



z/OS Central Storage Management

Kathy Walsh IBM

August 11, 2015 Session Number: 17634





#SHAREorg

SHARE is an independent volunteer-run information technology association that provides education, professional networking and industry influence.

Copyright (c) 2015 by SHARE Inc. 💿 🚯 🏵 Except where otherwise noted, this work is licensed under http://creativecommons.org/licenses/by-nc-sa/3.0/

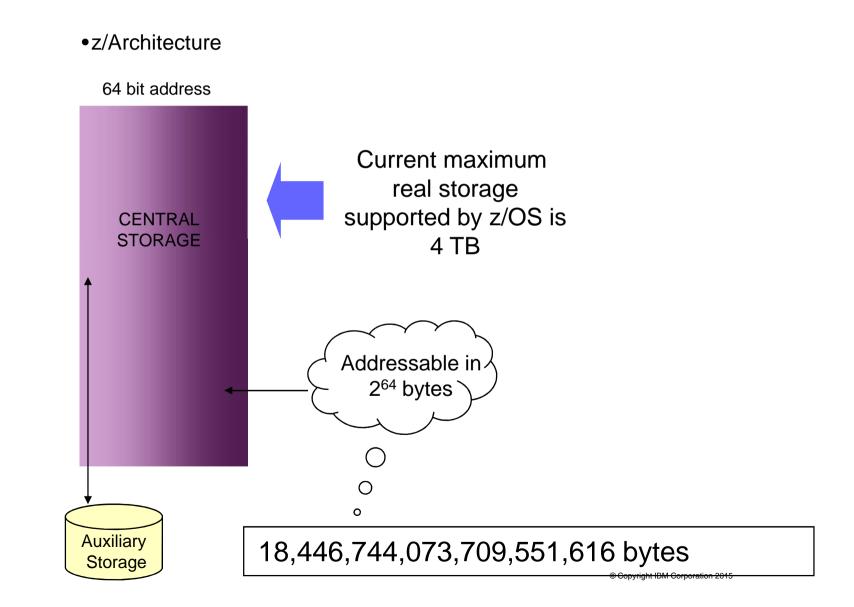
Agenda



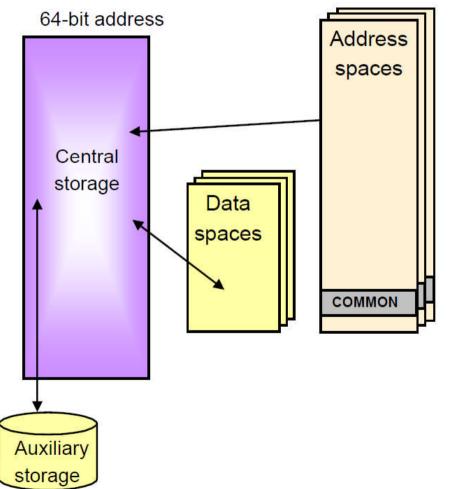
- Overview of Storage Management
- Architecture Overview
- Virtual Storage Areas
- Real Storage
- Value of Large Frames
- Storage Contention and Paging
- UIC
- Available Frame Queue
- Storage Shortage Management
- Pageable Storage Shortage Processing
- Auxiliary Storage Shortage Processing

