ times can be checked to see if they are high. If this is the case, then Parallel Access Volumes (PAVs) should be used for better performance. If PAVs are going to be used, then planning for the alias addresses must be done in order to define them.

HCD definitions

To access the logical volumes on an ESS from a zSeries system, an IODF is needed that includes the logical CKD subsystems of the ESS. This is done with the HCD dialog. These definitions and the corresponding ESS logical configuration definitions must match.

Volumes initialization

After the volume configuration in the ESS and in the HCD is done, the logical volumes must be initialized with ICKDSF. Only a Minimal INIT is required.

The ESS does not emulate Alternate Tracks, Device Support Tracks, or Service Tracks. This is similar to the implementation on IBM 3990 Model 6 with RAMAC models. The IBM RVA did emulate these Alternate, Device Support, and Service tracks. This sometimes caused some problems when migrating volumes one-to-one from one storage subsystem to another, when back-level system software was used that did not update the Volume Table of Contents (VTOC) to reflect the correct number of tracks. It is always a good idea to refresh the VTOC after a volume migration with the ICKDSF REFORMAT REFVTOC command. This refresh sets the number of tracks to the correct value.

TDMF is one of the programs that does not manage the alternate track difference; for this reason, a REFORMAT REFVTOC will be needed after migrating to an ESS if the source volume was an IBM RVA or a non-IBM subsystem.

For the base volumes on the ESS, the missing interrupt time should be set to 30 seconds.

Migration methods

There are several ways to do the data migration. Depending on the specific requirements of the processing environment, any of these methods may be more convenient. Further information on data migration can be found in the redbook *IBM TotalStorage Enterprise Storage Server: Implementing the ESS in Your Environment*, SG24-5420-01.

IBM migration services

In several countries IBM offers a migration service for data to be migrated from any previous S/390 storage subsystem to an ESS. IBM uses the Transparent Data Migration Facility (TDMF) tool to do the data migration, which enables data to be migrated while the normal production work continues. When all data is copied to the ESS, the systems can be restarted using the new volumes on the ESS.

Copy, and Dump/Restore

The classic approach is to dump all source volumes to tape and restore them to ESS volumes. This is the slowest approach and requires the application systems to be down during the dump/restore process. The advantage of this approach is that there is no need to attach both storage subsystems (the old one and the ESS) at the same time.

A much faster migration method is to do a volume copy from the old volumes to ESS volumes using, for example, the DFSMSdss program. This migration method also requires that both storage subsystems are online to the system that does the migration. Application systems must be down during the migration process.

While DFDS is by far the simplest way to move most data, the following alternatives are also available:

- ▶ IDCAMS EXPORT/IMPORT (VSAM)
- ▶ IDCAMS REPRO (VSAM, SAM, BDAM)
- ▶ IEBCOPY (PDSs - including load module libraries - and PDSEs)
- ▶ ICEGENER (SAM) - part of DFSORT
- ▶ IEBGENER (SAM)
- ▶ Specialized database utilities (for example, for CICS, DB2, or IMS)

Migrating data with XRC

If the data to be migrated currently resides on a IBM 3990 Model 6 or an IBM 9390 storage controller in a z/OS environment, then XRC can be used to migrate the volumes to ESS. This is a very convenient way to do data migration. The application systems can continue to run while the data is migrated. Once old and new volumes are synchronized, the systems can be shut down and then restarted using the new volumes on the ESS.

While a special enabling feature is required for XRC in ESS if it is going to be used as a primary control unit, this feature is not required when the ESS is a secondary control unit as in a migration scenario.

Migrating volumes with XRC is quite easy. The State Data Set must be allocated, for example hlq.XCOPY.session_id.STATE., and you must solve RACF demands. The use of XRC commands such as XSTART, XEND, XADDPAIR, and XRECOVER must be allowed. The Data Mover task ANTAS001 must be allowed by RACF to read from the source volumes, and it needs update authority to the State Data Set.

The system where the System Data Mover task runs needs access to both source and target storage control units. The target volumes must be online to the System Data Mover system. The volumes can have any *volser*.

The XRC session can be started with the command:

```
XSTART session_ID ERRORLEVEL VOLUME SESSIOntype(MIGRATE) HLQ(hlq)
```

Any name can be chosen to be used for session_ID, but it must match the session_ID in the State Data Set. Now the pairs to be synchronized can be added with the command:

```
XADDPAIR session_ID VOLUME(source target)
```

After all pairs are synchronized, this can be checked with the XQUERY command. Now a time must be chosen when the application systems can be shut down to do the switch to the new volumes. After the application systems have been stopped, this command sequence can be done:

```
XEND session_ID  
XRECOVER session_ID
```

The XRECOVER command will re-label the target volumes with the source volume's *volser*. If the source volumes are still online to the System Data Mover system, the target volumes will go offline. Now the systems can be restarted using the new volumes.

Detailed information on XRC use can be found in the publications *z/OS DFSMS Advanced Copy Services*, SC35-0428 and *DFSMS/MVS Remote Copy Guide and Reference*, SC35-0169.

9.7 Migrating data in z/VM

When migrating data to the ESS in a z/VM system, data migration can be greatly simplified by using DFSMS/VM, which is a no-charge feature of z/VM. This allows you to manage and control the use of VM disk space. It will also move minidisks from one media to another. DFSMS/VM includes such services as a data mover, an automated move process, and an interactive user interface.

Using DIRMAINT also provides tools that will manage the movement of CMS minidisks from one media to another.

DASD Dump Restore (DDR) is a service utility shipped with VM that can be used to dump data from disk to tape, restore data from tape to disk, and copy data between like disk drive volumes. DDR cannot be used to copy data between disk devices with different track formats.

CMDISK is a DIRMAINT command that can be used to move minidisks from any device type supported by VM to any other type.

COPYFILE is a CMS command that can be used to copy files between CMS formatted minidisks on devices with the same or different track formats.

SPXTAPE is a CP command that can be used to dump spool files to tape and to reload them from tape to disk.

9.8 Migrating data in VSE/ESA

Several dialogs in the VSE Interactive Interface can be used to set up the jobs to move data to the ESS. Data can be recognized and space fragmentation can be eliminated by using the backup/restore technique. The following Backup/Restore dialogs can be used:

- ▶ Export and import VSAM files
- ▶ Back up and restore VSAM files
- ▶ Back up and restore ICCF libraries
- ▶ Back up and restore the system history file
- ▶ Back up and restore the system residence library
- ▶ Create a loadable tape with the system residence library and system history file ready to restore

The following facilities can also be used:

- ▶ VSE/FASTCOPY can be used to move volumes and files between devices with identical track formats.
- ▶ VSE/DITTO can also be used to copy files.
- ▶ VSE/POWER commands can be used to transfer the SPOOL queue from one device to another.

- ▶ VSE/VSAM can move any VSAM data set using either REPRO or EXPORT/IMPORT functions.

9.9 Data migration in UNIX environment

No special tools or methods are required for moving data to IBM TotalStorage Enterprise Storage Server disks. The migration of data is done using standard host operating system commands. The UNIX hosts see the ESS logical devices (or logical volumes) just like normal physical SCSI disks.

Before putting any data on an ESS, first the logical FB volumes must be defined and host access configured. Configuration of the ESS is covered in Chapter 4, "Configuration" on page 93.

9.9.1 Migration methods

Logical Volume Manager software

- Most UNIX systems have Logical Volume Management software
 - AIX, HP-UX, Solstice for Solaris, Veritas VxVM
- AIX's **migratepv** command migrates a complete physical disk
- AIX's **cp1v -p** command copies logical volumes
- AIX's **mk1vcopy** command sets up a mirror; **splitvcopy** splits the mirror - can be used for data migration

Figure 9-3 UNIX data migration methods

For UNIX hosts, there are a number of methods of copying or moving data from one disk to another (see *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444 for a detailed discussion on the migration options). In this section we discuss some of the most common migration methods used.

Volume management software

Most UNIX systems provide specific tools for the movement of large amounts of data. These tools can directly control the disks attached to the system. AIX's Logical Volume Manager (LVM) is an example of such a tool. Logical Volume Management software is available for most of the UNIX systems, such as HP-UX, Solstice from Sun Microsystems for Solaris, and Veritas Volume Manager (VxVM) for Solaris. The LVM provides another layer of storage. It provides logical volumes that consist of physical partitions spread over several physical disks.

The AIX LVM provides a **migratepv** command to migrate complete physical volume data from one disk to another.

The AIX LVM also provides a command (**cp1v**) to migrate logical volumes to new logical volumes, created on an ESS, for example. Do not be confused by the term logical volume as it is used in UNIX and the term logical volume used in the ESS documentation for a logical disk, which is actually seen by the UNIX operating system as a physical disk.

One of the facilities of the AIX LVM is RAID 1 data mirroring in software. This facilitates data movement to new disks. The **mk1vcopy** command can be used to set up a mirror of the whole

logical volume onto another logical volume, defined on logical disks (we prefer this term here instead of logical volume) on an ESS. Once the synchronization of the copy is complete, the mirror can be split up by the `splitvcopy` command.

Standard UNIX commands for data migration

If a Logical Volume Manager is not available, then the standard UNIX commands to copy or migrate your data onto an ESS can be used.

Direct copy can be done with the `cpio -p` command. The `cpio` command is used for archiving and copying data. The `-p` option allows data to be copied between file systems without the creation of an intermediate archive. For a copy operation, the host must have access to the old disks and the new disks on an ESS. This procedure does not require application downtime.

The `backup` (in AIX) or `dump` (on other UNIX systems) and `restore` commands are commonly used to archive and restore data. They do not support a direct disk-to-disk copy operation, but require an intermediate device such as a tape drive or a spare disk to hold the archive created by the backup command.

There are other UNIX commands, such as the `tar` command, that also provide archival facilities that can be used for data migration. These commands require an intermediate device to hold the archive before you can restore it onto an ESS.

For more information about the use of these commands, see the publication *AIX Storage Management*, GG24-4484.

