



IBM Systems & Technology Group

z/VM System Limits

SHARE – Summer 2013 - Boston
Session 13512



Bill Bitner- bitnerb@us.ibm.com
z/VM Customer Focus and Care
[bitnerb @ us.ibm.com](mailto:bitnerb@us.ibm.com)

Trademarks

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml: AS/400, DBE, e-business logo, 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, zEC12, zBC12

The following are trademarks or registered trademarks of other companies

Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation
Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries
Linux is a registered trademark of Linus Torvalds
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.
Intel is a registered trademark of Intel Corporation
* 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.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use.

The 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.

Permission is hereby granted to SHARE to publish an exact copy of this paper in the SHARE proceedings. IBM retains the title to the copyright in this paper, as well as the copyright in all underlying works. IBM retains the right to make derivative works and to republish and distribute this paper to whomever it chooses in any way it chooses.

Agenda

- **Describe various limits**
 - Architected
 - Supported
 - Consumption
 - Latent
- **Show how to keep tabs on consumables**
- **Discuss limits that may be hit first**

Limits


- **Processors**
- **Memory**
- **I/O**
- **Others**
- **Latent limits**
- **Additional Disclaimer**
 - This presentation looks at individual limits, it is quite possible that you will hit one limit before you hit the next. We do it this way to help illustrate which limits Development will address first, but then to set expectations as to how much greater can one run before hitting that next limit.
 - This presentation talks about limits that are some times beyond the supported limits. This is meant to let the audience know what IBM did to determine where the supported limited should be and why it is the supported limit. It is not meant to imply it is safe to run up to that limit or that IBM knows everything that will go wrong if you do. So please stay at or below the supported limit.

Comments on Release Level

★ **z/VM 6.3 became Generally Available July 26, 2013**

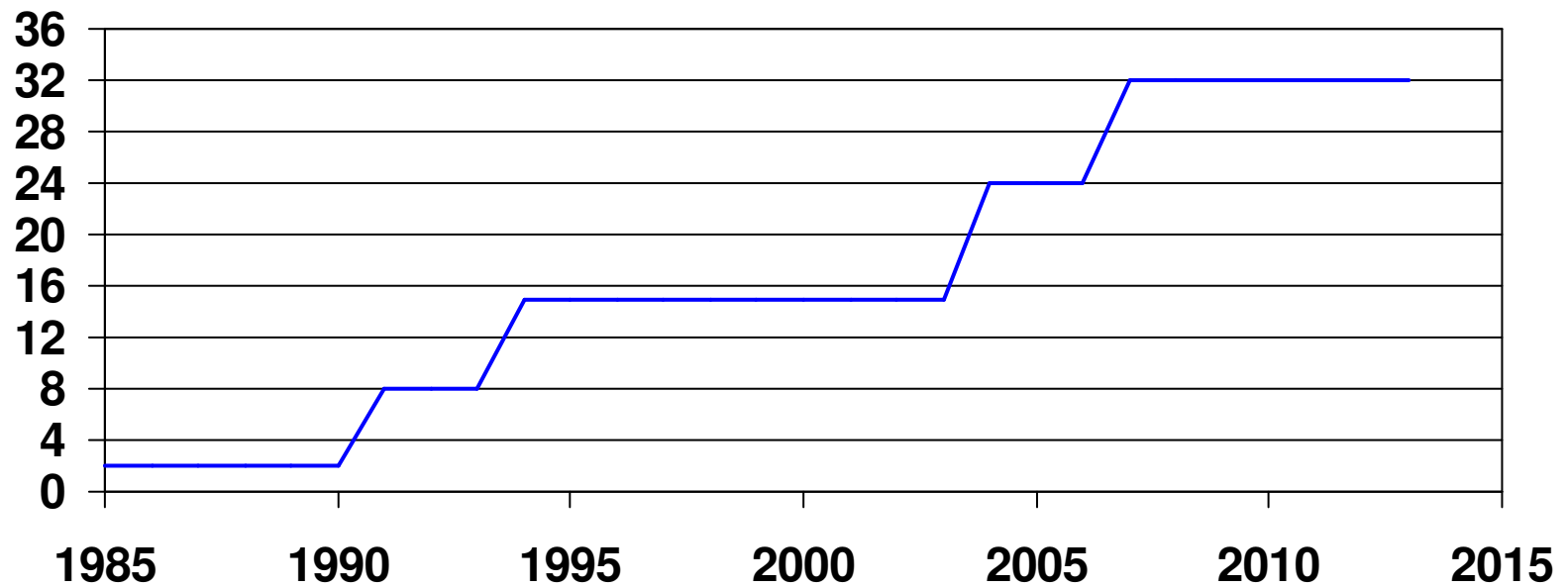
- **z/VM 6.1 went End of Service April 30, 2013 and is not called out in this presentation. In most cases, its limits are same as z/VM 5.4**

Processors

- **Processors (architected): 64**
 - Includes all engine types (CP, zAAP, zIIP, IFL...)
- **Processors (hardware – available to customer):**
 - z9: 54
 - z10: 64
 - z196: 80
 - zEC12: 101
- **Logical processors (unsupported): 64 (zEC12 & z10 EC); 54 (z9 EC)**
- **Logical processors in z/VM partition (support statement): 32**
- **Master processor (architected): 1**
 - 100%-utilized master is the issue
 - z/VM will elect a new master if master fails
 - In z/VM 6.3 Master may be reassigned to keep it as a Vertical High when running in Vertical Polarization Mode 
- **Virtual processors in single virtual machine (architected): 64**
 - But $N_{\text{Virtual}} > N_{\text{Logical}}$ is not usually practical
 - Interrupts presented to just 1 virtual CPU
- **Number of partitions: 60 (z9 through zEC12)**

Processor Scaling

Number of Supported Processors



Processors: FCX100 CPU

FCX100 Run 2007/09/06 14:00:28

CPU

General CPU Load and User Transactions

From 2007/09/04 09:07:00

To 2007/09/04 10:00:00

For 3180 Secs 00:53:00

CPU 2094-700

z/VM V.5.3.0 SLU 0701

CPU Load										Vector Facility		Status or		
PROC	TYPE	%CPU	%CP	%EMU	%WT	%SYS	%SP	%SIC	%LOGLD	%VTOT	%VEMU	REST	ded.	User
P00	IFL	16	2	14	84	2	0	84	16	
P15	IFL	18	2	16	82	1	0	80	18	
P14	IFL	18	2	16	82	1	0	80	18	
P13	IFL	18	2	16	82	1	0	80	18	
P12	IFL	18	2	16	82	1	0	81	18	
P11	IFL	18	2	17	82	1	0	80	19	
... truncated ...														