Enterprise Server Storage: Central and Aux



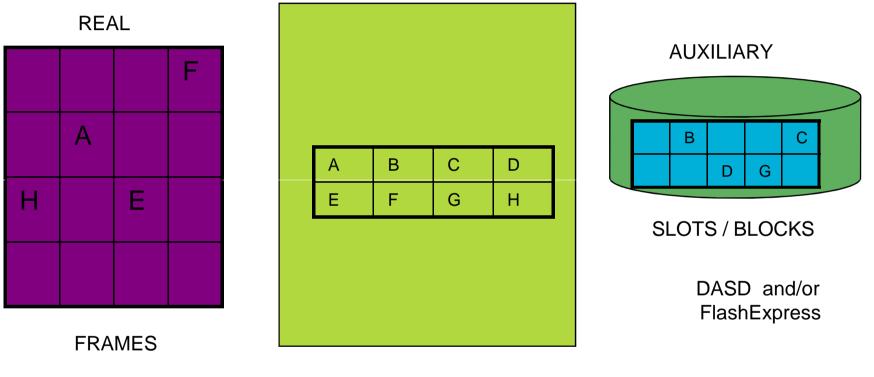


z/Architecture



Frames, Pages, and Slots

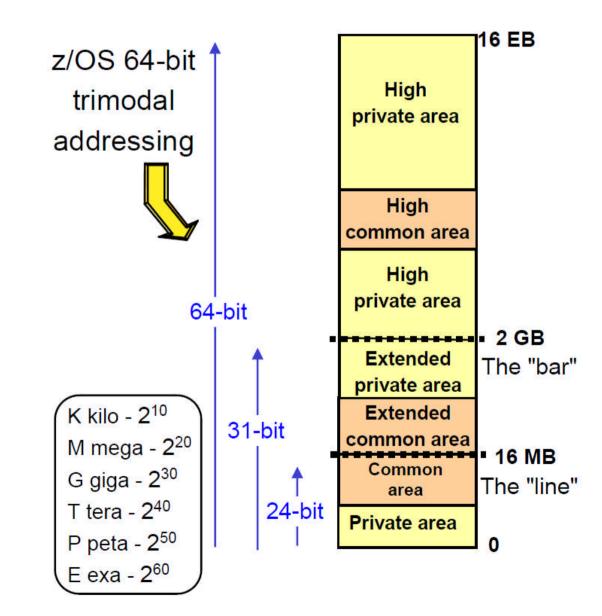




VIRTUAL

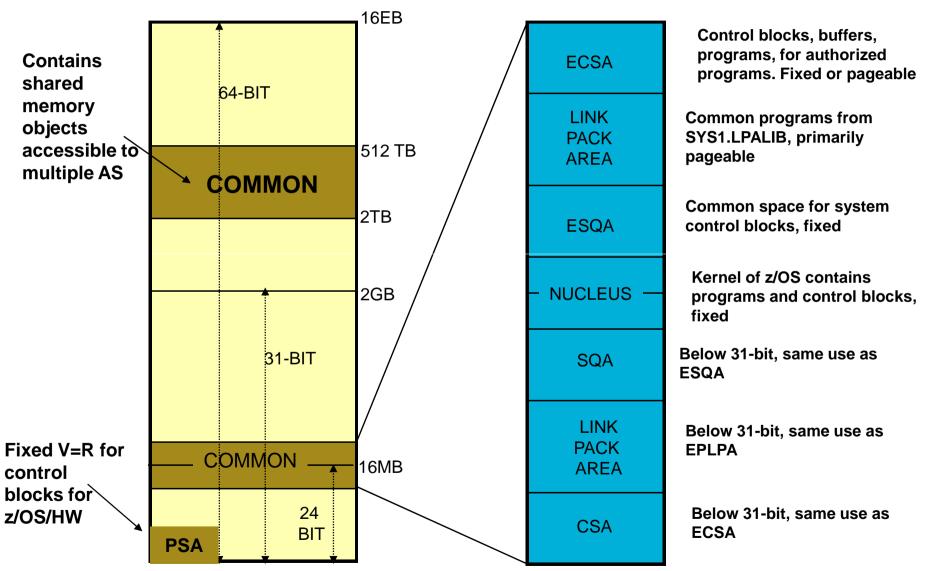
PAGES

Address Space Layout



Common Area Layout



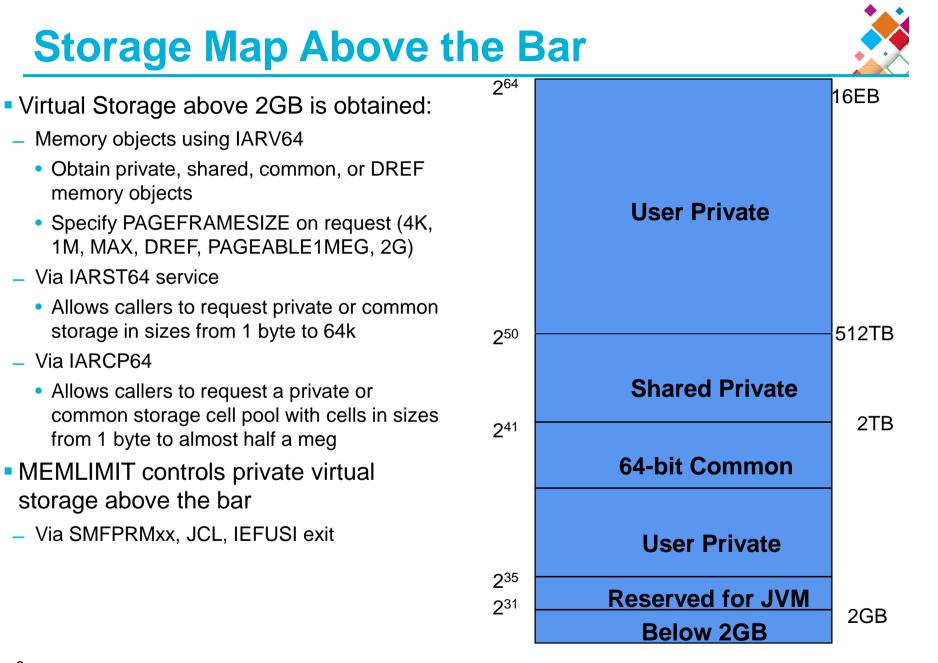


Virtual Storage

- RMF Report
- SMF 78 record
- REPORT(VSTOR)
- Size of (E)CSA and (E)SQA are defined at IPL via SYS1.Parmlib
- Private region is still important to monitor for 24-bit programs

	STATIC	STORAGE	MAP	
AREA	1	ADDRESS	SIZE	
EPVT	10	CD00000	1587M	
ECSA	1	8C20000	321M	
EMLPA		0	0K	
EFLPA	1	8C1D000	12K	
EPLPA		46DB000	69.3M	
ESQA		1BB7000	43.1M	
ENUC		1000000	11.7M	
16	5 MEG BO	JUNDARY		
NUCLEUS		FD6000	168K	
SQA		E95000	1284K	
PLPA		CE5000	1728K	
FLPA		CDA000	44K	
MLPA		0	0K	
CSA		900000	3944K	
PRIVATE		2000	9208K	
PSA		0	8K	





Memory Objects

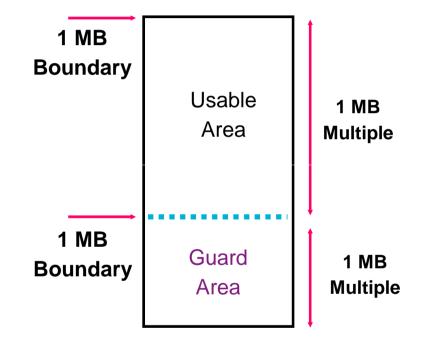


Memory Objects are

- Contiguous range of virtual addresses
- Allocated as a number of 1 MB increments of storage, on a 1 MB boundary
- Some of the memory is useable virtual storage, and the rest is not valid and is called the "guard" area
- Extent of useable virtual can be changed with a compensatory change in the extent of the guard area

Memory Object attributes:

- Allocated by a single request and freed in it's entirety
- Defined with a single storage protection key and fetch protection attribute
- Private memory objects are owned by a task
- Memory management operations must be performed within an object, i.e. cannot cross a memory object boundary



Allocating Virtual Storage



 The size of the High Common, High Shared and Large Frames areas are defined in the IEASYSxx parmlib member

- The LFAREA controls the amount of real storage allocated to support 1MB fixed and 2GB fixed pages
- Can only be changed by an IPL so plan and monitor closely
- LPAR must have more than 4GB of real storage to get an LFAREA, and must have 6.5GB of real storage to get a single 2GB page
- Backed by contiguous 4K pages at allocation time
- Free 1MB pages can be used if available 4K pages become constrained
- Available storage in the LFAREA can be used to support pageable 1MB pages
- Over-allocation of LFAREA can have a negative effect on the system if it reduces the quantity of 4K pages needed for normal system operations

1 MB Pageable Large Pages

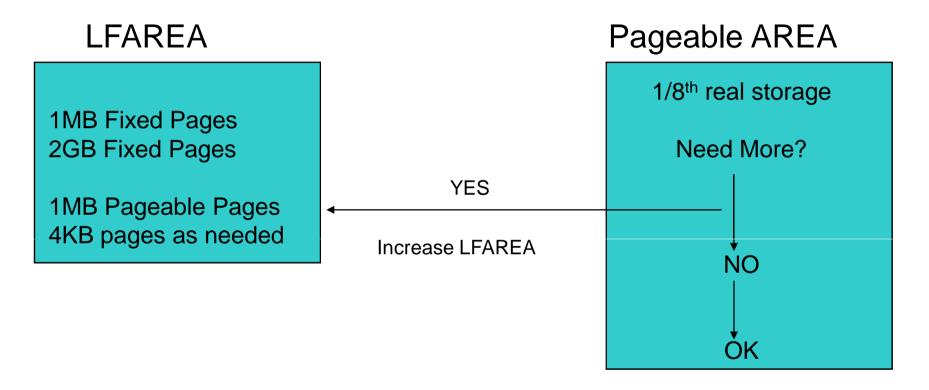


- The pageable 1MB page area is set to approximately 1/8th of the real storage
- Can be used to support 4KB page requests
- Obtained via the IARV64 GETSTOR|GETCOMMON macro
- Requires z/OS 2.1 or 1.13 with RSM Enablement Offering
- Requires zEC12, zBC12, or z13 processor
- Pageable large pages are backed when referenced
- Fixed large pages are backed at allocation time
- Pageable large pages can be moved to Aux
- With Flash, large pages are paged to Flash and retain their 'large page' attribute
- Without Flash, large pages will be demoted to 256 4KB pages, a page table will be built and pages sent to Aux
 - Pages lose their 'large page' attribute
 - Potential performance problem because no enhancements to AUX DASD I/O performance
- Every SCM capable LPAR will receive an allocation of 1 MB Pageable pages

- Allows for the dynamic addition (hot plugging) of Flash Express Cards







Large Frames require contiguous memory!

The Value of Large Memory



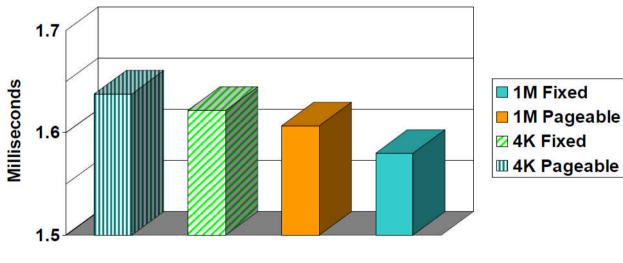
- Substantial Latency Reduction for OLTP workloads
- Significant response time reductions
- Increased transaction rates
- In-Memory Databases see dramatic gains in response time by avoiding I/O wait
- Batch Window Reduction
- More concurrent workloads
- Shorter elapsed times for jobs
- CPU performance improvements
- Increased application memory sizes due to 64-bit addressing has put pressure on the Hardware Translation Lookaside Buffer (TLB)
 - TLB sizes have remained relatively small due to low access time requirements and hardware space limitations
 - TLB coverage today represents a much smaller fraction of an applications working set leading to a larger number of TLB misses
 - Applications can incur a performance penalty due to the number of TLB misses and the increased CPU cost of each TLB miss
- Large pages increase TLB coverage without proportionally enlarging the TLB
 - Large pages allow for a single TLB entry to fulfill many more address translations

Benefits of 1MB Frames - Example



- Using DB2 10, brokerage workload, all buffer pools were backed by real storage
- zEC12 16 CPs, 5000-6000 tps simple to complex transactions
- 120GB real storage with LFAREA=70GB for 1MB measurements
- Results:
- 1MB Pageable frames are 2% better than 4K pageable frames for the workload*
- 1 MB Fixed Frames give the best overall performance

Total DB2 CPU Time per Transaction

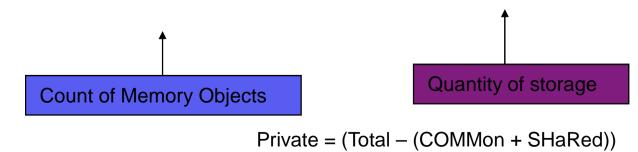


* Laboratory synthetic extensions used to create 1M Pageable Frames for buffer pools