9.10 Migrating from SCSI to Fibre Channel

The early IBM TotalStorage Enterprise Storage Servers were shipped with either SCSI or ESCON host adapters only. This was before native Fibre Channel attachment became available. If your installation still has data on the ESS that is accessed via direct SCSI attachment, and is currently planning to swap to Fibre Channel attachment, this section will help you start planning.

There are many possible ESS environments, and each environment requires a different detailed procedure for migration. This section presents the general procedure for migrating from SCSI to Fibre Channel. In general, migration from SCSI to Fibre Channel is a disruptive procedure.

If you are planning to migrate from SCSI to Fibre Channel, then the following material should be referred for detailed information:

- ▶ The white paper, “ESS Fibre Channel Scenarios”, which can be obtained from:
<http://www.storage.ibm.com/hardsoft/products/ess/support/essfcmig.pdf>
- ▶ *Implementing Fibre Channel Attachment on the ESS*, SG24-6113

Outlined below are the general steps for migrating from SCSI to Fibre Channel. Before starting the migration, a current backup of the data must be done.

1. First of all, both host and ESS must be ready for Fibre Channel attachment. The host software must be checked for adequate support of Fibre Channel adapters.
2. The Fibre Channel adapters must be installed on the ESS. Also if not yet installed, then the Fibre Channel adapters on the host(s) must also be installed and have the adapters recognized by the host(s).

3. All the operating system tasks to free the disks must be performed. A sample of this procedure is to unmount the file, vary off the disks, and export the volume group.
4. Using the ESS Specialist, the disks should be unassigned. This makes the disks inaccessible to the server. However, the disks are still configured inside the ESS and the data on the disks is still intact.
5. Using the ESS Specialist, the Fibre Channel adapters must be configured using the actual worldwide port name (WWPN) from the host's Fibre Channel adapters.
6. If multiple Fibre Channel hosts adapters are being connected to the ESS, the second and subsequent connections must be added as *new* Fibre Channel hosts using the actual worldwide port name from each of the host's Fibre Channel adapters. For this situation, you should consider installing in the application server the Subsystem Device Driver (SDD) program that comes with the ESS, which allows for the handling of shared logical volumes (see 5.8, "Subsystem Device Driver" on page 157 for more information).
7. Using the ESS Specialist, the volumes must be assigned to the *new* hosts.
8. The server is restarted and the volumes are made available to the server. This can be done by mounting the files.

9.11 Migrating from ESCON to FICON

Migration to FICON requires that sufficient Fibre Channel/FICON Host Adapters be installed in the ESS. These adapters must be configured to operate in FICON mode (see 4.19, "ESCON and FICON host adapters" on page 115 for how to do this), and be connected to FICON ports in the zSeries or S/390 processor.

FICON host adapters can be added to an ESS nondisruptively, provided sufficient spare slots are available in the host bays. If there are insufficient free slots available, then other adapters will have to be removed to make space.

Remember that you should always aim to distribute adapter cards evenly across the four host adapter bays of the ESS in order to maximize availability.

Note: Although ESCON and FICON paths can co-exist in the same path group (that is, the set of paths to one operating system image) on the ESS, the intermixed path group configuration is only supported for the duration of the migration. The ESS fully supports an intermixed configuration of ESCON and FICON host adapters. The limitation applies only to the intermix of ESCON and FICON paths in a path group, that is, intermixed paths to the same operating system image.

A sample ESCON configuration is shown in Figure 9-4 on page 257. Each image on each of the four CECs has eight ESCON channel paths to each ESS *logical control unit* (LCU). Four paths are cabled through each of the two ESCON directors.

- ▶ There are eight LCUs configured in the ESS.
- ▶ Each LCU has 256 devices defined, so only four LCUs are accessible per ESCON channel ($4 \times 256 = 1024$ — the maximum number of devices supported by ESCON).
- ▶ Hence the LCUs are partitioned into two groups, with each set of four LCUs addressable through eight of the 16 ESCON host adapters.
- ▶ The two sets of eight ESCON paths from the ESS are spread across the two ESCON directors.

- ▶ Each CEC has eight paths to each ESCON director, giving a total of 16 paths to the ESS. The CHPIDs are configured in the IOCP to give eight paths to each set of four LCUs.
- ▶ The paths are shared across the three images in each CEC.

This results in the maximum of eight paths per path group to all LCUs.

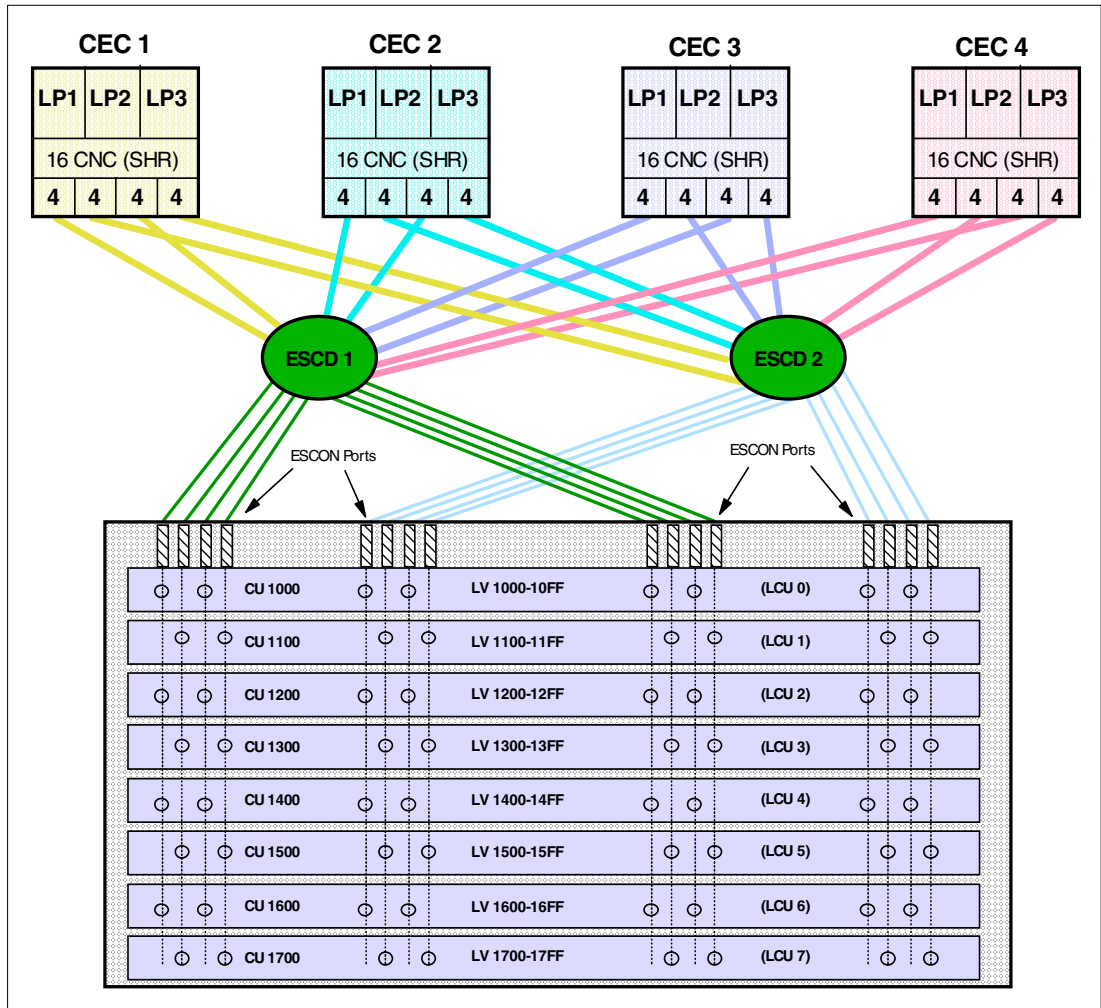


Figure 9-4 ESS configuration with ESCON adapters

If planning to install four FICON channels per CEC and four FICON host adapters on the ESS, then the interim configuration is shown in Figure 9-5 on page 258. In the interim configuration, the same images can access the ESS devices over a combination of both ESCON and FICON paths. This configuration is only supported for the duration of the migration to FICON.