1. $T/V \sim 18/16 = 1.13$ a little CP overhead here
2. Master does not seem unduly burdened

Processors: FCX144 PROCLOG

Interval	C P	<---- Percent Busy ---->					<--- Rates per Sec.--->					<----- PLDV ----->				
		U	Type	Total	User	Syst	Emul	Inst Siml	DIAG	SIGP	SSCH	pty	Mean Non-0	VMDBK Mastr	VMDBK Stoln	To Mastr
>>Mean>>	0	CP	54.5	53.9	.6	50.4	5608	28.2	1588	155.1	47	1	0	423.5	.1	
>>Mean>>	1	CP	61.1	60.7	.5	56.8	6304	30.2	1481	161.5	99	1	421.4	.0	
>>Mean>>	2	CP	62.3	61.7	.5	57.7	6444	30.6	1475	160.8	97	1	418.7	.0	
>>Mean>>	3	CP	63.9	63.5	.4	59.5	6534	30.0	1453	153.4	99	1	395.8	.0	
>>Mean>>	4	CP	58.3	57.7	.6	54.2	5744	27.2	1520	152.1	99	1	442.8	.0	
>>Mean>>	5	CP	60.2	59.8	.4	56.2	5860	26.7	1457	141.5	99	1	402.7	.0	
>>Mean>>	6	CP	61.8	61.3	.4	57.4	6356	30.6	1552	156.7	99	1	418.9	.0	
>>Mean>>	7	CP	60.1	59.7	.4	55.9	6173	30.6	1554	156.3	98	1	413.3	.0	
>>Mean>>	.	CP	60.2	59.8	.4	55.9	6128	29.2	1510	154.6	92	1	417.1	.0	

For z/VM 6.2 and older.

Processors: FCX304 PRCLOG (New for z/VM 6.3)

FCX304 CPU 2827 SER 15D37 Interval 02:57:04 - 14:13:0

<--- Percent Busy ---->											
Interval	C	P									
End Time	U	Type	PPD	Ent.	DVID	Pct	Time	Total	User	Syst	Emul
>>Mean>>	0	CP	vh	100	0000	0	53.4	51.3	2.1	39.0	
>>Mean>>	1	CP	vh	100	MIX	0	70.2	69.0	1.3	62.8	
>>Mean>>	2	CP	vh	100	MIX	0	63.9	62.5	1.4	55.9	
>>Mean>>	3	CP	vh	100	MIX	0	61.6	60.4	1.1	54.5	
>>Total>	4	CP	vh	400	MIX	0	249.1	243.2	5.8	212.2	
13:53:07	0	CP	vh	100	0000	0	98.3	96.3	2.0	68.6	
13:53:07	1	CP	vh	100	0001	0	97.4	95.8	1.6	81.6	
13:53:07	2	CP	vh	100	0002	0	97.4	95.6	1.8	80.9	
13:53:07	3	CP	vh	100	0003	0	97.3	95.7	1.6	81.4	

Processors: FCX114 USTAT

FCX114 Run 2007/09/06 14:00:28

USTAT

Page 186

wait State Analysis by User

From 2007/09/04 09:07:00

To 2007/09/04 10:00:00

For 3180 Secs 00:53:00

CPU 2094-700

z/VM V.5.3.0 SLU 0701

Userid	%ACT	%RUN	%CPU	%LDG	%PGW	%IOW	%SIM	<-SVM and->					%DM	%IOA	%PGA	%LIM	%OTH	<--%Time spent in-->					Nr of Users
								%TIW	%CFW	%TI	%EL	Q0						Q1	Q2	Q3	E0-3		
>System<	64	1	0	1	0	0	0	83	0	0	0	3	0	0	0	10	1	29	10	57	0	211	
TCPIP	100	0	0	0	0	0	0	0	0	3	0	97	0	0	0	0	3	0	0	0	0	0	
RSCSDNS1	100	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	
SNMPD	100	0	0	0	0	0	0	0	0	2	0	98	0	0	0	0	2	0	0	0	0	0	
SZVAS001	100	2	0	0	0	0	0	97	0	0	0	0	0	0	0	1	0	3	12	85	0	0	

1. %CPU wait is very low – nobody is starved for engine
2. %TIW is “test idle wait” – we are waiting to see if queue drop happens

Memory – Part 1

■ Central storage

- Supported central storage:
 - 256 GB with z/VM 5.4 and 6.2
 - 1TB with z/VM 6.3 ★
- Maximum LPAR size:
 - z9: 512 GB minus your HSA
 - z10 to zEC12: 1 TB

■ Expanded storage (architected): 16TB

- z/VM Limit: 128GB supported
 - Up to about 660GB unsupported (depends on other factors)
- See <http://www.vm.ibm.com/perf/tips/storconf.html>
- In z/VM 6.3, recommend configuring expanded storage as central storage. ★

Memory – Part 2


■ Virtual machine size:

- Supported/Tested 1 TB (2^{40})
 - Practical limit can be gated by:
 - Reorder Processing (z/VM 5.4 & 6.2)
 - VM Dump
 - Reorder Gone in z/VM 6.3 ★
 - Production level performance will require adequate real memory
- Hardware limits
 - zEC12 & zBC12 16TB
 - z196 & z114 16TB
 - z10 8TB
 - z9 1TB
 - z990 256GB
 - z900 256GB

Memory – Part 3

- **Active, or instantiated, guest real limit imposed by PTRM space limits (architected): 8 TB (64 TB with z/VM 6.3) ★**
 - 16 4-GB PTRM spaces; each PTRM space can map 512 GB of guest real
 - z/VM 6.3 – 128 PTRM all pre-allocated. ★
- **Virtual to real ratio (practical): about 2:1 or 3:1**
 - Warning: Different people have different definitions for “Virtual to real memory”. Here we are using total virtual machine size of started virtual machines to central storage.
 - 1:1 if you want to eliminate performance impact for production workloads.
 - As you get above 2:1, you really need to do your homework on your paging subsystem
 - Many factors come into play here, including:
 - Active:Idle Virtual machines
 - Workload/SLA sensitivity to delays
 - Exploitation of shared memory & other memory management techniques (e.g. CMM)