RMF Monitor 3 STORM report

		RMF V2	R1 St	orage M	emory	Objects		Li	.ne 1 c	of 62	
Samples: 1	00	System:	SYSB	Date: 0	9/20/1	4 Time	: 15.	.00.00	Range	e: 100	Sec
					O						
				- System	Summa	ry					
MemObj-		Frames		-1MB Me	mObj-	1MI	B Fix	ked	-1MB	Pageal	ble-
Shared	5	Shared 23	562	Total	0	Tota	1	0	Initi	al 7	056
Common	70	Common 16	496	Common	0	Commo	on	0	Dynam	nic	0
		%Used 1	3.1			%Usec	d	0.0	%Usec	i :	1.6
	Serv	 ice	N	 Memory C	 bjects	11	 MB Fr	 cames-		Bytes	
Jobname C				_	-					-	
Jobname C IXGLOGR S	Clas	s ASID	Total	_	-			gable		Comm	Shr
	Clas SYST	s ASID EM 0027	Total 256	Comm 0	Shr	1 MB Fi:	xed I	Pgable 4	Total	Comm 0	Shr 0
IXGLOGR S	Clas SYST OPSD	s ASID EM 0027 EF 0088	Total 256	Comm 0	Shr 0	1 MB Fi: 0 0	xed I 0 0	Pgable 4 101	Total 258M	Comm 0 0	Shr 0 64.0M



Managing Memory Objects



RMF Postprocessor Report

– REPORT(PAGING)

z/OS V2R1 RPT VERSION V2R1 RMF		M ID SYSB				
-OPT = IEAOPT00 LF						STORAGE FRAMES
MEMORY OBJECTS	COMMON	SHARED	1 MB			
MIN	70	5	 0			
MAX	70	5	0			
AVG	70	5	0			
1 MB FRAMES		FIXED			- PAGEABLE	
	TOTAL	AVAILABLE	IN-USE	TOTAL	AVAILABLE	IN-USE
MIN	0	0	0	7,056	2,496	4,560
MAX	0	0	0	7,056	2,496	4,560
AVG	0	0	0	7,056	2,496	4,560
HIGH SHARED FRAMES	TOTAL	CENTRAL	STORAGE		AUX DASD	AUX SCM
MIN	136902.1M		23,562		0	0
MAX	136902.1M		23,562		0	0
AVG	136902.1M		23,562		0	0
HIGH COMMON FRAMES	TOTAL	CENTRAL	STORAGE	FIXED 4K	AUX DASD	AUX SCM
MIN	17301504		16,493	1,718	0	0
MAX	17301504		16,493	1,718	0	0
AVG	17301504		16,493	1,718	0	0

Operator Command Support



d virtstor, hvcommon

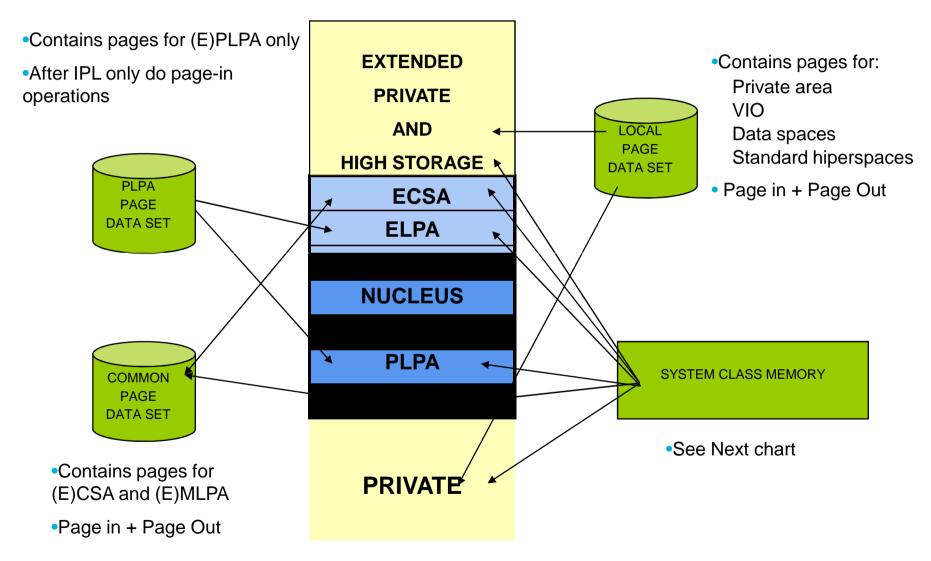
IAR019I SOURCE = DV TOTAL 64-BIT COMMON = 66G 64-BIT COMMON RANGE = 1982G-2048G 64-BIT COMMON ALLOCATED = 37M d virtstor,hvshare
IAR019I
SOURCE = DV
TOTAL SHARED = 66G
SHARED RANGE = 2048G-2112G
SHARED ALLOCATED = 0M

d virtstor,lfarea IAR019I SOURCE = DEFAULT TOTAL LFAREA = 900M , 0G LFAREA AVAILABLE = 1M , 0G LFAREA ALLOCATED (1M) = 608M LFAREA ALLOCATED (4K) = 0M MAX LFAREA ALLOCATED (1M) = 685M MAX LFAREA ALLOCATED (4K) = 0M LFAREA ALLOCATED (4K) = 0M LFAREA ALLOCATED (PAGEABLE1M) = 191M MAX LFAREA ALLOCATED (PAGEABLE1M) = 341M LFAREA ALLOCATED NUMBER OF 2G PAGES = 0 MAX LFAREA ALLOCATED NUMBER OF 2G PAGES = 0

If the high water mark for the number of fixed large pages used on behalf of 4K page requests is high decrease the LFAREA size or add additional real storage