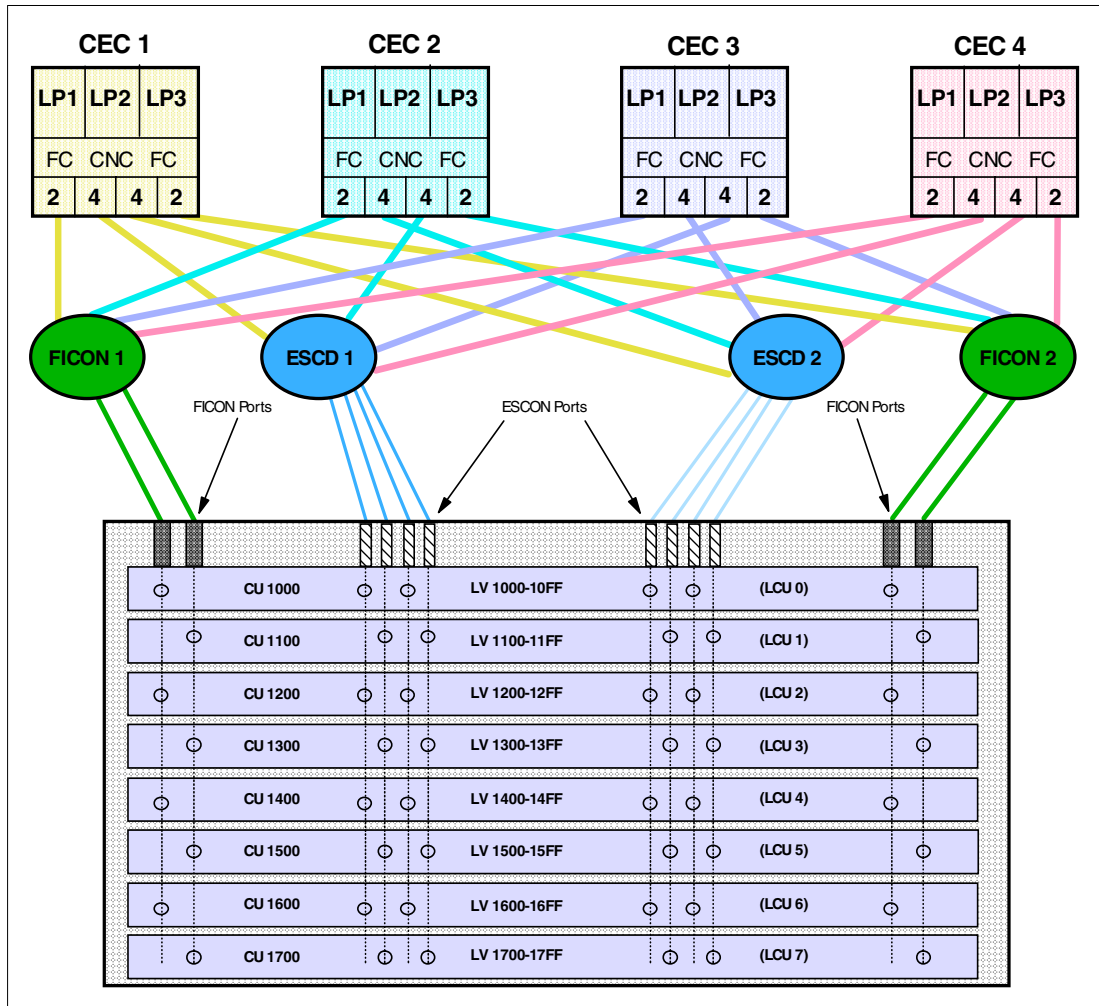


Figure 9-5 Interim configuration with ESCON and FICON intermix

The following steps must be followed:

- ▶ The software must be upgraded to levels that support FICON (see 8.2.6, “FICON support” on page 231).
- ▶ The FICON channels must be installed on the CECs. This is a nondisruptive procedure.
- ▶ The FICON host adapters must be installed on the ESS.
- ▶ The FICON directors must be installed. Note that the FICON channels could be connected point-to-point to the ESS, but this would require more ESS adapters.
- ▶ Half of the ESCON channel paths should be varied off.
- ▶ A dynamic I/O reconfiguration change must be performed to remove half of the ESCON channel paths and add the FICON channel paths to the ESS control unit definitions, while keeping the devices and ESCON paths online.

Note: At this stage the eight LCUs are still addressed as two groups of four through the FICON paths because the path group is shared with ESCON paths.

The final configuration is shown in Figure 9-6 on page 259.

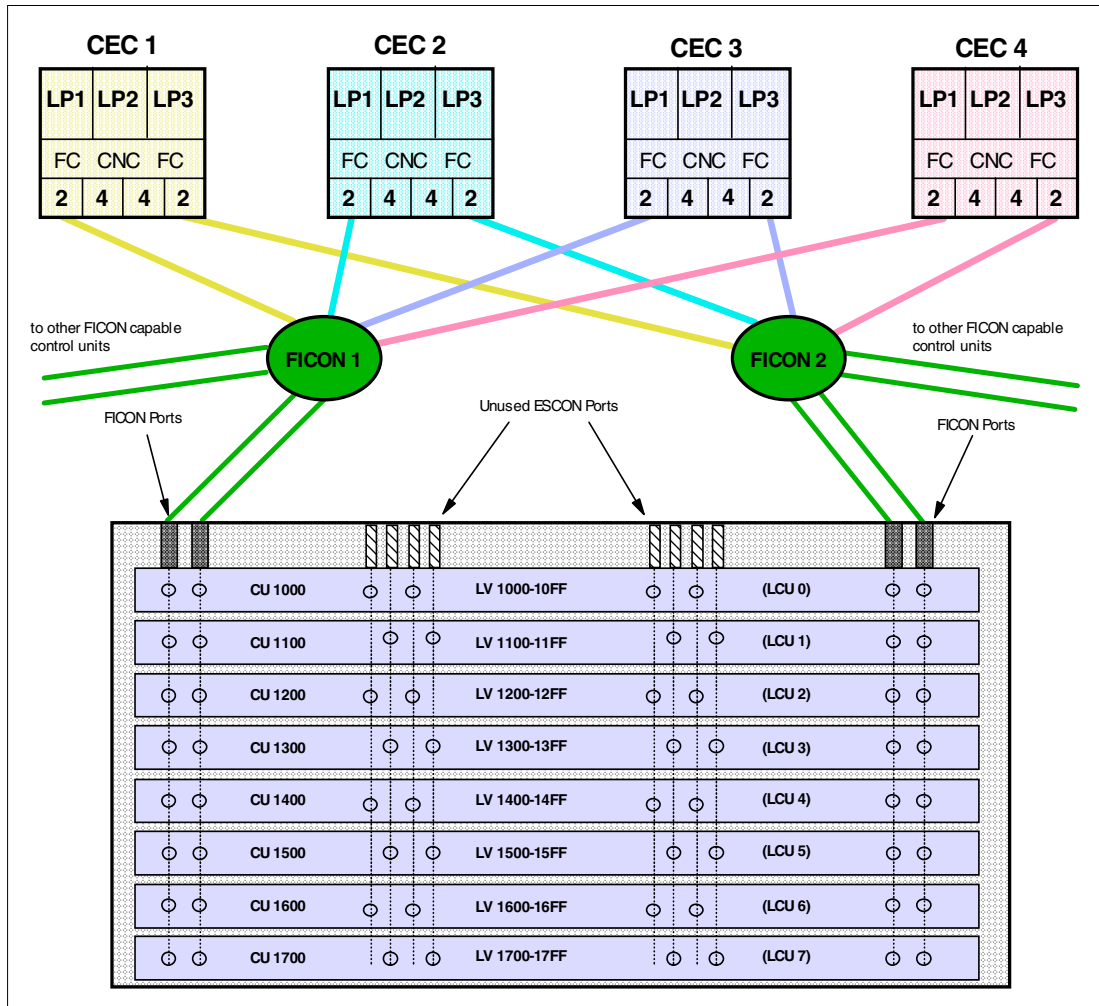


Figure 9-6 Target FICON ESS migration configuration

All eight LCUs are now addressed via all FICON channel paths, so all 2048 devices are addressable via all channels. As a final step in migrating from ESCON to FICON, another two pairs of FICON host adapters could be installed in the ESS to bring the total to eight. These would then also be connected to the two FICON directors.

Attention: Intermixed ESCON and FICON channel paths in the same path group are only supported to facilitate a nondisruptive migration, and should not be used for any extended length of time. Reconnects for devices in an intermixed configuration are not optimal from a performance perspective, and it is strongly recommended to move from an intermixed configuration as soon as possible.

9.12 IBM migration services

In several countries, IBM offers migration services for different environments. Check with your IBM Sales Representative, or contact your local IBM Global Services (IGS) representatives for additional information. You may also get information at:

<http://www.ibm.com/ibmlink>

Select your geographic area and then look for ServOffer under InfoLink.

9.12.1 Enhanced Migration Services - Piper

Piper is a hardware and software solution to accomplish the movement of system and application data from current disk storage media to the IBM TotalStorage Enterprise Storage Server. Piper can be used in CKD and FB environment for hardware-assisted data movement:

- ▶ Independent of the production host
- ▶ Copy workload moved to a hardware tool
- ▶ Minimizes the demand on application host MIPS
- ▶ TDMF or FDRPAS data mover used
- ▶ Rate of data movement can be tuned to maximize application performance during the copy
- ▶ Data migration concurrent or nonconcurrent

CKD environment

In a CKD environment, the migration is done using a Portable Rack Enclosure containing:

- ▶ 390 Multiprise
- ▶ ESCON director
- ▶ Preloaded TDMF or FDRPAS migration software
- ▶ Nondisruptive power configuration

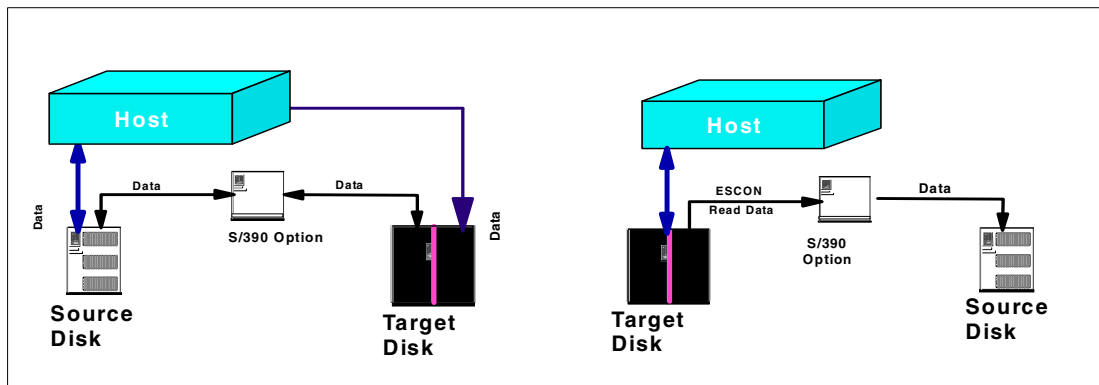


Figure 9-7 Piper CKD migration concurrent or nonconcurrent

FB environment

In a FB environment the migration is done by use of a Portable Rack Enclosure containing:

- ▶ Vicom SLIC routers
- ▶ McDATA SAN Switch
- ▶ Nondisruptive power configuration

To migrate data in an FB environment, there is no additional software required for the host or for the tool.