Memory – Part 4

- **z/VM 5.4 & 6.2: Paging space (architected) (optimal when $\leq 50\%$ allocated):**
 - 11.2 TB for ECKD (3390)
 - 15.9 TB for Emulated FBA on FCP SCSI (EDEV)
- **z/VM 6.3 Paging space will not require the $<50\%$ allocated,  though we will recommend some “buffer” space.**
- **255 CP Owned Volumes**
 - Above numbers based on using all of those for Paging
- **Do NOT mix ECKD and EDEV paging volumes on same system**
 - Various anomalies can occur
- **Concurrent paging I/Os per paging volume: 1 for ECKD, >1 for EDEV (Have observed 1.6)**

Memory – Part 5

- **System Execution Space (SXS) (architected): 2 GB**
 - For practical purposes it is 2GB, but there are structures in the space placed above 2GB
- **DCSS aggregate size (architected):**
 - Individual Segments up to 2047 MB
 - Segments must end prior to one 4KB page below 512GB
- **Minidisk Cache (architected): 8GB**
 - Practical 2GB
- **Installing z/VM: 2GB**
 - On some machines, there is a problem with having more than 2GB Central when doing the initial install of z/VM off the DVD.

Memory References

- **Memory Over Commit**

- <http://www.vm.ibm.com/perf/tips/memory.html>

- **Paging in General**

- <http://www.vm.ibm.com/perf/tips/prgpage.html>

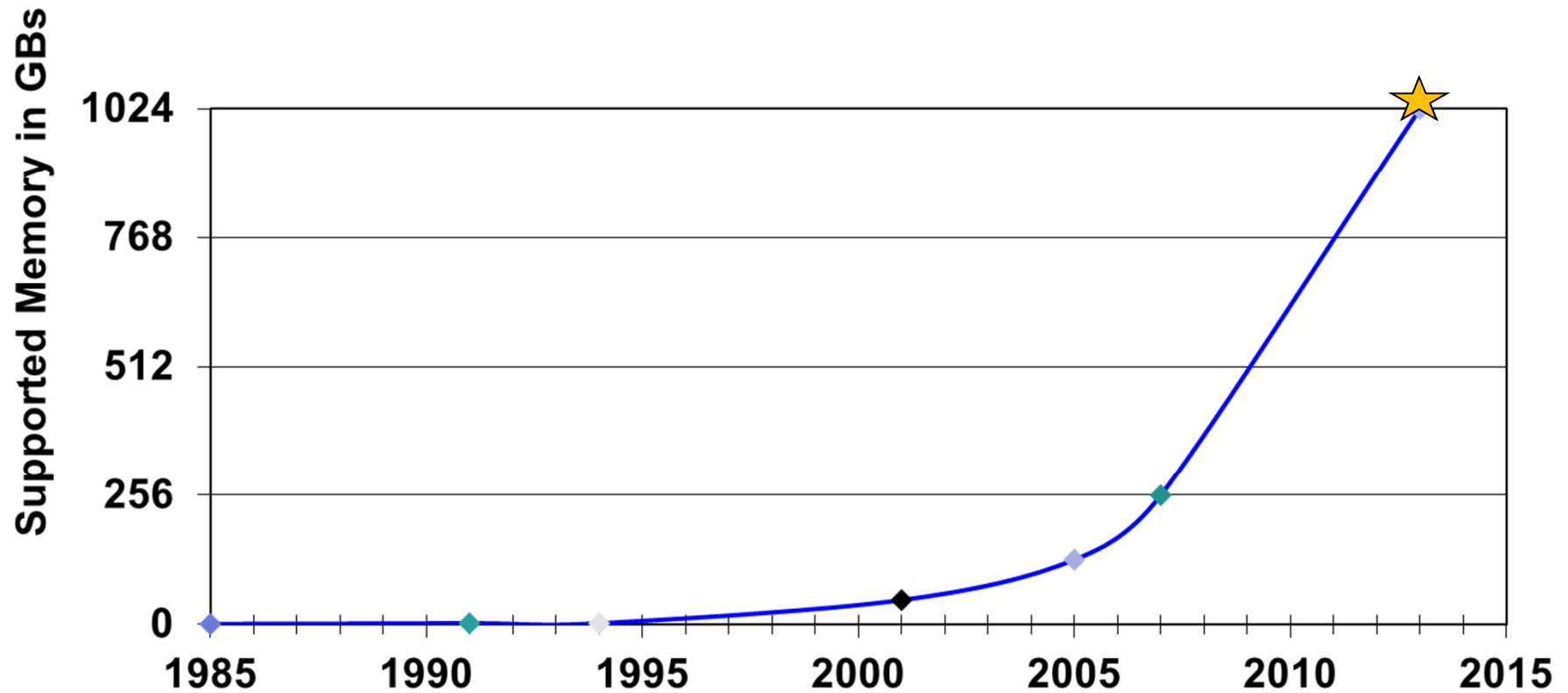
- **Reorder Processing**

- <http://www.vm.ibm.com/perf/tips/reorder.html>

- Goes away with z/VM 6.3 ★

Memory Scaling

Effective Real Memory Use Limits



Page Slots: FCX146 AUXLOG

FCX146 Run 2007/09/06 14:00:28

AUXLOG

Auxiliary Storage Utilization, by Time

From 2007/09/04 09:07:00

To 2007/09/04 10:00:00

For 3180 Secs 00:53:00

Interval End Time	<Page Slots>		<Spool Slots>		<Dump Slots>		<----- Spool Files ----->				<Average MLOAD>	
	Total Slots	Used %	Total Slots	Used %	Total Slots	Used %	<--Created-->		<--Purged-->		Paging msec	Spooling msec
>>Mean>>	87146k	44	5409096	52	0	..	54	.02	54	.02	2.8	.8
09:08:00	87146k	44	5409096	52	0	..	1	.02	1	.02	2.3	.8
09:09:00	87146k	44	5409096	52	0	..	1	.02	1	.02	3.9	.8
09:10:00	87146k	44	5409096	52	0	..	1	.02	1	.02	3.6	.8
09:11:00	87146k	44	5409096	52	0	..	1	.02	1	.02	2.8	.8
09:12:00	87146k	44	5409096	52	0	..	1	.02	1	.02	2.9	.8