Paging and Auxiliary Datasets







Storage Class Memory

Flash vs Disk Placement Criteria

Data Type	Data Page Placement
Pageable Link Pack Area (PLPA)	At IPL/NIP time PLPA pages will be placed both on flash and disk. PLPA data sets used for quick, and warm starts, SCM to resolve page faults.
VIO	VIO data will always be placed on disk (First to VIO accepting datasets with any spillover flowing to nonvio datasets)
HyperSwap Critical Address Space data	If flash space is available, all virtual pages belonging to a HyperSwap Critical Address Space will be placed on flash memory. If flash space is not available, these pages will be kept in memory and only paged to disk when the system is real storage constrained and no other alternatives exist
Pageable Large Pages	If contiguous flash space is available, pageable large pages will be preferentially written to flash.
All other data	If available space exists on both flash and disk then make a selection based on response time.





d	asm							
	SYS1	IEE200I						
	TYPE	FULL	STAT	DE	EV	DATASE	T NAM	Ξ
	PLPA	100%	FULL	028	E6	SYS1.P	LPA.P	AGCOM
	COMMON	60%	OK	028	E6	SYS1.C	OMMON	• PAGCOM
	LOCAL	0%	OK	048	31	SYS1.L	OCAL1	
	LOCAL	0%	OK	034	8	SYS1.L	OCAL7	
	LOCAL	0%	OK	048	BA	SYS1.L	OCAL8	
	SCM	12%	OK	N/	'A	N/A		
	PAGEDE	L COMMAN	ND IS	NOT	ACI	TIVE		

d asm, scm

SYS1 IEE207I

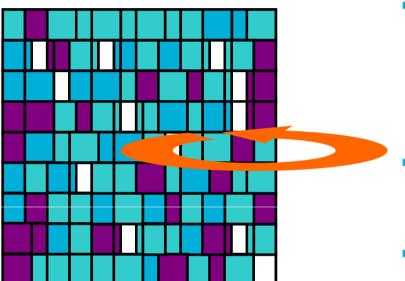
STATUS	FULL	SIZE	USED	IN-ERROR
IN-USE	12%	131,072	16,122	0

d m=scm

SYS1 IEE174I STORAGE-CLASS MEMORY STATUS 512M DEFINED ONLINE 0M-512M 0M OFFLINE-AVAILABLE SCM INCREMENT SIZE IS 1M

Measuring Central Storage Contention -UIC Calculation





- There are three different UIC's which can be displayed by performance monitors:
 - Current UIC
 - Minimum UIC
 - Maximum UIC

- UIC Unreferenced Interval Count
 - The higher the UIC value, the less contention for storage in the system
 - The lower the UIC value, the more contention for storage in the system
 - The page replacement algorithm z/OS uses was enhanced to more efficiently process large amounts of real storage
- Since z/OS 1.8, the UIC is defined as a single walk though all of central storage in seconds
- The UIC values seen in an RMF report will vary from 0 – 65535 (18 hours)
 - Values greater than 9999 are displayed as nnK

UIC Management



RMF Mon 3 – STORR or STORS panel

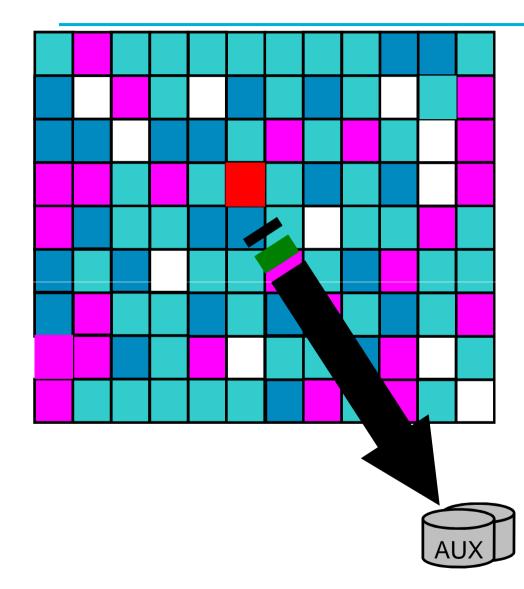
RMF	V2R1	Sto	rage 1	Resour		-	Stor	age	Summary		
			- % F:	rames				-	Frames	System	
NUC	SQA	CSA	LPA	ACTV	IDLE	AVA	IL SH	R	Online	UIC	
0	0	0	0	4	0		94	0	16749K	65535	
						-	_		_		
Volu	me DE	V	CU					PND	Pend	SPACE	- AVG Active Users-
Seri	al Ty	pe	Туре	e	PAV	8 8	%	%	Reasons	TYPE	TOTL LOCL SWAP COMM
ZOSP	G2 33	903	210'	7	1	0	0 0	0	None	LOCL	0.0 0.0 0.0 0.0

• RMF Postprocessor – REPORT(PAGING)

z/OS V2R1	SYSTEM ID SYSB	DATE 09/19/2014	INTERVAL 05.00.001
OPT = IEAOPT00 LFAREA	SIZE = 0 CENTRAL STORAGE	MOVEMENT AND REQUEST RATES	- IN PAGES PER
SYSTEM UIC: MIN = 65	535 MAX = 65535 AV	G = 65535	

Stealing in z/OS





- Pageable frame stealing is the process of taking an assigned central storage frame away from an address space to make it available for other purposes
- When there is a demand for pageable frames, RSM will steal frames which are <u>unreferenced</u> for a long time and return them to the system – (Available Frame Queue)
 - -No demand, then there is no stealing
- Since z/OS 1.8 stealing works against the entire storage range (Global LRU)
- Still protect frames in address spaces if Storage Critical or has a protective target