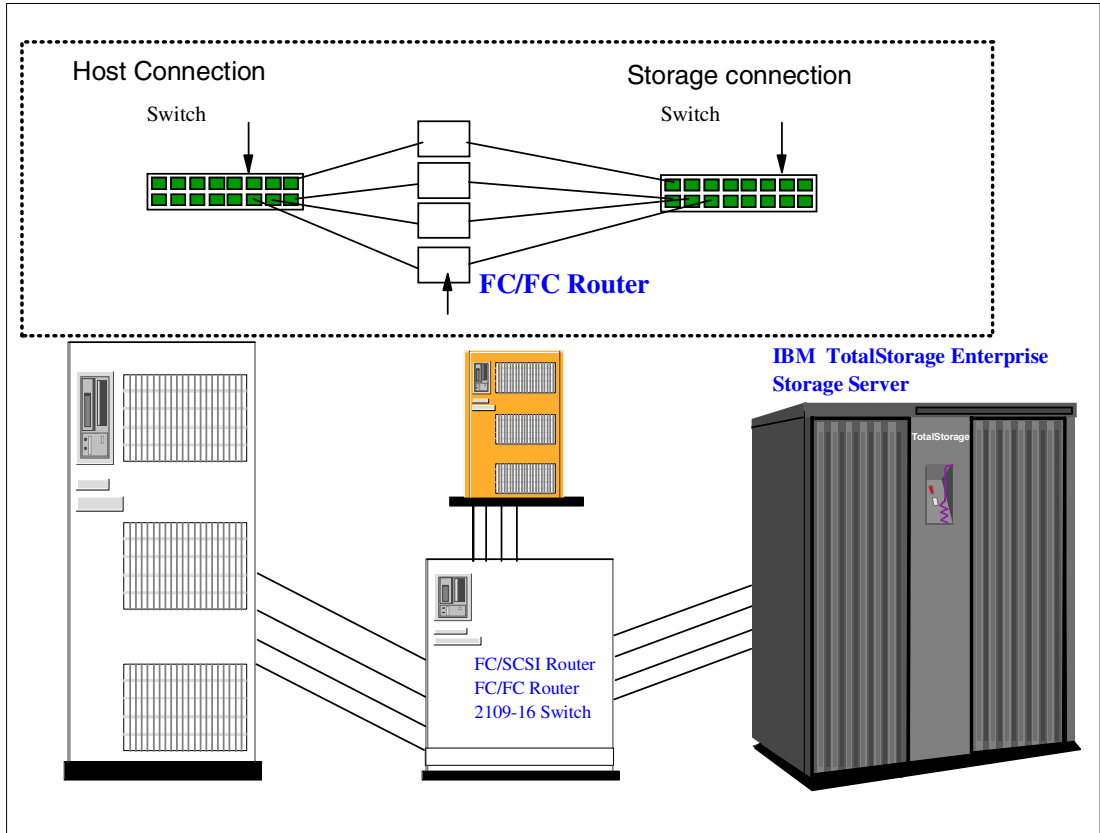


Figure 9-8 Piper FB migration concurrent or nonconcurrent



A

Feature codes

This part of the book summarizes and gives information on the most relevant features to be considered when doing the hardware configuration of the ESS Model 800 and its available options.

Features are described for technical illustration purposes, to help when dealing with the physical configuration of the ESS. Not all the ESS Model 800 features are listed in this section. The complete list of features can be found in the IBM TotalStorage Enterprise Storage Server Model 800 hardware announcement letter 102-201.

A.1 Overview

The IBM TotalStorage Enterprise Storage Server Model 800 provides a wide range of configurations scaling from 582 GB to 27.9 TB, with many standard and optional features that must be specified when ordering the hardware configuration.

There are many options that must be specified, such as the cluster processors, cache size, number of disk eight-packs and the speed (RPM) and capacity of the disk drives, number and type of host adapters, advanced functions, power supply specifications, modem country group, cables, etc. This allows you to configure the ESS according to your specific workload and environment requirements.

All these options, either with a charge or available at no additional cost, are identified by a corresponding feature number on the ESS Model 800 whose machine type and model designation is 2105-800.

Important: When reading the information in this section, these considerations apply:

- ▶ Features are described in this section for technical illustration purposes. Information about pre-requisites and co-requisites among the features is not always included.
- ▶ The descriptions do not include information on whether the features are priced or available at no additional cost.
- ▶ Only the more relevant feature codes at the time of doing the hardware configuration are summarized (not all administrative features are listed in this section).
- ▶ For current detailed information, refer to the announcement letter for the IBM TotalStorage Enterprise Storage Server Model 800.

A.2 Major feature codes

When doing the hardware configuration of the ESS Model 800, the clusters processors, the cache size, and the type and quantity of host adapters need to be specified.

A.2.1 Processors

These features designate the type of processors installed in the ESS Model 800. The default configuration is the Standard processors (fc 3604). The Standard processors can be field upgraded to the Turbo processors.

Table A-1 ESS processor options

Processor option	Model 800 feature
Standard	3604
Turbo	3606

A.2.2 Cache sizes

The ESS can be configured with five different cache capacities, according to Table A-2 on page 265. These features are used to designate the total amount of cache installed in the ESS. The default configuration is 8 GB cache (fc 4012). The installed cache can be upgraded in the field to any larger capacity.

Table A-2 ESS cache capacities (Gigabytes)

Capacity	Feature
8 GB	4012
16 GB	4014
24 GB	4015
32 GB	4016
64 GB	4020

A.2.3 Host adapters

The ESS supports the intermix of FICON, ESCON, SCSI, and Fibre Channel host attachments, making the ESS the natural fit for server consolidation requirements. Table A-3 shows the feature codes for the different host adapters that can be ordered with the ESS Model 800. The ESS can be configured with up to 16 of these host adapters.

For the highest availability, it is recommended that adapter cards (of the same type) be installed in pairs and into different host adapter bays. See 2.10, “Host adapters” on page 35.

Table A-3 ESS host adapters

Host Adapter	Feature	Number of ports
2 Gb Fibre Channel/FICON (long wave)	3024	1
2 Gb Fibre Channel/FICON (short wave)	3025	1
Enhanced ESCON	3012	2
SCSI	3002	2

A.2.3.1 2 Gb Fibre Channel/FICON (long wave) host adapter (3024)

Each Fibre Channel/FICON (long wave) host adapter provides one port with an LC type connector. The interface supports 200 MBps and 100 MBps, full-duplex data transfer over long-wave fiber links. The adapter supports the SCSI-FCP ULP (Upper Layer Protocol) on point-to-point, fabric, and arbitrated loop (private loop) topologies, and the FICON ULP on point-to-point and fabric topologies. SCSI-FCP and FICON are not supported simultaneously on an adapter.

- ▶ Minimum number of this feature per ESS: none.
- ▶ Maximum number of this feature per ESS: 16.
- ▶ No cable order is required, since one single mode cable is included with each feature, as specified by a cable feature code 9751 - 9753 (see Table A-4 on page 266).

A.2.3.2 2 Gb Fibre Channel/FICON (short wave) host adapter (3025)

Each Fibre Channel/FICON (short wave) host adapter provides one port with an LC type connector. The interface supports 200 MBps and 100 MBps, full-duplex data transfer over short-wave fiber links. The adapter supports the SCSI-FCP ULP (Upper Layer Protocol) on point-to-point, fabric, and arbitrated loop (private loop) topologies, and the FICON ULP on point-to-point and fabric topologies. SCSI-FCP and FICON are not supported simultaneously on an adapter.

- ▶ Minimum number of this feature per ESS: none.
- ▶ Maximum number of this feature per ESS: 16.

- ▶ No cable order is required, since one multimode cable is included with each feature, as specified by a cable feature code 9761 - 9763 (see Table A-4).

A.2.3.3 Fibre Channel/FICON host attachment cables

Cables are required to attach the ESS host adapters to servers and fabric components. Table A-4 shows the Fibre Channel/FICON host attachment cables features.

Features 975x-976x are used to specify the cables shipped with the ESS. One feature must be specified for each Fibre Channel/FICON host adapter feature. The additional cables can be ordered using the 285x-286x features.

The 2 meter cables (9753, 2853, 9763, 2863) with female SC connector should be used only for attaching the 2 Gb Fibre Channel/FICON host adapters (fc 3024 and 3025) of the ESS Model 800 to SC/SC connector cables previously used with older types of Fibre Channel/FICON host adapters.

Table A-4 2 Gb Fibre Channel/FICON cable feature codes

Description	Feature	Additional cables
9 Micron, 31 meter, single mode, SC/LC connectors	9751	2851
9 Micron, 31 meter, single mode, LC/LC connectors	9752	2852
9 Micron, 2 meter, single mode, SC/LC connectors	9753	2853
50 Micron, 31 meter, multimode, SC/LC connectors	9761	2861
50 Micron, 31 meter, multimode, LC/LC connectors	9762	2862
50 Micron, 2 meter, multimode, SC/LC connectors	9763	2863

A.2.3.4 Enhanced ESCON host adapter (3012)

This feature provides one Enhanced ESCON host adapter for ESS attachment to ESCON channels on zSeries and S/390 servers or ESCON directors, and for PPRC connection with another ESS. The adapter supports two ESCON links, with each link supporting 64 logical paths. The ESCON attachment uses an LED-type interface. ESCON cables must be ordered separately.

- ▶ Minimum number of this feature per ESS: none.
- ▶ Maximum number of this feature per ESS: 16.

A.2.3.5 SCSI host adapter (3002)

This feature provides one SCSI host adapter for ESS attachment to SCSI servers. The adapter has two ports. This dual-port Ultra SCSI host adapter supports the Ultra SCSI protocol (SCSI-2 Fast/Wide differential is a subset of Ultra SCSI and is therefore supported as well). If the server has SCSI-2 (fast/wide differential) adapters and/or Ultra SCSI adapters, they can attach to the ESS.

- ▶ Minimum number of this feature per ESS: none.
- ▶ Maximum number of this feature per ESS: 16.

A.2.3.6 SCSI host attachment cables

Cables are required to attach the ESS host adapters to servers and fabric components. Table A-5 on page 267 lists the feature codes for the different SCSI cables that can be ordered for the ESS.