1. This system is using 44% of its page slots.



DASD I/O: FCX109 DEVICE CPOWNED

FCX109 Run 2007/09/06 14:00:28

DEVICE CPOWNED

Page 152

Load and Performance of CP Owned Disks

From 2007/09/04 09:07:00

To 2007/09/04 10:00:00

For 3180 Secs 00:53:00

CPU 2094-700

z/VM V.5.3.0 SLU 0701

Page / SPOOL Allocation Summary

PAGE slots available	87146k	SPOOL slots available	5409096
PAGE slot utilization	44%	SPOOL slot utilization	52%
T-Disk cylinders avail.	DUMP slots available	0
T-Disk space utilization	...%	DUMP slot utilization	..%

< Device Descr. ->		<----- Rate/s ----->										User		Serv		MLOAD		Block	
%Used																			
Addr	Devtyp	Volume Serial	Area Type	Area Extent	Used %	<--Page-->		<--Spool-->		Total	SSCH +RSCH	Inter feres	Queue Length	Time /Page	Resp Time	Page Size	for Alloc		
F08B	3390	VS2P49	PAGE	0-3338	45	2.6	1.7	4.4	1.6	1	.02	2.4	2.4	7	89		
F090	3390	VS2P69	PAGE	0-3338	45	2.7	1.6	4.3	1.6	1	0	2.7	2.7	7	84		

V:R Ratio: FCX113 UPAGE

Userid	<----- Paging Activity/s ----->							<----- Number of Pages ----->							Stor	Nr of
	<Page Rate>		Page	<-Page Migration-->				WSS	<-Resident-->		<--Locked-->			Size		
>System<	Reads	Write	Steals	>2GB>	X>MS	MS>X	X>DS		R<2GB	R>2GB	L<2GB	L>2GB	XSTOR	DASD		
>System<	1.7	1.1	4.1	.0	2.4	3.7	1.4	122050	2347	106962	6	24	12240	179131	1310M	212
DATAMOVF	.0	.0	.0	.0	.0	.1	.0	13	0	0	0	0	483	254	32M	
DATAMOVA	.0	.0	.0	.0	.5	.5	.0	147	0	0	0	0	220	368	32M	
DATAMOVB	.0	.0	.0	.0	.6	.6	.0	192	0	0	0	0	220	366	32M	
DATAMOVC	.0	.0	.0	.0	.6	.6	.0	191	0	0	0	0	220	369	32M	
DATAMOVD	.0	.0	.0	.0	.6	.6	.0	189	0	0	0	0	220	362	32M	

1. Resident Guest Pages = (2347 + 106962) * 212 = 88.3 GB
2. V:R = (1310 MB * 212) / 91 GB = 2.98
3. For z/VM 6.2 and older

Report FCX292 UPGUTL is new for z/VM 6.3

FCX292 Run 2013/04/10 07:38:36

UPGUTL
User Page Utilization Data

Page 103

From 2013/04/09 16:02:10
To 2013/04/09 16:13:10
For 660 Secs 00:11:00

SYSTEMID
CPU 2817-744 SN A6D85
z/VM V.6.3.0 SLU 0000

"This is a performance report for SYSTEM XYZ"

Storage																				
Resident																		Base		
Invalid But Resident																		Space		Nr of
AgeList																		Size		Users
Data	Owned	WSS	Inst	Resvd	T_All	T<2G	T>2G	L<2G	L>2G	U<2G	U>2G	P<2G	P>2G	A<2G	A>2G	XSTOR	AUX	Size	Users	
>>Mean>>	.0	5284M	6765M	5611	5286M	27M	5259M	1010	232K	6565	2238K	59588	26M	53080	107M	.0	1815M	7108M	73	
User Class Data:																				
CMS1_USE	.0	3320K	19M	.0	484K	.0	484K	.0	4096	.0	69632	.0	244K	.0	344K	.0	19M	2047M	1	
LCC_CLIE	.0	364M	485M	.0	365M	11264	365M	.0	208K	.0	325K	.0	2686K	.0	8177K	.0	164M	1024M	8	
LXA_SERV	.0	7974M	10G	.0	7978M	41M	7937M	.0	206K	9984	3327K	90624	39M	80725	161M	.0	2719M	10240M	48	
User Data:																				
DISKACNT	.0	4976K	5156K	0	4K	0	4K	0	0	0	4K	0	0	0	0	0	5152K	32M		
DTCVSW1	.0	184K	11M	0	196K	8K	188K	8K	4K	0	4K	0	0	0	168K	0	11M	32M		
DTCVSW2	.0	180K	11M	0	184K	0	184K	0	4K	0	4K	0	0	0	164K	0	10M	32M		
EREP	.0	4912K	4944K	0	4K	0	4K	0	0	0	4K	0	0	0	0	0	4940K	32M		
FTPSEVE	.0	84K	5764K	0	88K	0	88K	0	4K	0	4K	0	0	0	76K	0	5760K	32M		
GCSXA	.0	204K	208K	0	8K	0	8K	0	4K	0	4K	0	0	0	0	0	200K	16M		
LCC00001	.0	364M	488M	0	365M	0	365M	0	204K	0	228K	0	2884K	0	8660K	0	192M	1024M		
LCC00002	.0	369M	492M	0	371M	20K	371M	0	204K	0	224K	0	2312K	0	7736K	0	159M	1024M		
LCC00003	.0	363M	484M	0	364M	0	364M	0	204K	0	252K	0	2852K	0	8372K	0	215M	1024M		
LCC00004	.0	363M	483M	0	363M	16K	363M	0	204K	0	228K	0	2724K	0	8512K	0	185M	1024M		

- Look for the new concepts: Inst IBR UFO PNR AgeList
- Amounts are in bytes, suffixed. Not page counts!
- FCX113 UPAGE is still produced.