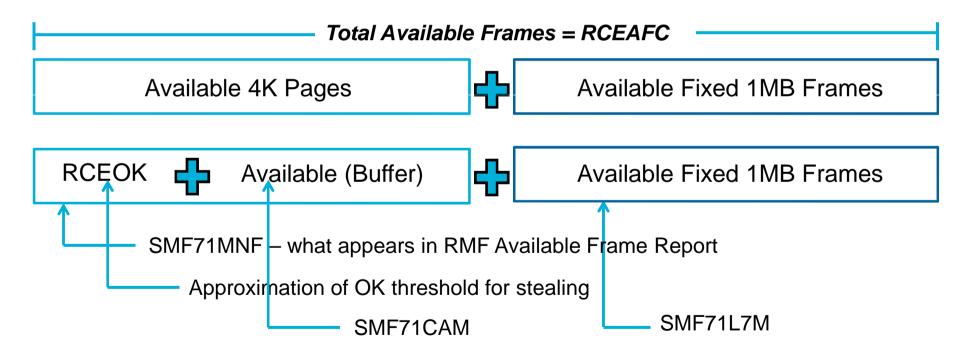
MCCAFCTH=(lowvalue,okvalue)

- Specifies the low and the OK threshold values for storage
- The *lowvalue* indicates the number of frames on the available frame queue when stealing begins
- The okvalue indicates the number of frames on the available frame queue when stealing ends
- SRM manages the queue and starts the stealing process when the number of frames falls below the threshold
- The defaults are:
- LOW will vary between MAX(MCCAFCTH lowvalue, 400, 0.2% of pageable storage)
- OK will vary between MAX(MCCAFCTH okvalue, 600, 0.4% of the pageable storage)
 - SRM will automatically adjust the actual threshold values based on measurements of storage usage but doesn't let values get lower than MCCAFCTH low threshold
- Typically no need to specify this parameter

Available Frame Calculations



- Available frames are tracked in RCEAFC
- New IEASYSxx parm INCLUDE1MAFC changes RCEAFC calculation by including Available 1MB Frames in count
- Min, Max, and Average values are tracked Minimums shown below



z/OS 2.2 will include Available Fixed 1MB Frames in both SMF71MNF and SMF71CAM



RMF Post Processor – REPORT(PAGING)

	FRAME	AND SLOT CO	UNTS			
(31 SAMPLES) OCENTRAL STORAGE FRAMES	TOTAL	AVAILABLE	SQA	LPA	CSA	LSQA
MIN	16748558	14620193	7,243	5,761	9,470	77,480
MAX	16748558	14620757	7,244	5,761	9,481	77,498
AVG	16748558	14620474	7,243	5,761	9,478	77 , 491

RMF Monitor 2 – SRCS panel

CPU= 1/	1 UI	C= 65	SK PR:	= 0		(L	Syste	m= SY	CSD TO	otal					
		HI	SQA	LPA	LPA	CSA	L+C	PRI	LSQA	LSQA (CPU	IN	OUT	OUT	OUT
TIME	AFC	UIC	F	F	FF	F	FF	FF	CSF	ESF U	JTL	Q	LOG	RQ	WQ
22:25:57	16M	65K	7.6K	22K	80	11K	12K	38K	77K		1	68	26	0	26
22:26:28	16M		7.6K		80		12K	38K			1	68	26	0	26
22:26:28	16M	65K	7.6K	22K	80	11K	12K	38K	77K		1	68	26	0	26
22:26:28	16M	65K	7.6K	22K	80	11K	12K	38K	77K		1	68	26	0	26

Demand Paging

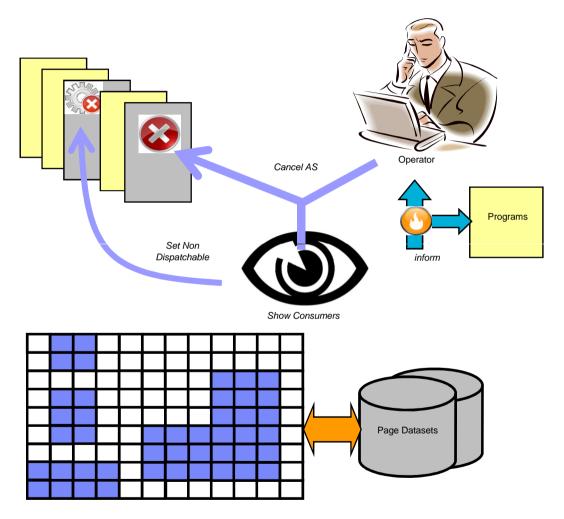


- Paging rates are counted in terms of 4KB pages even when moving 1MB Pageable pages to Flash
- Verify it's 1MB pageable pages by looking at I/O rates to Local Page Data Sets

OPT = IEAOPT01	LFAREA	SIZE =	C) CENTR	AL STOP	RAGE PAGIN	IG RATES	- IN PA	GES PER	SECOND
			PAGE IN				- PAGE	OUT		
		NON	SWAP	TOT	AL			TOT	'AL	
			NON				NON			
CATEGORY	SWAP	BLOCK	BLOCK	RATE	8	SWAP	SWAP	RATE	8	
TOTAL SYSTEM										
HIPERSPACE		0.00		0.00	0		0.00	0.00	0	
VIO		0.00		0.00	0		0.00	0.00	0	
NON-VIO	0.00	0.00	15,243	15,243	100	0.00	6,814	6,814	100	
SUM	0.00	0.00	15,243	15,243	100	0.00	6,814	6,814	100	
SHARED			0.00	0.00	*		0.00	0.00		
					· · · · · ·	~				
PAGE MOVEMENT WI	COMPANY AND	NTRAL STO	ORAGE	1,2	26.83	Page	In		Pa	age Out
PAGE MOVEMENT TI					0.0	Juge				ige out
AVERAGE NUMBER O		PER BLO	CK		0.0	/				
BLOCKS PER SECON				15.0	0.00	¥				
PAGE-IN EVENTS (PAGE FAU	ULT RATE))	15,2	43.46					

Storage Shortage Management





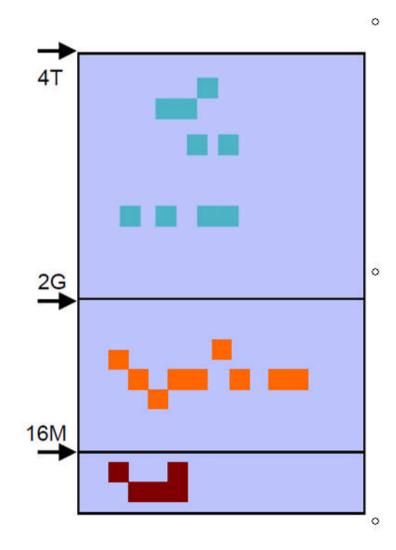
- Monitors
 - Fixed Storage consumption
 - Auxiliary Storage consumption
 - Every 2 seconds
- Informs in case of problems
 - Operator via messages
 - Programs via ENF55

Takes Actions

- To set Address Spaces non dispatchable
- To cancel address spaces on operator request

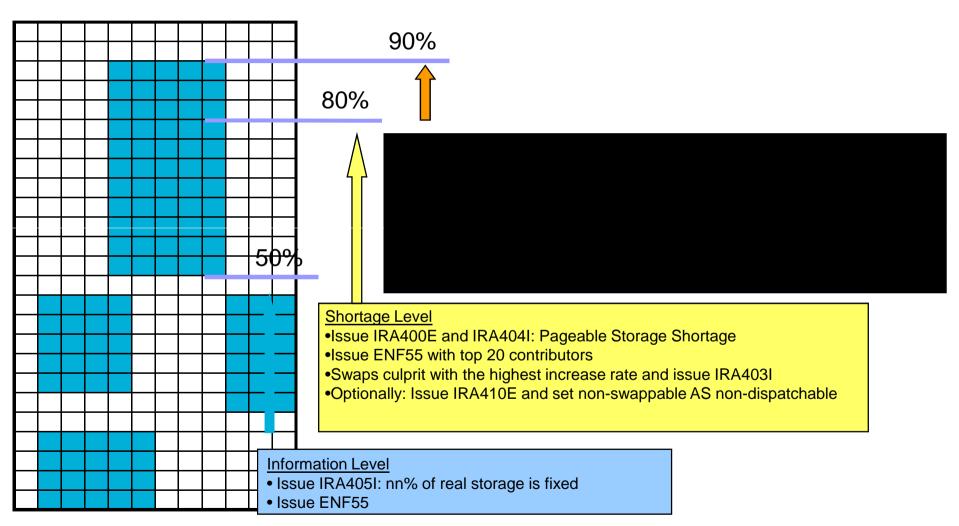
In-Real Swap

- Since z/OS 1.8 pageable storage shortages are handled by an In-Real Swap
- Frames in the shortage area are exchanged with other frames
- Pageable storage shortage between 16MB -> 2GB, frames will be exchanged with frames above 2GB
- Pageable storage shortages below 16MB, frames will be exchanged with frames above 2GB or above 16MB
- Message IRA404I is issues and lists the five largest users of fixed frames in the shortage area





Pageable Storage Shortages



Note: for Below 16M the shortage targets are 92% and 96%

Pageable Storage Shortages



IRA420I ! ## ! USER	! ASID ! PAGES	! O/W FIXED !
IRA420I ++	-++	++
IRA420I ! 01 ! I9A2GH11	! 0081 ! 0000065939	! 0000065851 !
IRA420I ! 02 ! I9A2GH10	! 0087N! 0000065936	! 0000065517 !
IRA420I ! 03 ! I9A2GH13	! 0088 ! 0000066002	! 0000062839 !
IRA420I ! 04 ! XCFAS	! 00065! 0000020870	! 0000001436 !
IRA420I ! 05 ! TRACE	! 00045! 0000008262	! 000000706 !
*73 IRA421D REPLY M FOR	MORE, E TO END, ## I	O CANCEL A USER

These messages get issued after the system runs into a critical storage shortage. When the operator enters a M, the next 5 address spaces get presented at the console. A reply of the line number (##) terminates the address space.

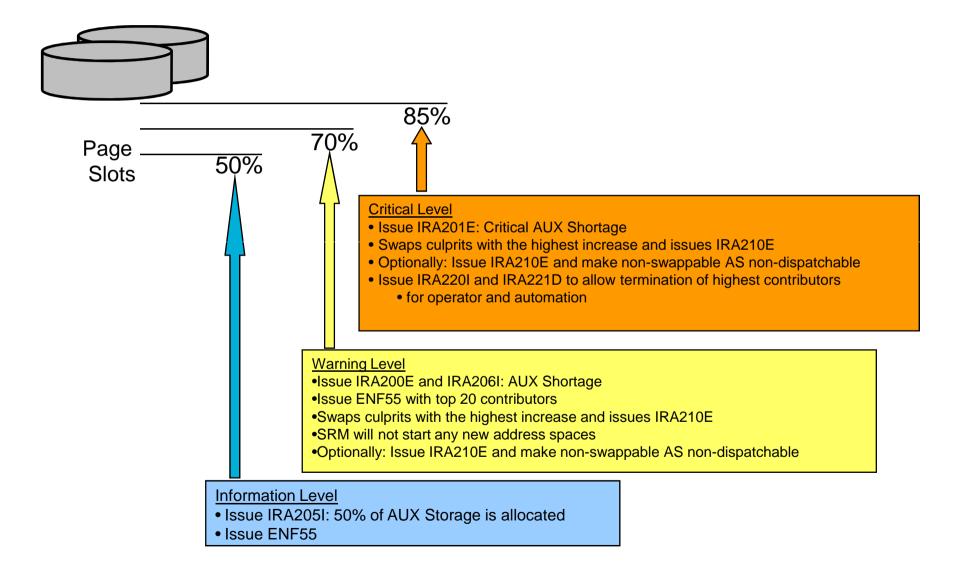
IEAOPTxx keyword to disable the WTOR IRA421D or change the number of rows in message IRA420I:

STORAGEWTOR=xxxx

Specifies how the system handles address spaces during a critical storage shortage. Where xxxx is either YES (default), NO or AUTO

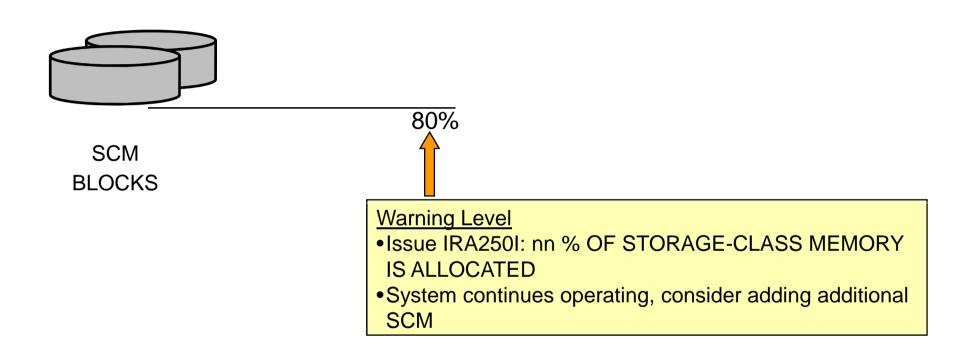
Auxiliary Storage Shortages





Auxiliary Storage Shortages - Details





Auxiliary Storage Shortages



IRA220I	!	##	!	USER	!	ASID !	PAGES	!	SLOTS	!
IRA220I	+-		•+•		• + •	+		-+-		+
IRA220I	!	01	!	I9A2GH11	!	0081 !	0000065939	!	0000065851	!
IRA220I	!	02	ļ	I9A2GH10	!	0087N!	0000065936	!	0000065517	!
IRA220I	!	03	ļ	I9A2GH13	!	0088 !	0000066002	!	0000062839	!
IRA220I	!	04	!	XCFAS	!	0006S!	0000020870	!	0000001436	!
IRA220I	!	05	ļ	TRACE	!	0004s!	0000008262	!	0000000706	!
*73 IRA2	221	DF	E]	PLY M FOR	M	ORE, E	TO END, ## :	0	CANCEL A US	ER

These messages get issued after the system runs into a critical storage shortage. When the operator enters a M, the next 5 address spaces get presented at the console. A reply of the line number (##) terminates the address space.

OPT keywords to disable the WTOR IRA221D or change the number of rows in message IRA220I:

STORAGEWTOR=xxxx

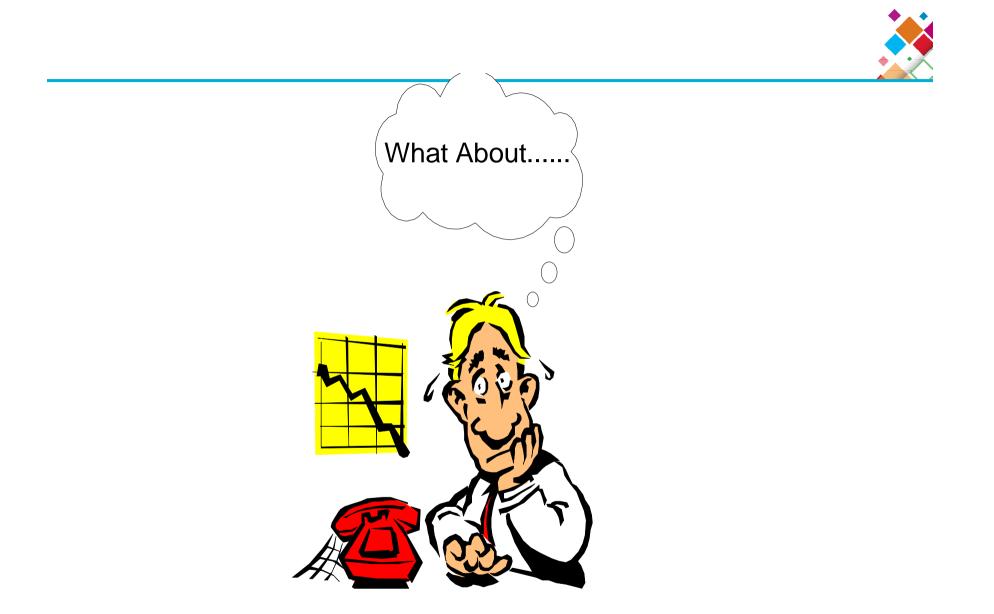
Specifies how the system handles address spaces during a critical storage shortage. Where *option* is either YES (default), NO or AUTO.

Storage Shortage Management: Summary



	AUX	Pageable Below 16M	Pageable Above 16M
Warning	50%	70%	50%
		IRA405I	IRA405I
Shortage	70%	92%	80%
Unonage	1078	MCCFXEPR	MCCFXTPR
Critical	050/	96%	90%
Shortage	85%	100 - (100- MCCFXEPR)/2	100 - (100- MCCFXTPR)/2

© Copyright IBM Corporation 2015



Trademarks



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

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by (are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System j, System p, System p5, System z, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. 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.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the 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.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.