Two SCSI host attachment cables are required for each SCSI host adapter. When configuring the ESS you specify the feature codes 9701 to 9710 to request the necessary cables that will ship with the ESS. You can specify up to 32 of the 97xx cables in any combination, with quantity tied to feature 3002. For ordering additional cables, use the feature codes 2801 to 2810.

For the cable feature codes that correspond to the different servers SCSI adapters, see the Host Adapters and Cables section of the *ESS attachment support matrix* document, which can be located at:

<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>

Table A-5 SCSI cable feature codes

Description	Feature code	Additional cables
10 Meter Cable - Ultra SCSI	9701	2801
20 Meter Cable - Ultra SCSI	9702	2802
10 Meter Cable - SCSI-2 Fast/Wide	9703	2803
20 Meter Cable - SCSI-2 Fast/Wide	9704	2804
10 Meter Cable - SCSI-2 Fast/Wide (AS/400)	9705	2805
20 Meter Cable - SCSI-2 Fast/Wide (AS/400)	9706	2806
10 Meter Cable - SCSI-2 Fast/Wide (Sun/HP (dual) PCI)	9707	2807
20 Meter Cable - SCSI-2 Fast/Wide (Sun/HP (dual) PCI)	9708	2808
10 Meter Cable - SCSI-2 Fast/Wide (HP PCI)	9709	2809
20 Meter Cable - SCSI-2 Fast/Wide (HP PCI)	9710	2810

A.2.4 ESS Master Console

The IBM TotalStorage Enterprise Storage Server Master Console (ESS Master Console) serves as a single point of control for up to seven ESSs.

A.2.4.1 ESS Master Console (2717)

This feature provides the hardware to support ESS configuration, Call Home, and remote support capabilities. The feature consists of a dedicated console (processor, modem, monitor, keyboard, multiport serial adapter) and networking components (switch and Ethernet cables). This feature must be ordered with the first ESS installed at each location.

A.2.4.2 Master Console Remote Support Cables (2716)

Up to seven ESS machines are supported per ESS Master Console. Six additional ESS machines can be connected into the ESS Master Console (fc 2717) with the Remote Support Cable feature. The feature provides Ethernet cables to attach the ESS to an existing Master Console, thus enabling the sharing of the service and configuration functions provided by the console.

A.2.4.3 Modem specify features for Remote Support (9301-9318)

Feature number 2717 supplies a modem for the ESS. To designate the country-specific modem requirements, one of the 9301 to 9318 features is used. The minimum to specify is one; the maximum to specify is one.

A.3 Disk storage configurations

The IBM TotalStorage Enterprise Storage Server Model 800 provides a selection of disk drive capacities and speeds (RPM) that can be installed within the ESS to meet specific workload requirements. The disk configurations options are ordered by feature codes as detailed in this section.

A.3.1 Capacity range (physical capacity)

The total ESS disk storage capacity is expressed as a raw storage capacity number (physical capacity). To determine the raw capacity of an ESS, perform the following calculation for *each* disk eight-pack capacity installed:

- ▶ Multiply the number of disk eight-packs by 8 to determine total number of disks
- ▶ Multiply the total number of disks by the disk capacity (18.2 GB, 36.4 GB, 72.8 GB)

Then add together the total capacities of each disk type.

Feature codes 9060 to 9067 are used to indicate the physical capacity of a new machine (excluding Step Ahead eight-packs) and is used for administrative purposes only.

A.3.2 Disk eight-packs

A disk eight-pack contains eight disk drives, all of which are of the same capacity, either 18.2 GB, 36.4 GB, or 72.8 GB, and 10,000 or 15,000 rpm. The ESS Model 800 supports the intermix of disk drive capacity and speed, subject to disk intermix limitations.

Disk intermix limitations

For a given disk drive capacity, the 15,000 rpm disk eight-packs cannot be intermixed with 10,000 rpm disk eight-packs of the same capacity. IBM has, however, issued a statement of direction as follows:

Statement of Direction: IBM plans to support the intermix of 15,000 rpm drives with lower rpm drives of the same capacity within an ESS Model 800.

All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Availability, prices, ordering information, and terms and conditions will be provided when the product is announced for general availability.

Disk eight-packs (212x and 214x)

The minimum number of eight-packs installable is four, and the maximum is 48 — if the Step Ahead feature code 9500 is not present. The ESS base frame supports 16 eight-packs, after which the Expansion Enclosure (fc 2110) is required.

The ESS Model 800 supports the intermix of eight-packs of different capacity and speed disk, subject to disk intermix limitations (see “Disk intermix limitations” on page 268).

Eight-packs are installed in pairs of same capacity. Features 212x and 214x are used to order the required disk eight-packs (Table A-6 shows a detailed list of eight-pack features).

Table A-6 Disk eight-pack feature codes

Description	Disk Eight-Packs	Step Ahead Disk Eight-Packs
18.2 GB (10,000 rpm)	2122	2132
36.4 GB (10,000 rpm)	2123	2133

Description	Disk Eight-Packs	Step Ahead Disk Eight-Packs
72.8 GB (10,000 rpm)	2124	2134
18.2 GB (15,000 rpm)	2142	2152
36.4 GB (15,000 rpm)	2143	2153

A.3.2.1 Capacity upgrades

Field-installed capacity upgrades are available for all disk drive sizes (18.2 GB, 36.4 GB, and 72.8 GB) and speeds (10,000 rpm, 15,000 rpm), subject to disk intermix limitations (see “Disk intermix limitations” on page 268). The disk eight-pack features must be installed in pairs of the same type. A capacity upgrade (adding eight-packs) can be performed concurrently with normal I/O operations on the ESS.

A.3.2.2 Step Ahead Flexible Capacity option

The ESS provides for the pre-installation of disk eight-packs via the Step Ahead program. The Step Ahead feature 9500 must be specified, followed by the required eight-pack Step Ahead features (fc 213x or 215x) shown in Table A-6 on page 268. The same disk intermix limitations apply with the Step Ahead eight-packs (see “Disk intermix limitations” on page 268). The maximum number of Step Ahead eight-packs installed is two of the same drive type.

A.3.2.3 Flexible Capacity option

Feature 9600 is used to identify machines not participating in the Step Ahead program. When you exit the Step Ahead program, the 9500 feature is replaced with the 9600 Flexible Capacity option, and the Step Ahead eight-pack features 213x or 215x are converted to standard eight-packs with feature codes 212x or 214x.

A.3.2.4 Disk eight-pack conversions

Disk eight-pack exchange is available for disk eight-pack capacity and rpm conversions. Eight-packs can be converted to a higher capacity or rpm by removing the existing feature and replacing them with the required drive feature code, for example 18.2 GB 10,000 rpm (fc 2122) to 72.8 GB 10,000 rpm (fc 2124). The eight-packs conversions must be ordered in pairs and are subject to the disk intermix limitations (see “Disk intermix limitations” on page 268).

A.3.2.5 Eight-pack count (900x-904x)

The total number of eight-packs (features 212x, 213x, 214x, 215x) installed in the ESS is recorded using the feature codes 9004 to 9048, the last two digits representing the actual count. Features 9004 to 9016 are associated with the base ESS. Features 9018 to 9048 need an Expansion Enclosure.

A.3.2.6 Eight-pack mounting kit

The mounting kit is used to house up to 8 eight-packs and consists of the sheet metal cage, power supplies, fans, and cables. All ESS base enclosures require a minimum of two mounting kits. An additional four mounting kits (cages) can be installed in the Expansion Enclosure.

Feature 9102 (disk eight-pack mounting kit) is used to specify the amount of cages required for housing the configured eight-packs (fc 212x, 213x, 214x, 215x). Feature 2102 (disk eight-pack mounting kit pre-install) is used to order additional cages for pre-installation into the Expansion Enclosure. Then the mounting kit count features 915x indicate how many mounting kits (fc 2102 plus 9102) are installed in the ESS.

A.4 Advanced Functions

The IBM TotalStorage Enterprise Storage Server (ESS) Advanced Functions enhance the capabilities of the ESS Model 800:

- ▶ Parallel Access Volumes (PAV) offer significant performance enhancements in the zSeries and S/390 environments by enabling simultaneous processing for multiple I/O operations to the same logical volume. PAV is described in detail in 6.2, “Parallel Access Volume” on page 166.
- ▶ Extended Remote Copy (XRC) is a combined hardware and software business continuance solution for the zSeries and S/390 environments providing asynchronous mirroring between two ESSs at global distances. XRC is described in detail in 7.10, “Extended Remote Copy (XRC)” on page 220.
- ▶ Peer-to-Peer Remote Copy (PPRC) is a hardware-based business continuance solution designed to provide synchronous mirroring between two ESSs that can be located up to 103 km from each other. PPRC is described in detail in 7.5, “Peer-to-Peer Remote Copy (PPRC)” on page 203. This feature includes the PPRC Extended Distance (PPRC-XD) remote copy function, for nonsynchronous mirroring between two ESSs over continental distances (the distance only limited by the network and channel extenders technology capabilities). PPRC-XD is described in detail in 7.7, “PPRC Extended Distance (PPRC-XD)” on page 212.
- ▶ FlashCopy is designed to provide a point-in-time instant copy capability for logical volumes in the ESS. FlashCopy is described in detail in 7.4, “FlashCopy” on page 200.

The ESS Function Authorization features, together with the corresponding ESS features for the Advanced Functions (80xx, 81xx, 82xx, 83xx) allow you to order these functions.

A.4.1 ESS Advanced Functions

The ESS Function Authorization feature numbers (see Table A-7) provide a set of pricing tiers for the ESS Advanced Functions. These tiers provide increased granularity (as compared to earlier models) with pricing matched to the physical capacity of the ESS Model 800 (refer to A.4.2, “Capacity tier calculation” on page 271 for tier determination).