Zoom in on FCX292 UPGUTL report new for z/VM 6.3

```

----- Storage -----
<----- Resident ----->
<----- Invalid But Resident ----->
<----- Total -----> <-Locked--> <-- UFO --> <-- PNR --> <-AgeList->
Inst  Resvd  T_All  T<2G  T>2G  L<2G  L>2G  U<2G  U>2G  P<2G  P>2G  A<2G  A>2G  XSTOR  AUX
6765M 5611 5286M  27M 5259M 1010 232K 6565 2238K 59588 26M 53080 107M  .0 1815M

 19M  .0 484K  .0 484K  .0 4096  .0 69632  .0 244K  .0 344K  .0 19M
485M  .0 365M 11264 365M  .0 208K  .0 325K  .0 2686K  .0 8177K  .0 164M
10G   .0 7978M 41M 7937M  .0 206K 9984 3327K 90624 39M 80725 161M  .0 2719M 1
    
```

- Look for the new concepts: Inst IBR UFO PNR AgeList
- Amounts are in bytes, suffixed. Not page counts!
- FCX113 UPAGE is still produced.

Report FCX290 UPGACT is new for z/VM 6.3

FCX290 Run 2013/04/10 07:38:36

UPGACT
User Page Activity

Page 102

From 2013/04/09 16:02:10
To 2013/04/09 16:13:10
For 660 Secs 00:11:00

SYSTEMID
CPU 2817-744 SN A6D85
z/VM V.6.3.0 SLU 0000

"This is a performance report for SYSTEM XYZ"

<----- Storage ----->														
<----- Movement/s ----->														
Stl	<--- Transition/s --->					<-Steal/s->			<Migrate/s>				Nr of	
Userid	Wt	Inst	Relse	Inval	Reval	Ready	NoRdy	PGIN	PGOUT	Reads	Write	Mwrit	Xrel	Users
>>Mean>>	1.0	143K	5142	849K	718K	999K	.0	.0	.0	958K	761K	.0	.0	73
User Class Data:														
CMS1_USE	1.0	15515	15801	2377	1632	5145	.0	.0	.0	.0	1980	.0	.0	1
LCC_CLIE	1.0	658K	20875	488K	486K	60875	.0	.0	.0	54212	22869	.0	.0	8
LXA_SERV	1.0	108K	1095	1191K	994K	1506K	.0	.0	.0	1447K	1153K	.0	.0	48
User Data:														
DISKACNT	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0
DTCVSW1	1.0	0	0	3072	2855	0	0	0	0	0	0	0	0	0
DTCVSW2	1.0	0	0	3004	2780	0	0	0	0	0	0	0	0	0
EREP	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0
FTPSERVE	1.0	0	0	1434	1434	0	0	0	0	0	0	0	0	0
GCSXA	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0
LCC00001	1.0	601K	18686	501K	498K	65139	0	0	0	49866	23670	0	0	0
LCC00002	1.0	657K	24955	487K	486K	54725	0	0	0	44522	18991	0	0	0
LCC00003	1.0	565K	23012	485K	481K	64065	0	0	0	44783	19859	0	0	0
LCC00004	1.0	602K	24104	499K	495K	63178	0	0	0	48811	24588	0	0	0
LCC00005	1.0	717K	25675	500K	499K	65865	0	0	0	66002	28753	0	0	0

- Look for the new concepts: Inst Relse Inval Reval Ready NoRdy

Zoom in on Report FCX290 UPGACT new for z/VM 6.3



FCX290 Run 2013/04/10 07:38:36

UPGACT
User Page Activity

Page 102

From 2013/04/09 16:02:10
To 2013/04/09 16:13:10
For 660 Secs 00:11:00

"This is a performance report for SYSTEM XYZ"

SYSTEMID
CPU 2817-744 SN A6D85
z/VM V.6.3.0 SLU 0000

Stl	<--- Transition/s ---->				<-Steal/s->		
Userid	Wt	Inst	Relse	Inval	Reval	Ready	NoRdy
>>Mean>>	1.0	143K	5142	849K	718K	999K	.0
User Class Data:							
CMS1_USE	1.0	15515	15801	2377	1632	5145	.0
LCC_CLIE	1.0	658K	20875	488K	486K	60875	.0
LXA_SERV	1.0	108K	1095	1191K	994K	1506K	.0
LCC00001	1.0	601K	18686	501K	498K	65139	0
LCC00002	1.0	657K	24955	487K	486K	54725	0
LCC00003	1.0	565K	23012	485K	481K	64065	0
LCC00004	1.0	602K	24104	499K	495K	63178	0
LCC00005	1.0	717K	25675	500K	499K	65865	0

- Look for the new concepts: Inst Relse Inval Reval Ready NoRdy



PTRM Space: FCX134 DSPACESH

Data Space Name	<----- Rate per Sec. ----->						<-----Number of Pages----->									
	Pgstl	Pgrds	Pgwrt	X-rds	X-wrt	X-mig	Total	Resid	R<2GB	Lock	L<2GB	Count	Lockd	XSTOR	DASD	
-----	.075	.093	.015	.043	.074	.022	147k	1842	93	0	0	0	0	75	2998	
FULL\$TRACK\$CACHE\$1	.000	.000	.000	.000	.000	.000	524k	0	0	0	0	0	0	0	0	
ISFCDATASPACE	.000	.000	.000	.000	.000	.000	524k	112	74	100	74	112	100	0	41	
PTRM0000	14.79	1.733	.752	14.05	14.43	.039	1049k	596k	30116	0	0	0	0	5879	54074	
REAL	.000	.000	.000	.000	.000	.000	40M	0	0	0	0	0	0	0	0	
SYSTEM	.023	.000	.037	.019	.023	.004	524k	41	1	0	0	41	0	17	6410	

1. PTRM space = (596,000 + 5879 + 54,074) = 655,953 = 2.5 GB

Real Memory: FCX254 AVAILLOG

FCX254 Run 2007/09/06 14:00:28

AVAILLOG

Page 190

Available List Management, by Time

From 2007/09/04 09:07:00

To 2007/09/04 10:00:00

For 3180 Secs 00:53:00

CPU 2094-700

z/VM V.5.3.0 SLU 0701

Available List Management																				
Interval	Thresholds				Page Frames						Times		Replenishment						Perct	
	<Low	>2GB	<2GB	>2GB	<Available	<Obtains/s	<Returns/s	<Empty	<Scan1	<Scan2	<Em-Scan	Scan	Emerg	Comp1	Pages	Comp1	Pages	Comp1	Pages	Fail
>>Mean>>	20	7588	5820	13388	5130	7678	323.3	857.4	311.5	844.8	0	0	27	1381k	63	1380k	58	84490	82	88
09:08:00	20	7680	5820	13480	6665	15122	353.3	838.5	353.2	1007	0	0	0	43091	3	26491	0	0	3	100
09:09:00	20	7680	5820	13480	3986	5496	163.1	640.2	108.9	442.7	0	0	1	14528	0	0	0	0	0	0
09:10:00	20	7681	5820	13481	6622	9542	222.4	556.1	257.0	598.3	0	0	0	30103	2	8868	0	0	1	100
09:11:00	20	7681	5820	13481	4982	6710	292.1	615.2	248.8	533.6	0	0	0	21246	0	8547	1	3989	1	100
09:12:00	20	7681	5820	13481	4769	1560	284.9	946.9	254.4	830.0	0	0	0	18253	0	22438	2	656	1	100

1. Pct ES = 88% generally this system is tight on storage
2. Scan fail >0 generally this system is tight on storage
3. Times Empty = 0 this indicates it isn't critical yet (you do not need to wait for things to be critical).
4. Meant for z/VM 6.2 and older.

Report FCX295 AVLA2GLG is new for z/VM 6.3

FCX295 Run 2013/04/10 07:38:36

AVLA2GLG

Available List Data Above 2G, by Time

From 2013/04/09 16:02:10

To 2013/04/09 16:13:10

For 660 Secs 00:11:00

"This is a performance report for SYS

Interval	<----- Storage ----->						<--Times-->		<-Frame Thresh-->		
	<Available>		<Requests/s>		<Returns/s>		<-Empty/s-->		Sing	<-Contigs-->	
End Time	Sing	Cont	Sing	Cont	Sing	Cont	Sing	Cont	Low	Low	Prot
>>Mean>>	23M	267M	47M	59M	47M	51M	.0	.0	1310	15	15
16:02:40	0	938M	32M	126M	502K	30310	.0	.0	1332	15	15
16:03:10	152K	4556K	50M	89M	49M	59M	.0	.0	1168	15	15

- Times Empty/s should be zero
- FCX254 AVAILLOG is no longer produced in z/VM 6.3

SXS Space: FCX261 SXSAVAIL

FCX261 Run 2007/09/06 14:00:28

SXSAVAIL

Page 261

System Execution Space Page Queues Management

From 2007/09/04 09:07:00

To 2007/09/04 10:00:00

For 3180 Secs 00:53:00

CPU 2094-700

z/VM V.5.3.0 SLU 0701

Interval	<-- Backed <2GB Page Queue -->					<-- Backed >2GB Page Queue -->					<----- Unbacked Page Queue ----->									
	Avail	<-Pages/s-->	<Preferred>	Pages Taken	Return	Avail	<-Pages/s-->	<Preferred>	Pages Taken	Return	Used	Empty	Pages	Taken	Return	Used	Empty	Thres	Att/s	Stolen
>>Mean>>	26	.513	.509	.513	.000	3	1.798	1.804	1.798	4.114	466946	130.3	130.1	126.2	.000	128	.000	128	...	
09:08:00	26	.483	.383	.483	.000	0	1.650	1.650	1.650	3.667	467829	128.2	127.3	124.5	.000	128	.000	128	...	
09:09:00	26	.500	.500	.500	.000	0	.583	.583	.583	3.067	465679	120.8	84.98	117.8	.000	128	.000	128	...	
09:10:00	27	.517	.533	.517	.000	0	1.183	1.183	1.183	4.000	467657	109.1	142.1	105.1	.000	128	.000	128	...	
09:11:00	27	.517	.517	.517	.000	0	1.633	1.633	1.633	2.917	467632	137.2	136.8	134.3	.000	128	.000	128	...	
09:12:00	29	.450	.483	.450	.000	0	2.000	2.000	2.000	3.383	467654	129.9	130.2	126.5	.000	128	.000	128	...	
09:13:00	27	.517	.483	.517	.000	0	2.483	2.483	2.483	3.550	467698	139.3	140.0	135.7	.000	128	.000	128	...	
09:14:00	25	.550	.517	.550	.000	0	2.000	2.000	2.000	2.750	465651	119.0	84.92	116.3	.000	128	.000	128	...	

1. How we touch guest pages: (1) 64-bit; (2) AR mode; (3) SXS.
2. There are 524,288 pages in the SXS.
3. This system has 466,000 SXS pages available on average.

MDC: FCX178 MDCSTOR

```

<----- Main Storage Frames ----->
Interval          <--Actual-->   Min    Max  Page    Steal
End Time         Ideal  <2GB  >2GB   Set    Set Del/s  Invokd/s  Bias
>>Mean>>       5839k  82738 1354k    0  7864k    0    .000    1.00
09:57:41        5838k 119813 1932k    0  7864k    0    .000    1.00
09:58:11        5838k 119813 1932k    0  7864k    0    .000    1.00
09:58:41        5838k 119825 1932k    0  7864k    0    .000    1.00
09:59:11        5838k 119825 1932k    0  7864k    0    .000    1.00
09:59:41        5838k 119825 1932k    0  7864k    0    .000    1.00
10:00:11        5838k 119837 1932k    0  7864k    0    .000    1.00
    
```

- Xstore not used for this configuration so edited out from report.
- Add up the pages in Main Storage and you get ~8GB

MDC Spaces: FCX134 DSPACESH

Owning		Users	<-----Number of Pages----->									
userid	Data Space Name	Permt	Total	<--Resid--> Resid	R<2GB	<-Locked--> Lock	L<2GB	<-Aliases--> Count	Lockd	XSTOR	DASD	
>System<	-----	0	1507k	5665	101	0	0	100	0	0	0	
SYSTEM	FULL\$TRACK\$CACHE\$1	0	524k	0	0	0	0	0	0	0	0	
SYSTEM	FULL\$TRACK\$CACHE\$2	0	524k	0	0	0	0	0	0	0	0	
SYSTEM	FULL\$TRACK\$CACHE\$3	0	524k	0	0	0	0	0	0	0	0	
SYSTEM	FULL\$TRACK\$CACHE\$4	0	524k	0	0	0	0	0	0	0	0	
SYSTEM	ISFCDATASPACE	0	524k	0	0	0	0	0	0	0	0	
SYSTEM	PTRM0000	0	1049k	44489	0	0	0	0	0	0	0	
SYSTEM	REAL	0	7864k	0	0	0	0	0	0	0	0	
SYSTEM	SYSTEM	0	524k	805	787	0	0	800	0	0	0	
SYSTEM	VIRTUAL\$FREE\$STORAGE	0	524k	23	23	0	0	0	0	0	0	

- You'll see the address spaces used for MDC (track cache)
- Values here are zero for page counts, ignore.
- More than one FULL\$TRACK\$CACHE\$# space should be investigated to see if it the MDC settings are higher than needed.

I/O

- **Number of subchannels in a partition (aka device numbers) (architected): 65,535**
- **Device numbers per disk volume**
 - Without PAV, 1
 - With PAV or HyperPAV, 8 (base plus seven aliases)
- **Virtual Devices per Virtual Machine:**
 - 24576 (24K)
- **Concurrent real I/Os per ECKD disk volume: 1 usually, but 8 with PAV or HyperPAV if of guest origin**
- ★ ■ **GDPS Environments can have secondary devices defined in SSID with Multiple subchannel set support.**

I/O: DASD Volume Sizes

- **ECKD minidisk for a CMS file system:**
 - 32768 cylinders (22.5 GB)
 - 65520 cylinders (~45 GB) with CMS EAV APAR VM64711
- **Largest EFBA minidisk for a CMS file system: 381 GB**
 - Practical limit of 22GB due to file system structure under 16MB, unless there are very few files.
- **Largest ECKD volume:**
 - 65536 cylinders (45 GB)
 - 262,668 cylinders (~180 GB) with EAV APAR VM64709
 - CP use limited to first 64K cylinders
- **Largest EDEV CP can use: 1024 GB (but PAGE, SPOL, DRCT must be below 64 GB line on volume)**
- **Largest EDEV, period: 2³² FB-512 blocks (2048 GB)**

I/O

- **VDISK size (architected): 2 GB (minus eight 512-byte blocks)**
- **Total VDISK (architected): 2TB**
- **Single VSWITCH OSAs: 8**
- **Real HiperSockets VLAN IDs: 4096**



DASD I/O: FCX108 DEVICE

FCX108 Run 2007/09/06 14:00:28

DEVICE

Page 110

General I/O Device Load and Performance

From 2007/09/04 09:07:00

To 2007/09/04 10:00:00

For 3181 Secs 00:53:01

CPU 2094-700 SN

z/VM V.5.3.0 SLU 0701

<-- Device Descr. -->		Mdisk Pa-	<-Rate/s->				<----- Time (msec) ----->					Req.	<Percent>		SEEK	Recov	<-Throttle->		
Addr	Type	Label/ID	Links	ths	I/O	Avoid	Pend	Disc	Conn	Serv	Resp	CUwt	Qued	Busy	READ	Cyls	SSCH	Set/s	Dly/s
>>	All	DASD	<<5	.4	.2	.1	3.4	3.7	3.7	.0	.0	0	17	1173	00
F024	3390	VS2426	1	4	12.9	147.0	.2	.7	.4	1.3	1.3	.0	.0	2	91	193	0
0C20	CTCA		...	1	12.63	.2	.6	1.1	1.1	.0	.0	1	0
F685	3390	VS2W01	290	4	11.8	.3	.2	.0	.3	.5	.5	.0	.0	1	84	89	0
F411	3390	VS2613	1	4	10.6	.5	.2	.3	.4	.9	.9	.0	.0	1	1	1303	0

Other

- **Number of spool files (architected):**
 - 9999 per user
 - 1.6 million spool files per system
 - 1024 files per warm start block * (180 * 9) warm start blocks
- **Number of logged-on virtual machines (approximate): about 100,000 (per designers)**

Metrics for Formal Spin Locks

FCX265 CPU 2094 SER 19B9E Interval 02:31:51 - 12:34:01 GDLVM7

```

<----- Spin Lock Activity ----->
<----- Total -----> <--- Exclusive ---> <----- Shared ----->
Interval                Locks Average   Pct   Locks Average   Pct   Locks Average   Pct
End Time LockName      /sec   usec   Spin  /sec   usec   Spin  /sec   usec   Spin
>>Mean>> SRMATDLK      1.9    .539   .000   1.9    .539   .000   .0    .000   .000
>>Mean>> RSAAVCLK       .0    2.015   .000   .0    2.015   .000   .0    .000   .000
>>Mean>> FSDVMLK        .0   24.97   .000   .0   24.97   .000   .0    .000   .000
>>Mean>> SRMALOCK        .0    .000   .000   .0    .000   .000   .0    .000   .000
>>Mean>> HCPTRQLK       4.1    .195   .000   4.1    .195   .000   .0    .000   .000
>>Mean>> SRMSLOCK      34.0   1.096   .001  32.7   1.037   .001   1.3   .001   .000
    
```

Changes in Limits with Single System Image Clusters

- **Clustering four z/VM systems allows horizontal scaling**
- **Balance that with whitespace that might be required for Live Guest Relocation (LGR)**
- **If MP or Scaling effects for one large z/VM system have negative impact, splitting into multiple smaller z/VM systems in an SSI Cluster could be beneficial.**

SSI Cluster Effect on Processors Limits

- **Real Processors:**

- 32 x 4 = 128 processors
- Consider white space
- Low processor requirements for cross member communication as long as system resource (device) access is stable
- Perhaps greater efficiency by running smaller n-way
 - Example: One 32-way vs. Four 8-ways
 - Gives 4 master processors from one perception.

- **Virtual Processors:**

- If splitting z/VM system into smaller systems, remember to ensure no virtual machine has more virtual CPUs than logicals on the system.

z/VM 6.2 Effect on Memory Limits

- **Real Memory:**
 - 256 GB x 4 = 1 TB
 - With z/VM 6.3 1TB x 4 = 4TB! ★
 - Consider white space, cannot share like processors
 - Low memory costs to duplicate z/VM kernel and most control structures.
- **Virtual Machine Memory:**
 - No change
- **Paging Space**
 - Some slots lost due to sharing across members
 - But can reuse paging slots on each member, so it scales well.

Other SSI Cluster Effects on Limits

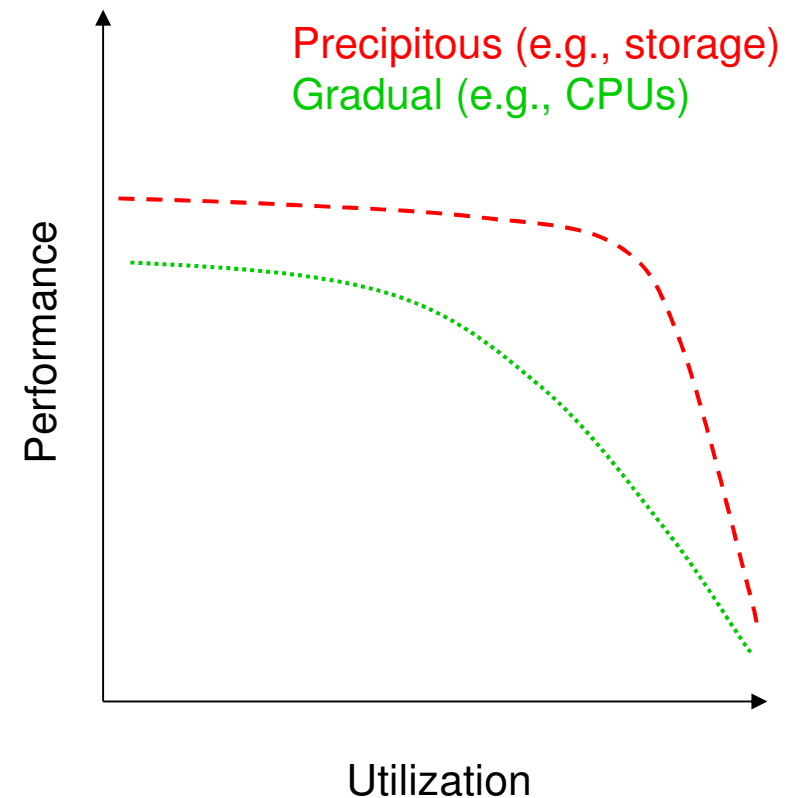
- **Distance for limit on DASD on SSI Cluster is 100km, unless using repeater technology.**
- **Distance for limit on Network on SSI Cluster is 10km, unless using repeater technology.**
 - Can double that if physical switches are placed at 10km from each CEC.
 - Remember, all members have to be in same LAN segment (or segments).

Latent Limits

- **Sometimes it's not an architected limit**
- **Sometimes it's just “your workload won't scale past here, because...”**
- **In our studies of z/VM 5.3, we found these kinds of latent limits:**
 - Searching for a below-2-GB frame in lists dominated by above-2-GB frames (storage balancing functions)
 - Contention for locks, usually the scheduler lock
- **These kinds of phenomena were the reasons we published the limits to be 256 GB and 32 engines**
 - We wanted to publish supported limits we felt would be safe in a very large variety of workloads and environments
 - Many of our measurement workloads scaled higher than this (for example, 440 GB and 54 engines)

Other Notes on z/VM Limits

- **Sheer hardware:**
 - z/VM 5.2: 24 engines, 128 GB real
 - z/VM 5.3: 32 engines, 256 GB real
 - zSeries: 65,000 I/O devices
- **Workloads we've run in test have included:**
 - 54 engines
 - 1 TB real storage
 - 128 GB XSTORE
 - 240 1-GB Linux guests
 - 8 1-TB guests
- **Utilizations we routinely see in customer environments**
 - 85% to 95% CPU utilization without worry
 - Tens of thousands of pages per second without worry
- **Our limits tend to have two distinct shapes**
 - Performance drops off slowly with utilization (CPUs)
 - Performance drops off rapidly when wall is hit (storage)



Keeping Tabs on Consumption Limits

- **Processor**
 - CPU utilization: FCX100 CPU, FCX114 USTAT
- **Memory & Paging**
 - Page slots in use: FCX146 AUXLOG
 - DASD I/O: FCX109 DEVICE CPOWNED
 - V:R Memory ratio: FCX113 UPAGE
 - PTRM space consumed: FCX134 DSPACESH
 - Storage in use for segment tables: FCX113 UPAGE
 - Consumption of SXS space: FCX261 SXS AVAIL
 - MDC: FCX178 MDCSTOR, FCX134 DSPACESH
 - Consumption of real memory: FCX103 STORAGE, FCX254 AVAILLOG
 - Consumption of expanded storage: FCX103 STORAGE
- **I/O**
 - DASD I/O: FCX108 DEVICE
 - Concurrency on FICON chpids: FCX131 DEVCONF, FCX215 INTERIM FCHANNEL, FCX168 DEVLOG

What Consumption Limits Will We Hit First?

- **Depends on workload**
 - Guest-storage-intensive:
 - page slots on DASD... at 5-6 TB things start to get interesting... mitigate by paging to SCSI
 - utilization on paging volumes and chpids: watch for MLOAD elongation... mitigate by spreading I/O
 - Page Reorder Processing
 - mitigation by application tuning... perhaps smaller guests
 - Real-storage-intensive:
 - Ability of the system to page will limit you: ensure adequate XSTORE and paging capacity
 - You can define > 256 GB of real storage, but we are aware that some workloads cannot scale that high
 - Mitigation by application tuning or by using CMM
 - CPU-intensive:
 - FCX100 CPU and FCX 114 USTAT will reveal CPU limitations
 - You can define > 32 engines, but we are aware that some workloads cannot scale that high
 - Mitigation by application tuning
 - I/O-intensive:
 - Device queueing: consider whether PAV or HyperPAV might offer leverage
 - Chpid utilization: add more chpids per storage controller
 - Ultimately partitions can be split, but we would prefer you not have to do this (too complicated)
- **Without trend data (repeated samples) for *your* workloads it is difficult to predict which of these limits *you* will hit first**

Summary

- **Knowing Limits:**
 - Real resource consumption
 - Limits to managing the virtualization of real resources
- **Measuring Limits:**
 - Knowing where to watch for these limits
 - Including these in capacity planning
- **Managing Limits**
 - Tuning and configuring
 - Planning for growth

Contact Information

Bill Bitner

z/VM Customer Focus and Care

bitnerb@us.ibm.com

+1 607-429-3286

Please remember to do an online evaluation at
www.share.org/bostoneval