Table A-7 ESS Function Authorization features

Advanced Function	IBM 2240 ESS Function Authorization	
	Machine type and model	features
PAV	2240-PAV	8000 to 8012
XRC	2240-XRC	8100 to 8112
PPRC	2240-PRC	8200 to 8212
FlashCopy	2240-FLC	8300 to 8312

The IBM 2240 ESS Function Authorization feature numbers are for billing purposes only and authorize the use of ESS Advanced Functions at a given capacity level on a specific ESS Model 800 (refer to Table A-8 on page 271).

Table A-8 ESS Function Authorization - Capacity tiers and features

Physical Capacity tier	IBM 2240 ESS Function Authorization (Machine type-model and features)			
	2240-PAV	2240-XRC	2240-PRC	2240-FLC
Up to 1 TB	8000	8100	8200	8300
Up to 2 TB	8001	8101	8201	8301
Up to 3 TB	8002	8102	8202	8302
Up to 4 TB	8003	8103	8203	8303
Up to 5 TB	8004	8104	8204	8304
Up to 6 TB	8005	8105	8205	8305
Up to 8 TB	8006	8106	8206	8306
Up to 10 TB	8007	8107	8207	8307
Up to 12 TB	8008	8108	8208	8308
Up to 16 TB	8009	8109	8209	8309
Up to 20 TB	8010	8110	8210	8310
Up to 25 TB	8011	8111	8211	8311
Up to 30 TB	8012	8112	8212	8312

A.4.2 Capacity tier calculation

ESS Advanced Functions are enabled and authorized based upon the physical capacity of the ESS:

- ▶ PAV and XRC enabling and authorization must be equal to or greater than the physical capacity within the ESS that will be logically configured as count-key-data (CKD) volumes for use with the zSeries and S/390 servers.
- ▶ PPRC and FlashCopy enabling and authorization must be equal to or greater than the total physical capacity of the ESS.

A.3.1, “Capacity range (physical capacity)” on page 268 explains how to calculate the physical capacity on a disk eight-pack basis.

A.4.3 Ordering Advanced Functions

Advanced Functions require the selection of IBM 2105 Model 800 feature numbers and the purchase of the matching IBM 2240 ESS Function Authorization feature numbers:

- ▶ The ESS Model 800 feature numbers (80xx, 81xx, 82xx, 83xx) enable a given function on the ESS at a given capacity level.
- ▶ The ESS Function Authorization feature numbers (80xx, 81xx, 82xx, 83xx) authorize use of the given Advanced Function at the given capacity level on the ES machine for which it was purchased.

The ESS Model 800 feature numbers (8xxx) and the ESS Function Authorization feature numbers (8xxx) are co-requisites and must always correspond to one another (refer to Table A-9 on page 272).

Table A-9 ESS Model 800 and ESS Function Authorization features correspondence

ESS Advanced Function	ESS Model 800 feature	IBM 2240 ESS Function Authorization feature
Parallel Access Volume (PAV)	2105-800 features 80xx	2240-PAV features 80xx
Extended Remote Copy (XRC)	2105-800 features 81xx	2240-XRC features 81xx
Peer-to-Peer Remote Copy (PPRC)	2105-800 features 82xx	2240-PRC features 82xx
FlashCopy	2105-800 features 83xx	2240-FLC features 83xx

The ESS Function Authorizations (IBM 2240 FLC, PAV, PRC, and XRC) must be ordered with the ESS Model 800. The ESS Model 800 order must include the specification of the matching 8xxx Advanced Function feature (see Table A-10). Refer to A.4.2, “Capacity tier calculation” on page 271 for tier determination.

ESS Function Authorization and ESS Model 800 capacity tiers feature numbers are similar, as you can see comparing Table A-8 on page 271 and Table A-10.

Table A-10 ESS Model 800 - Capacity tiers and features

Physical Capacity tier	ESS Model 800 (2105-800) features			
	Parallel Access Volumes	Extended Remote Copy	Peer-to-Peer Remote Copy	FlashCopy
Up to 1 TB	8000	8100	8200	8300
Up to 2 TB	8001	8101	8201	8301
Up to 3 TB	8002	8102	8202	8302
Up to 4 TB	8003	8103	8203	8303
Up to 5 TB	8004	8104	8204	8304
Up to 6 TB	8005	8105	8205	8305
Up to 8 TB	8006	8106	8206	8306
Up to 10 TB	8007	8107	8207	8307
Up to 12 TB	8008	8108	8208	8308
Up to 16 TB	8009	8109	8209	8309
Up to 20 TB	8010	8110	8210	8310
Up to 25 TB	8011	8111	8211	8311
Up to 30 TB	8012	8112	8212	8312

A.5 Additional feature codes

There are additional features used to configure the ESS. Some of these features are so-called *specifies*, and they allow the manufacturing plant to set up the ESS with the appropriate cables and parts characteristics for the location where the ESS will be operating.

A.5.0.1 Expansion Enclosure (2110)

The Expansion Enclosure is used for adding additional disk eight-packs into an ESS configuration. The enclosure can support up to 32 disk eight-packs and four disk mounting

kits (cages); a minimum of one mounting kit must be ordered with the enclosure. Only one Expansion Enclosure can be added to an ESS. One power cord (feature code 98xx or 998x) must be specified and must be of the same type as the base ESS enclosure.

A.5.0.2 Reduced shipping weight (0970)

This feature ensures that the maximum shipping weight of the ESS base enclosure and Expansion Enclosure (feature 2110) not exceed 2,000 pounds during the initial shipment.

The weight of the ESS is reduced by removing selected components, which will be shipped separately. The components will be installed into the machine by the IBM System Service Representative during machine installation. This feature will increase the machine installation time and should be ordered if required.

A.5.0.3 Operator panel language groups

These features are used to select the operator panel language. The default is English. The specify codes 2928 to 2980 can be used if an alternative language is required.

A.5.0.4 Power cords

The ESS power control subsystem is fully redundant. Power is supplied through two 30, 50, or 60 amp line cords, with either line cord capable of providing 100% of the needed power. One feature is required for the base and one for the Expansion Enclosure, and the feature must be the same for both. Replacement cords can be ordered using feature 2651 to 2655 and 2687 to 2689.

- ▶ 9851 for three phase, 50/60 Hz, 50 amp
- ▶ 9852 for three phase, 50 Hz, 50 amp (EMEA)
- ▶ 9853 for three phase, 50/60 Hz, 60 amp (US and Japan)
- ▶ 9854 for three phase, 50/60 Hz, 60 amp
- ▶ 9855 for three phase, 50/60 Hz, 30 amp
- ▶ 9987 for three phase, 50/60 Hz, 50 amp (Chicago)
- ▶ 9988 for three phase, 50/60 Hz, 60 amp (Chicago)
- ▶ 9989 for three phase, 50/60 Hz, 30 amp (Chicago)

A.5.0.5 Input voltage

The input voltage feature determines the type of AC power supply used within the ESS base and Expansion Enclosures and must be selected using one of the 9870 or 9871 specify codes.

- ▶ 9870 for nominal AC Voltage: 200V-240V
- ▶ 9871 for nominal AC Voltage: 380V-480V

A.5.0.6 Remote Power Control (1011)

Provides the components to enable the ESS power to be controlled remotely.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

IBM Redbooks

For information on ordering these publications, see “How to get IBM Redbooks” on page 276.

- ▶ *IBM TotalStorage Enterprise Storage Server: Implementing the ESS in Your Environment*, SG24-5420-01
- ▶ *Implementing ESS Copy Services on UNIX and Windows NT/2000*, SG24-5757
- ▶ *Implementing ESS Copy Services on S/390*, SG24-5680
- ▶ *IBM TotalStorage Enterprise Storage Server PPRC Extended Distance*, SG24-6568
- ▶ *IBM TotalStorage Solutions for Disaster Recovery*, SG24-6547
- ▶ *IBM TotalStorage Expert Hands-On Usage Guide*, SG24-6102
- ▶ *AIX Storage Management*, GG24-4484
- ▶ *Implementing Fibre Channel Attachment on the ESS*, SG24-6113
- ▶ *Introduction to Storage Area Network, SAN*, SG24-5470
- ▶ *IBM @server iSeries in Storage Area Networks A Guide to Implementing FC Disk and Tape with iSeries*, SG24-6220 - redpiece available at <http://www.ibm.com/redbooks> (expected redbook publish date September 2002)
- ▶ *FICON Native Implementation and Reference Guide*, SG24-6266

Other resources

These publications are also relevant as further information sources:

- ▶ *IBM TotalStorage Enterprise Storage Server Web Interface User's Guide*, SC26-7448
- ▶ *IBM TotalStorage Enterprise Storage Server Copy Services Command-Line Interface Reference*, SC26-7449
- ▶ *IBM TotalStorage Enterprise Storage Server Host System Attachment Guide*, SC26-7446
- ▶ *IBM TotalStorage Enterprise Storage Server Introduction and Planning Guide*, GC26-7444
- ▶ *IBM TotalStorage Enterprise Storage Server User's Guide*, SC26-7445
- ▶ *IBM TotalStorage Enterprise Storage Server DFSMS Software Support Reference*, SC26-7440
- ▶ *z/OS DFSMS Advanced Copy Services*, SC35-0428
- ▶ *IBM TotalStorage Subsystem Device Driver User's Guide*, SC26-7478

Referenced Web sites

These Web sites are also relevant as further information sources:

- ▶ ESS support information
<http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm>
- ▶ ESS Web site
<http://www.storage.ibm.com/hardsoft/products/ess/ess.htm>
- ▶ ESS technical support
<http://ssddom02.storage.ibm.com/techsup/webnav.nsf/support/2105>
- ▶ IBM TotalStorage Expert
<http://www.storage.ibm.com/software/storwatch/ess/index.html>

How to get IBM Redbooks

You can order hardcopy Redbooks, as well as view, download, or search for Redbooks at the following Web site:

ibm.com/redbooks

You can also download additional materials (code samples or diskette/CD-ROM images) from that site.

IBM Redbooks collections

Redbooks are also available on CD-ROMs. Click the CD-ROMs button on the Redbooks Web site for information about all the CD-ROMs offered, as well as updates and formats.

Index

Numerics

- 2 Gb Fibre Channel/FICON
 - host adapters 7, 9, 20, 35
- 2 GB NVS 5–6, 20, 30, 50, 52
- 2N 53, 58
- 3+3+2S array 26, 65, 105, 120
- 32,760 cylinder volume support 184
- 3380 77, 97, 127
- 3390 77, 97, 127
- 4+4 array 26, 66, 105, 120
- 6+P+S array 26, 64, 104, 120
- 64 GB cache 20, 29, 156
- 7+P array 26, 64, 104, 120

A

- AC dual power 58
- access mode 74, 114, 132
- Add Volumes panel 126, 128
- adding CKD volumes 126
- addressing 72
 - CKD hosts 75
 - FCP 74
 - FICON specifics 75
 - SCSI host 73
- alias address 129, 171
- AMS LISTDATA command 234
- ANTRQST macro 195, 202, 208, 221
- arbitrated loop 40, 137
- architecture 49, 51
 - addressing characteristics 72
 - CKD 13
 - FB 13
 - Seascape 2
 - z/Architecture 16
- array 17
 - 3+3+2S 65, 105
 - 4+4 66, 105
 - 6+P+S 64, 104
 - 7+P 64, 104
 - data striping 152
- asset management 160
- automation 216
- availability
 - design 51
 - features 53
 - Fibre Channel connection 137
 - multipathing 159

B

- back end 6, 17, 151
- balancing with RAID 10 66
- base address 129, 171
- base enclosure 21, 24, 98

- battery 44, 58
- bays 35

C

- cache 20, 29, 85–86, 264
 - algorithms 150
 - management 6
 - performance considerations 156
 - read operations 87
 - RMF reporting 164
 - sequential read operations 90
 - sequential write operations 91
 - size 20, 29, 101
- cages 22–24, 98
- Call Home 46, 54
- capacity
 - CKD logical devices 127
 - disk drives 25
 - intermix 106
 - RAID 5 vs RAID 10 107
 - raw 95
 - tier calculation 271
- capacity management 161
- catch-up transition 215
- CD-ROM 45
- channel extenders 8, 37, 217
- CKD
 - architecture 13
 - base and alias 129, 171
 - defining logical devices 128
 - host addressing 75
 - host view of ESS 76
 - logical device capacities 127
 - LSS 68, 77, 116
 - server 15
 - server mapping 71
- clusters 29, 50, 149
 - failover and failback 62
- combination of RAID 5 and RAID 10 67
- command-line interface 195, 198, 209
- commands
 - AMS LISDATA 234
 - CP users native FlashCopy 196
 - DEVSERV QPAVS 173
 - ESS Copy Services CLI 198
 - ESS Copy Services TSO commands 195, 207
 - FlashCopy commands 202
 - LISTDATA 164
 - PPRC TSO commands 207
 - Report LUNs 80
 - rsQueryComplete 215
- Common Parts Interconnect, *See* CPI
- Concurrent Copy 218
- concurrent maintenance 56, 58

- configuration
 - logical 94
 - performance recommendations 155
 - physical 94, 101
 - planning 248
 - reconfiguration of ranks 107
- Configure Disk Groups panel 123
- Configure Host Adapter Ports panel 113, 115
- Configure LCU panel 116
- connectivity
 - ESCON 138
 - Fibe Channel 136
 - FICON 140
 - PPRC 216
 - SCSI 134
- connectors
 - LC and SC 39
- consistency group 207
- consolidation of storage 4
- Control Unit Initiated Reconfiguration, *See* CUIR
- controller image, *See* LCU
- copy functions 194
 - Concurrent Copy 218
 - FlashCopy 200
 - management 195, 202, 207
 - PPPC 203
 - PPRC-XD 212
 - XRC 220
- Count-key-data, *See* CKD
- CPI 50, 83, 149
- CUIR 54, 63, 232, 236
- custom volumes 183

D

- daisy chaining 135
- DASD 13
- data flow
 - host adapters 83
 - read 84
 - write 85
- data migration
 - UNIX systems 254
 - VSE/ESA systems 253
 - z/OS systems 250
 - z/VM systems 253
- data protection 7, 52, 55
 - RAID implementation 63
- data striping 152
- DC power supply 58
- DDM 14, 118
 - eight-pack 17
 - replacement 60
- design
 - data flow 84
 - fault tolerant 7
 - for data availability 51
- destage 17
- device 14
- device adapters 31, 50, 68, 85–86, 102
 - SSA 22

- DEVSERV command 173
- DFSMSdss utility 195, 202, 218
- dimensions 23
- Direct Access Storage Device, *See* DASD
- disconnect time 186, 188
- disk
 - destage 17
 - disk group 17
 - eight-pack 17
- disk array 17
 - 3+3+2S 26, 65, 105, 120
 - 4+4 26, 66, 105, 120
 - 6+P+S 26, 64, 104, 120
 - 7+P 26, 64, 104, 120
- disk drive module, *See* DDM
- disk drives 22, 25
 - 15,000 rpm 20, 26, 154, 156
 - conversions 28
 - eight-pack 26
 - intermixing 27
 - limitations 27
 - per loop 32
 - performance considerations 156
 - replacement 60
- disk group 17, 103, 118
- diskette drives 45
- distances
 - ESCON 36
 - fiber distances 42
 - Fibre Channel 40
 - PPRC supported distances 217
- duplex pending XD volume state 206, 214
- duplex volume state 205
- DWDM 8, 36, 218
- dynamic PAV tuning 172

E

- eight-pack 17, 26
 - base enclosure fill up 98
 - effective and physical capacity 26
 - Expansion Enclosure fill up 99
 - upgrades 100
- Enterprise Storage Server Network, *See* ESSNet
- ESCON
 - addressing 72
 - cache hit operation 187
 - cache miss operation 188
 - FICON intermix 144
 - host 15
 - host adapters 11, 35–36
 - host attachment 138
 - host view of ESS 76
 - logical paths 139
 - PPRC ports 211
- ESS Advanced Functions 270
- ESS Copy Services 103, 194
 - command-line interface (CLI) 195, 198, 209
 - Concurrent Copy 218
 - iSeries 224
 - management 195, 202, 207

- setup 199
- TSO commands 195, 199
- Web user interface (WUI) 195–196, 202, 209
- XRC 221
- ESS Copy Services Welcome screen 196
- ESS Expert 160
 - support information 242
- ESS features
 - 2110 21
 - 3024 39, 42, 265
 - 3025 39, 42, 265
- ESS local area network, *See* ESSNet
- ESS Master Console 45–46
 - ESSNet 108
 - user's data protection 55
 - Web browser 109
- ESS Model 800
 - characteristics 3, 5, 20
 - internal diagram 50
 - major components 23
 - performance 4
 - photograph 22
 - remote copy functions 8
- ESS Specialist
 - Add Volumes panel 126, 128
 - Configure Disk Groups panel 123
 - Configure Host Adapter Ports panel 113, 115
 - Configure LCU panel 116
 - Fixed Block Storage panel 125
 - Modify Host Systems panel 113
 - Modify PAV Assignments panel 173
 - Storage Allocation panel 110
 - support information 242
 - Web user interface 109
- ESSNet 46–47, 55
 - setup 108
- Expansion Enclosure 21, 24, 99
- extended long busy 207
- Extended Remote Copy, *See* XRC

F

- failback 53, 63
- failover 53, 62
- fast write 89
- fault tolerant 7, 53
- FB
 - architecture 13
 - assigning logical volumes 128
 - FCP attached LSS 81
 - FCP logical devices 131
 - logical devices 130
 - LSS 68, 118
 - SCSI attached LSS 79
 - server 15
- features 264
 - 2110 21
 - 3024 39, 42
 - 3025 39, 42
- fiber distances 42
- Fibre Channel

- addressing 72, 74
- availability connection 137
- host adapters 13, 35, 39
- host attachment 136
- host mapping 71
- host view of ESS 80
- WWPN 80
- FICON
 - addressing 72, 75
 - benefits 185, 190
 - cache hit operation 189
 - cache miss operation 190
 - control unit images 141
 - ESCON intermix 144
 - host adapters 11, 35, 41
 - host attachment 140
 - host view of ESS 76
 - logical paths 142
 - resources exceeded 144
 - RMF support 164
 - support information 231, 233, 236–237
 - XRC support 224
- Fixed Block architecture, *See* FB
- Fixed Block Storage panel 125
- FlashCopy 8, 200, 231
 - iSeries 225
 - management 202
 - support information 233, 242
- floating spare 60
- front end 17
- Function Authorization feature numbers 270

G

- GDPS 227
- Goal mode 174, 177
- go-to-SYNC 215

H

- hard disk drive, *See* hdd
- hardware
 - base enclosure 98
 - battery backup 44
 - cages 22–24
 - characteristics 20
 - device adapters 102
 - disk drives 25
 - Expansion Enclosure 99
 - host adapters 35
 - host adapters sequence 101
 - power supplies 43
 - requirements 230
- Hardware Configuration Definition, *See* HCD
- HCD 128, 177
 - PAV definition 173
- hdd 14, 118
 - eight-pack 17
 - replacement 60
- host adapters 22, 35, 50, 85–86, 265
 - 2 Gb Fibre Channel/FICON 7, 9, 20

- bays 35
- configuration recommendations 155, 191
- data flow 83
- ESCON 11, 36
- Fibre Channel 13, 39
- FICON 11, 41
- installation sequence 101
- logical configuration 113, 115
- SCSI 12, 37
- host mapping to LSS 71

I

- I/O drawer 22, 50, 149
- I/O operations 150
- IBM TotalStorage Expert 160
- ICKDSF utility 195, 208, 234
- initial configuration 248
- installation planning 246
- Intel-based servers 15
- interfaces 45
- intermixing
 - disk drives 27, 106
 - FICON and ESCON 144
 - RAID 5 and RAID 10 ranks 106, 120
 - spare pool 60
- IOSQ time 169, 186
- iSeries 40, 82
 - 2105 and 9337 volumes 97
 - assigning volumes 128
 - ESS Copy Services 224
 - FlashCopy 225
 - LUNs 82
 - PPRC 226
 - server 15

L

- large volume support 184, 232, 234, 236
- LC connector 39
- LCU 17, 68
- Least Recently Used, *See* LRU
- Licensed Internal Code (LIC) 230
- limitations
 - disk drive intermix 27
- Linux
 - support information 239
- LISTDATA command 164
- load balancing 158
- Load Source Unit (LSU) of iSeries 224
- logical configuration 94, 97
 - Add Volumes panel 126, 128
 - assigning iSeries volumes 128
 - base and alias address 129, 171
 - CKD devices 128
 - CKD LSS 116
 - common terms 111
 - Configure Disk Groups panel 123
 - Configure Host Adapter Ports panel 113, 115
 - Configure LCU panel 116
 - configuring CKD ranks 123

- configuring FB ranks 125
- defining PAVs 173
- FB devices 130
- FB FCP devices 131
- FB LSS 118
- Fixed Block Storage panel 125
- Modify Host Systems panel 113
- process 112
- standard logical configuration 97, 111, 145
- Storage Allocation panel 110

Logical Control Unit, *See* LCU

logical devices 14

logical paths 36

- ESCON establishment 139

- FICON establishment 141

- PPRC paths 211

- resources exceeded 144

Logical Storage Subsystem, *See* LSS

logical unit number, *See* LUN

logical volumes 14

long wave 39, 42

loop

- availability 33

- configuration 100, 103

- RAID 10 arrays 105

- RAID 5 arrays 104

LRU 86, 89

LSS 68

- CKD logical configuration 116

- CKD logical subsystems 77

- FB logical configuration 118

- FB-SCSI logical subsystems 79

- FCP logical subsystems 81

LUN 14, 38, 78, 80

- access modes 74

- affinity 73–74

M

mainframe 15

maintenance

- concurrent logic maintenance 56

- concurrent power maintenance 58

maintenance strategy 54

major components 23

management

- Concurrent Copy 218

- FlashCopy 202

- PPRC 207

- PPRC-XD 216

- XRC 221

mapping

- CKD server 71

- device adapter to LSS 68

- ESCON server view 76

- Fibre Channel host 71

- Fibre Channel server view 80

- FICON server view 76

- host to LSS 71

- ranks 69

- SCSI host 71

- SCSI server view 78
- Master Console, *See* ESS Master Console
- measurement tools 159
- migration 249
 - from ESCON to FICON 256
 - from SCSI to Fibre Channel 255
 - services 259
- mirrored internal DASD for iSeries 225
- Modify Host Systems panel 113
- MOST terminal 45
- Multiple Allegiance 6, 92, 179, 231
 - support information 233

N

- N+1 44, 58
- non-volatile storage, *See* NVS
- NVS 20, 30, 50, 52
 - LRU 89
 - RMF reporting 164
 - sequential write operations 91
 - write operations 88

O

- open systems 16
 - support information 240

P

- Parallel Access Volume, *See* PAV
- paths
 - failover 159
 - load balancing 158
- PAV 6, 92, 129, 166, 187, 231
 - assignment 174
 - querying 173
 - support information 233
 - tuning 171
- Peer-to-Peer Remote Copy, *See* PPRC
- pending time 186
- performance 4
 - accelerators 92, 103, 148
 - back end 151
 - cache algorithms 150
 - custom volumes 183
 - disk drives 154, 156
 - FICON benefits 185, 190
 - host adapters considerations 191
 - management 161
 - Multiple Allegiance 179
 - PAV 166
 - PAV tuning 171
 - Priority I/O Queuing 181
 - RAID 5 vs RAID 10 107, 153
 - recommendations 155
 - sequential pre-fetch 150
 - SMP processors 3, 50, 101, 149
 - software tools 159
 - third-generation hardware 5, 20, 148
 - WLM 174

- zSeries accelerators 92, 103, 166
- photograph 22
- physical capacity 20, 95
- physical configuration 94–95, 101
- physical installation 248
- physical planning 246
- PIPER 260
- Priority I/O Queuing 92
- point-in-time copy 8, 200
- point-to-point 40, 116, 137
- power supplies 58
- PPRC 8, 203, 231
 - command-line interface (CLI) 209
 - configuration guides 210
 - data consistency 215
 - ICKDSF 208
 - iSeries 226
 - management 207
 - support information 233, 241
 - supported distances 217
 - TSO commands 207
 - volume states 205
 - Web user interface 209
- PPRC Extended Distance, *See* PPRC-XD
- PPRC-XD 8, 37, 212, 231
 - for initial establish 216
 - support information 233
- Priority I/O Queuing 181
- processors 20, 23, 29, 85, 149, 264
 - drawer 22, 50
 - Turbo option 101, 264
- pSeries
 - server 16

R

- RAID 10 7, 20
 - combination with RAID 5 67
 - data striping 153
 - implementation 65
 - loop configuration 105
 - rank effective capacity 26
 - storage balancing 66, 122
 - vs RAID 5 106, 153
- RAID 5 7
 - combination with RAID 10 67
 - data striping 152
 - implementation 64
 - loop configuration 104
 - rank effective capacity 26
 - vs RAID 10 106, 153
- ranks 63
 - assigning logical volumes 125
 - configuration example 133
 - configuring CKD ranks 123
 - configuring FB ranks 124
 - data striping 152
 - disk group 118
 - performance recommendations 156
 - RAID 10 105
 - RAID 5 104

- RAID 5 and RAID 10 intermix 106, 120
- RAID 5 vs RAID 10 106, 153
- RAID reconfiguration 107
- raw capacity 95
- read
 - cache operation 87
 - data flow 84
 - sequential operation 88, 90
- Redbooks Web site 276
 - Contact us xxi
- remote copy 8
- Remote Service Support 47, 54
- replacement disk drive 60
- Report LUNs command 80
- RIO 149
- RMF 164
- rsQueryComplete command 215

S

- SAN 9, 39–40
- SC connector 39
- SCSI
 - addressing 72–73
 - host adapters 12, 35, 37
 - host attachment 134
 - host mapping 71
 - host view of ESS 78
 - SCSI ID 14
- SDD 35, 157, 241
- SDM 220
- Seascope Architecture 2
- sequential
 - detection 90
 - reads 88
- sequential operations
 - pre-fetch 150
 - read 90
 - write 91, 152
- server
 - CKD 15
 - FB 15
 - Intel-based 15
 - iSeries 15
- SETCACHE command 164
- short wave 39, 42
- simplex volume state 205
- Single Level Storage of iSeries 82
- site requirements 247
- SMP processors 20, 23, 29, 50, 101, 149
- software requirements 230
- software tools
 - ESS Expert 160
 - RMF 164
- sparing 32, 59
 - capacity intermix 106
 - spare pool 104–105
- spatial reuse 33–34
- SSA
 - back end throughput 151
 - device adapters 22, 31, 68

- loop 32
 - operation 33
 - sparing 32, 59
- stage 18
- static volumes 207
- Step Ahead option 28, 96
- Storage Allocation panel 110
- Storage Area Network, *See* SAN
- storage consolidation 4
- strategy for maintenance 54
- striping 152
- Subsystem Device Driver, *See* SDD
- support information 230
- supported servers
 - ESCON 37
 - Fibre Channel 40
 - FICON 42
 - SCSI 38
- suspended volume state 205
- switched fabric 40, 116, 137
- sysplex I/O management 6
- System Data Mover, *See* SDM

T

- terminology 10
- third-generation hardware 5, 20, 148
- tools
 - ESS Expert 160
 - RMF 164
- topologies 42, 137
- TPF
 - support information 238
- TSO
 - ESS Copy Services commands 195, 199
 - FlashCopy commands 202
 - PPRC commands 207
- TSO commands 199
- tuning PAVs 171
- Turbo option 29, 50, 101

U

- UCB busy 167
- UNIX
 - server 16
- upgrades
 - eight-pack 100
- utilities
 - DFSMSdss 195, 202, 218
 - ICKDSF 195, 208, 234

V

- volume 14
 - 3390 and 3380 77
 - assigning to a rank 125
 - custom 183
 - large volume support 184
 - PPRC volume states 205
- VSE/ESA

- data migration 253
- support 236

W

- Web user interface 109, 195–196
- WLM 174
- Workload Manager, *See* WLM
- worldwide port name, *See* WWPN
- write
 - cache and NVS operation 88
 - data flow 85
 - sequential operations 91, 152
- WWPN 80

X

- XRC 220
- xSeries
 - server 16

Z

- z/Architecture 16
- z/OS
 - data migration 250
 - software tools 163
 - support information 232
- z/VM 16
 - data migration 253
 - native FlashCopy commands 196
 - support information 235
- zSeries
 - performance accelerators 92, 166
 - server 16
 - support information 230



Redbooks

IBM TotalStorage Enterprise Storage Server Model 800

(0.5" spine)
0.475" x 0.873"
250 -> 459 pages



IBM TotalStorage Enterprise Storage Server Model 800



Learn about the characteristics of the new ESS Model 800

Know the powerful functions available in the ESS

Discover the advanced architecture of the ESS

This IBM Redbook describes the IBM TotalStorage Enterprise Storage Server Model 800, its architecture, its logical design, hardware design and components, advanced functions, performance features, and specific characteristics.

The information contained in this redbook will be useful for those who need a general understanding of this powerful model of disk enterprise storage server, as well as for those looking for a more detailed understanding on how the ESS Model 800 is designed and operates.

In addition to the logical and physical description of the ESS Model 800, also the fundamentals of the configuration process are described in this redbook. This is all useful information for the IT storage person for the proper planning and configuration when installing the ESS, as well as for the efficient management of this powerful storage subsystem.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks