z/VM System Limits

version 8
see https://www.vm.ibm.com/library/presentations/ for latest version

Jacob Gagnon
Software Engineer
z/VM Development Lab
Endicott, NY

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

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 which limit-related performance metrics to review

- Discuss limits that may be hit first

Acknowledgments:
- Tim Greer, Xenia Tkatschow, Bill Bitner, Jacob Gagnon, Emily Hugenbruch

```
                              / VV           VVV MM          MM
                             /  VV          VVV  MMM        MMM
    ZZZZZZ        /       VV          VVV   MMMM      MMMM
       ZZ       /       VV      VVV      MM MM MM MM
      ZZ      /       VV   VVV       MM  MMM  MM
     ZZ     /       VVVVV        MM  M    MM
    ZZ    /       VVV          MM          MM
   ZZZZZZ /         V          MM        MM

      built on IBM Virtualization Technology
```

Limits
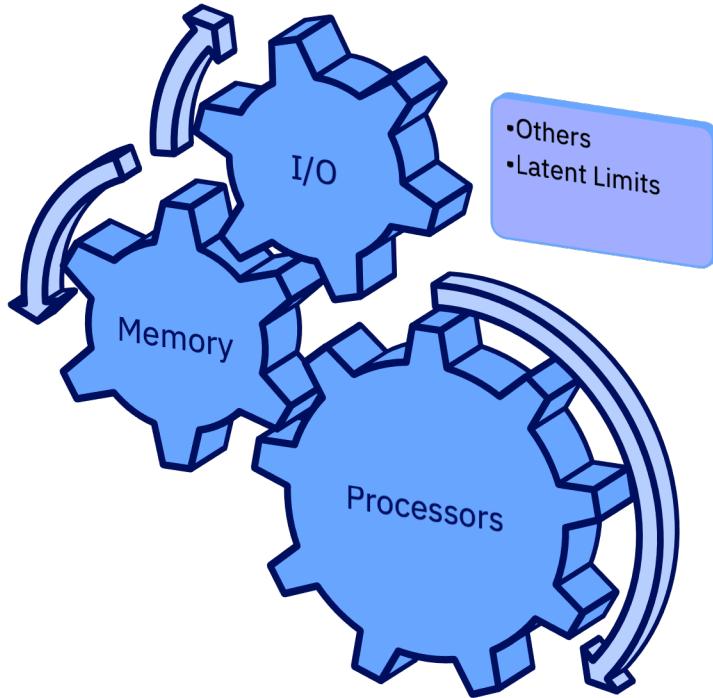
ADDITIONAL DISCLAIMERS:

– 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 sometimes 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.

Key Notes for Presentation

| 6.4 | 6.4+ | 7.1 | 7.1+ | 7.2 | 7.2+ |

- z/VM Continuous Delivery Strategy

- Presentation will show limits affected based on:

6.4	z/VM 6.4 GA November 11, 2016
6.4+	z/VM 6.4 plus service
7.1	z/VM 7.1 became GA September 21, 2018
7.1+	z/VM 7.1 plus service
7.2	z/VM 7.1 became GA September 18, 2020
7.2+	z/VM 7.1 plus service

- z/VM 6.3 went End of Service December 31, 2017 and is not called out in this presentation.
- z/VM 6.4 goes End of Service March 31, 2021
- IBM Z references apply to equivalent LinuxONE machines except if noted separately

Key Notes for Presentation

- Throughout this presentation, limits highlighted in:

RED are PRACTICAL Limits

YELLOW are SUPPORTED limits

GREEN are ARCHITECTED limits

Processors (Part 1 of 2)

| 6.4 | 6.4+ | 7.1 | 7.1+ |

- Processors hardware architected:
  - Includes all processor types (CP, zIIP, IFL, etc)

- Processors hardware available to customer:
  - z15:   **190** (model M05 only)
  - z14:   **170** (model M05 only)
  - z13:   **141**
  - zEC12:   **101**
  - z196:   **80**

- PR/SM Logical processors:

- Logical processors in a z/VM partition supported:
  - is **80** on z14 and newer with z/VM 7.1 + VM66265
  - is **64** on z13 and newer with z/VM 7.1 or z/VM 6.4 + VM65586
  - is **32** on zEC12 and older with z/VM 6.4

- **Note:** with SMT-1 or SMT-2, the limit to number of cores supported is half the logical processors as each possible logical processor would be associated with a thread on an IFL core. So logical 80-way would be a limit of 40 IFL cores even with SMT-1.

Processors (Part 2 of 2)

6.4

- z/VM master processor (z/VM design): 1
  - Some z/VM work is serialized by running on a "master" processor
  - Watch for 100%-utilized, rare in Linux workloads
  - z/VM will elect a new master if master fails or is varied off
  - Master may be reassigned to keep it as a vertical high processor when running in vertical polarization mode

- Virtual CPUs in a single virtual machine (z/VM design): 64
  - But $N_{Virtual} > N_{Logical}$ is usually not practical
  - Most interrupts presented to just 1 virtual CPU

- Number of logical partitions
  - z196   **60**
  - zEC12   **60**
  - z13   **85**
  - z14   **85**
  - z15   **85**

Topology and Vertical CPU Management

How much REAL processor is my LOGICAL processor guaranteed?

VH – Vertical High CPUs are entitled to 100% of a real CPU
VM – Vertical Medium CPUs are entitled to some of a real CPU (50%-100%)
VL – Vertical Low CPUs are not entitled to any real CPU

z/VM tries to map LOGICAL processors to REAL processors as closely as possible and move those mappings as little as possible.

Processor Scaling

Number of Supported Logical Processors in z/VM
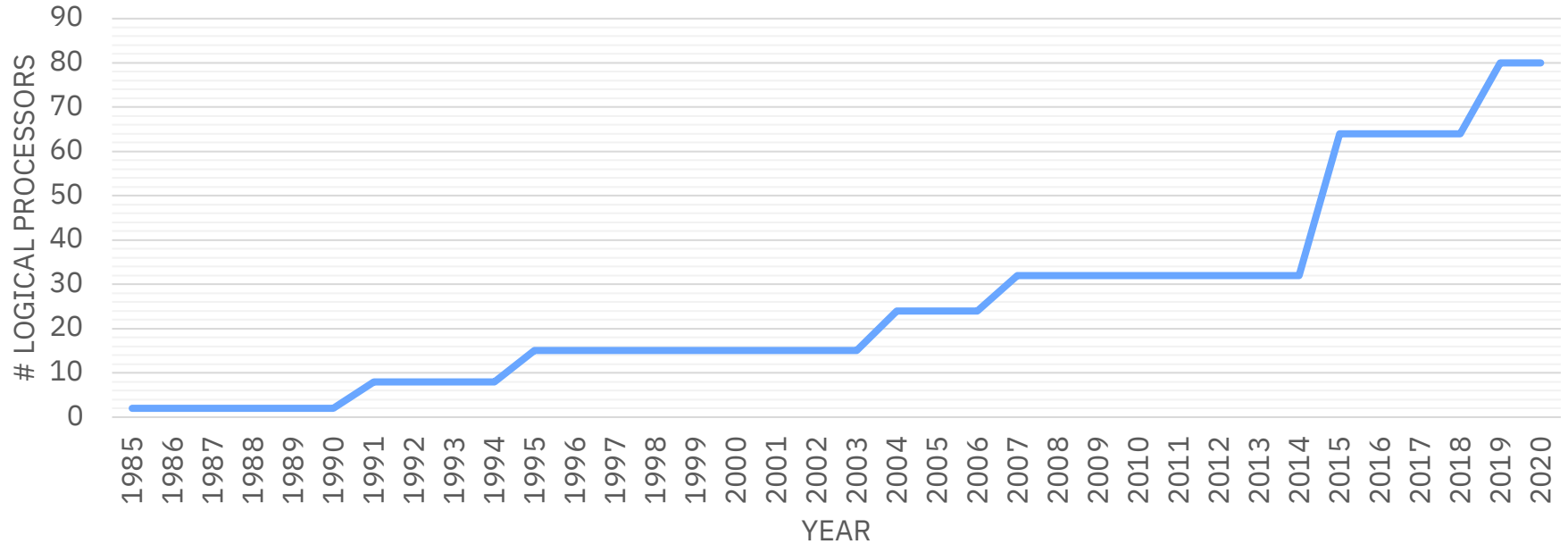
Processors: FCX100 CPU

```
1FCX100   Run 2019/06/13 09:30:42        CPU                                              Page   15
                                         General CPU Load and User Transactions

 From 2019/06/13 03:51:22                                                      A10Z6040
 To   2019/06/13 04:01:52                                                      CPU 3906-M05    SN 146E7
 For    630 Secs 00:10:30                  Result of A10Z6040 Run              z/VM    V.7.1.0 SLU 0000
 _____

 CPU Load                                                        Status or
 PROC TYPE %CPU   %CP %EMU WT  %SYS  %SP %SIC %LOGLD  %PR %ENT   ded. User
 P00  IFL    42    14   28 58     3    0   97     42    0  100   Alternate
 P01  IFL    42    14   28 58     3    0   97     42    0  100   Alternate
 P02  IFL    42    14   28 58     3    0   97     42    0  100   Alternate
 P03  IFL    42    14   28 58     3    0   96     42    0  100   Alternate
 P04  IFL    42    14   28 58     3    0   96     42    0  100   Alternate
 P05  IFL    42    14   28 58     3    0   96     42    0  100   Alternate
 P06  IFL    41    14   27 59     3    0   96     41    0  100   Alternate
 P07  IFL    42    14   27 58     3    0   96     42    0  100   Alternate
 P08  IFL    36    14   22 64     3    0   94     36    0  100   Master
```

1. T/V ~ **42**/**28** = 1.5   a good chunk of CP overhead here

2. Master does not seem unduly burdened

Processors: FCX304 PRCLOG

```
1FCX304  Run 2019/06/13 09:30:42          PRCLOG                                      Page    5
                                          Processor Activity, by Time
 From 2019/06/13 03:51:22                                                    A10Z6040
 To   2019/06/13 04:01:52                                                    CPU 3906-M05    SN 146E7
 For   630 Secs 00:10:30                   Result of A10Z6040 Run           z/VM    V.7.1.0 SLU 0000
 _____

                                      <--- Percent Busy ---->  <-- Rates per Sec. --->  <----- Paging ------->  <Co>  < Di>
                              Pct                                                                         Fast   Page <mm>  < ag>
 Interval                     Park                              Inst                      <2GB  PGIN Path  Read  Msgs X'9C'  Core/
 End Time  CPU Type PPD Ent. DVID Time %Susp Total  User   Syst  Emul  Siml  DIAG  SIGP  SSCH   /s    /s   %    /s    /s    /s  Thrd
 >>Mean>>  00 IFL  VhD 100 0000    0   .0  42.4  39.4   3.0  27.9 28954 1749 16254 445.7    .0   .0 ....  4186   .0    .0  00/0
 >>Mean>>  01 IFL  VhD 100 0001    0   .0  42.2  39.4   2.8  27.9 29114 1775 15636 466.4    .0   .0 ....  4031   .0    .0  01/0
 >>Mean>>  02 IFL  VhD 100 0002    0   .0  41.7  38.9   2.8  27.6 28751 1763 15329 491.7    .0   .0 ....  3893   .0    .0  02/0
 >>Mean>>  03 IFL  VhD 100 0003    0   .0  41.9  39.1   2.7  27.7 28826 1736 15070 492.2    .0   .0 ....  3974   .0    .0  03/0
 >>Mean>>  04 IFL  VhD 100 0004    0   .0  41.8  39.0   2.8  27.7 28805 1748 15120 496.7    .0   .0 ....  3971   .0    .0  04/0
 >>Mean>>  05 IFL  VhD 100 0005    0   .0  41.7  38.8   2.9  27.5 28540 1727 14901 535.1    .0   .0 ....  3930   .0    .0  05/0
 >>Mean>>  06 IFL  VhD 100 0006    0   .0  41.4  38.5   3.0  27.3 28326 1722 14889 542.0    .0   .0 ....  3820   .0    .0  06/0


 03:51:52  00 IFL  VhD 100 0000    0   .0  23.9  22.2   1.7  16.3 14930 727.6 8259 225.0    .0   .0 ....  4708   .1    .0  00/0
 03:51:52  01 IFL  VhD 100 0001    0   .0  23.5  21.9   1.6  16.0 15068 741.5 8051 239.8    .0   .0 ....  4390   .5    .0  01/0
 03:51:52  02 IFL  VhD 100 0002    0   .0  22.7  21.1   1.6  15.5 14345 754.6 7977 241.1    .0   .0 ....  4116   .0    .0  02/0
 03:51:52  03 IFL  VhD 100 0003    0   .0  23.3  21.8   1.5  16.0 14793 734.9 7730 234.0    .0   .0 ....  4361   .0    .0  03/0
 03:51:52  04 IFL  VhD 100 0004    0   .0  23.4  21.7   1.7  16.0 14671 755.0 7794 240.0    .0   .0 ....  4273   .1    .0  04/0
 03:51:52  05 IFL  VhD 100 0005    0   .0  23.3  21.7   1.6  16.0 14741 730.8 7605 262.0    .0   .0 ....  4200   .0    .0  05/0
 03:51:52  06 IFL  VhD 100 0006    0   .0  22.9  21.2   1.6  15.7 14196 745.5 7698 274.1    .0   .0 ....  4023   .0    .0  06/0
```

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                                                       CPU 2094-700
For   3180 Secs 00:53:00                                                       z/VM   V.5.3.0 SLU 0701
_____


  .      ____      .    .    .    .    .    .    .    .    .    .    .    .    .    .   .   .   .  .   .
                                                  <-SVM and->                 <--%Time spent in--> Nr of
 Userid   %ACT   %RUN %CPU %LDG %PGW %IOW %SIM %TIW %CFW %TI %EL %DM %IOA %PGA %LIM %OTH  Q0  Q1  Q2  Q3 E0-3 Users
 >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
 RSCSDNS1 100     0    0    0    0    0    0    0    0    0   0  100    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
 SZVAS001 100     2    0    0    0    0    0   97    0    0   0    0    0    0    0    1    0   3  12  85    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

3. %LIM is limit list and Resource Pool related

Processors: FCX302 PHYSLOG Report

```
1FCX302   Run 2019/06/13 09:30:42          PHYSLOG
                                           Real Core Utilization Log

 From 2019/06/13 03:51:22
 To   2019/06/13 04:01:52
 For    630 Secs 00:10:30                  Result of A10Z6040 Run

 _____

 Interval      <PhCore> Shrd  Total
 End Time  Type Conf Ded Log.  Weight %LgclC %Ovrhd LCoT/L %LPmgt %Total TypeT/L
 >>Mean>> CP      4    0   0       0   .000   .000   ...    .000   .000    ...
 >>Mean>> IFL   128   16   0       0 1599.8   .038  1.000   .040 1599.8  1.000
 >>Mean>> ICF     4    0   0       0   .000   .000   ...    .000   .000    ...
 >>Mean>> ZIIP    4    0   0       0   .000   .000   ...    .000   .000    ...
 >>Mean>> >Sum  140   16   0       0 1599.8   .038  1.000   .040 1599.8  1.000

 03:51:52 CP      4    0   0       0   .000   .000   ...    .000   .000    ...
 03:51:52 IFL   128   16   0       0 1596.2   .035  1.000   .031 1596.3  1.000
 03:51:52 ICF     4    0   0       0   .000   .000   ...    .000   .000    ...
 03:51:52 ZIIP    4    0   0       0   .000   .000   ...    .000   .000    ...
 03:51:52 >Sum  140   16   0       0 1596.2   .035  1.000   .031 1596.3  1.000
```

Memory (Part 1 of 6)

Real Memory (Central Storage)

- CPC Total maximum customer memory:
  - z15   **40 TB**
  - z14   **32 TB**
  - z13   **10 TB**
  - zEC12   **3 TB**
  - z10 EC   **1.5 TB**

- Maximum LPAR size:
  - z15   **16 TB**
  - z14   **16 TB**
  - z13   **10 TB**
  - zEC12   **1 TB**
  - z196   **1 TB**
  - z10 EC   **1 TB**

- z/VM supported limit: **2 TB**
                        4 TB with z/VM 7.2 + VM66173

Memory (Part 2 of 6)

6.4

z/VM virtual machine size supported: **1 TB**

- Practical limit can be gated by performance of:
    - **Dumping** a VM system
    - **Live Guest Relocation** requirements
    - **Production level performance** requirements

Active, or instantiated, total virtual machines limit imposed by DAT structure limits:
- **64 TB**
    - 128 PTRM pre-allocated spaces each 2 GB-space can map 512 GB of guest-real memory (host-virtual).

Memory (Part 3 of 6)

6.4

7.2+

- Virtual to real memory ratio (z/VM design): **64 TB : 2 TB = 32:1**
    64 TB : 4 TB = 16:1 (with z/VM 7.2 + VM66173)

- Virtual to real memory 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 real memory configured to z/VM.

    - **1:1** if you want to eliminate performance impact for production workloads.

    - Consider maximum ratio due to:
        - Workload growth
        - Live Guest Relocation

    - Practical over commitment dependent on:
        - Active:Idle virtual machines
        - Workload/Service Level Agreement sensitivity to delays
        - Performance of paging subsystem (e.g. flash, HyperPAV, channels, etc.)
        - Accuracy of sizing of the virtual machines
        - Exploitation of memory saving/exploitation capabilities (e.g. CMM, DIM)

Memory (Part 4 of 6)

6.4   7.1+

- z/VM design CP Owned volumes: **255**
  - Only a subset can be used for paging
  - SSI configurations paging is not shared, but other CP-owned slots are.

Release	ECKD (3390)	EDEV (SCSI)
z/VM 7.1 + APAR VM66263	202 TB	15.9 TB
z/VM 6.4	11.2 TB	15.9 TB

- Maximum paging space design limits (if you could use all volumes)

- Concurrent paging I/Os per paging volume:
  - ECKD without HyperPAV: **1**
  - ECKD with HyperPAV: **8**
  - EDEV: >1 (Have observed average of 1.6 in heavy workloads)

Memory (Part 5 of 6)

6.4   7.1+

- Rules of thumb:

  - Do not cheat on calculating paging space required!

  - Do not allow page space to become full (avoid PGT004 abends)

- Do not mix ECKD and EDEV paging volumes on same system

- Keep volumes dedicated to paging

- In environments with virtual to real ratio of 1, consider turning off early writes and keep slot

  - CP command: **SET AGELIST EARLYWRITES NO KEEPSLOT NO**

  - In system config file: **STORAGE AGELIST EARLYWRITES NO KEEPSLOT NO**

Memory (Part 6 of 6)

6.4   7.1

- System Execution Space (SXS) z/VM design limit: **2 GB**
  - For practical purposes it is 2GB, but there are structures in the space placed above 2GB

- DCSS
  - Individual Segments up to 2047 MB
    - Segments must end prior to one 4KB page below 512GB

- Minidisk Cache (z/VM design): **8 GB**
  - Recommended limit **2 GB**
  - Recommend fixing MDC size rather than letting arbiter change it dynamically

- Installing z/VM: minimum of **768 MB**

- Minimum memory to run z/VM second level:
  - z/VM 6.4:   **32 MB**
  - z/VM 7.1:   **128 MB**

Memory References

Memory over commitment
– http://www.vm.ibm.com/perf/tips/memory.html

Paging in general
– http://www.vm.ibm.com/perf/tips/prgpage.html

Real Memory Scaling

Real Memory Supported by a z/VM system
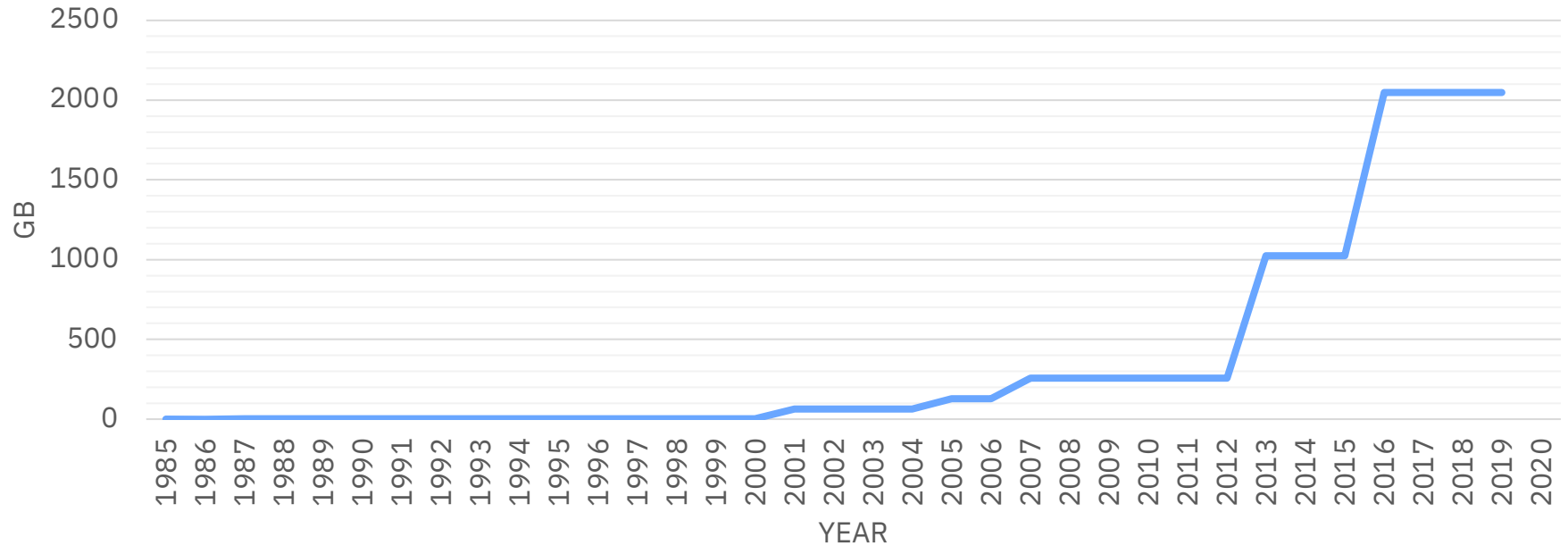
VIR2REAL Tool

- Displays the ratio of total virtual storage to LPAR real storage of your z/VM system.

  - Too high a ratio, and your system my underperform

  - Too low a ratio, you may be able to handle more workload.

- Displays your defined paging space (Indicates if paging or dump space is inadequate)

- Great as a quick check tool.

- Does not indicate HOW your system is paging.

NOTE: VIR2REAL is an aid but it can't tell you everything!

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

_____


           <Page Slots> <Spool Slots> <Dump Slots> <----- Spool Files -----> <Average MLOAD>
Interval    Total Used    Total  Used  Total Used  <-Created--> <--Purged--> Paging Spooling
End Time    Slots   %     Slots    %   Slots   %   Total   /s  Total    /s    msec    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
```

DASD I/O: FCX109 DEVICE CPOWNED

```
1FCX109  Run 2019/06/13 09:30:42         DEVICE   CPOWNED                                        Page   34
                                         Load and Performance of CP Owned Disks

 From 2019/06/13 03:51:22                                                      A10Z6040
 To   2019/06/13 04:01:52                                                      CPU 3906-M05    SN 146E7
 For    630 Secs 00:10:30              Result of A10Z6040 Run                   z/VM   V.7.1.0 SLU 0000
 _____

 Page / SPOOL Allocation Summary
 PAGE slots available        2642m      SPOOL slots available      7810920
 PAGE slot utilization          3%      SPOOL slot utilization          0%
 T-Disk space avail. (MB) .......       DUMP slots available        23587k
 T-Disk space utilization    ...%       DUMP slot utilization          21%


 .     .        .                       .     .     .      .      .      .  ____    .    .     .    . .
 < Device Descr. ->                     <------------ Rate/s ------------>  I/O       Serv MLOAD Block %Used I
       Volume Area   Area       Used    <--Page---> <--Spool-->     SSCH Inter Queue  Time  Resp. Page  for O
 Addr Devtyp Serial Type        Extent    %  P-Rds P-Wrt S-Rds S-Wrt Total +RSCH feres Lngth /Page  Time Size Alloc M
 C005 3390-9 ATP033 PAGE      11793420   3  359.1 373.4  ...   ... 732.4  40.4    0 148.9   .9  24.8   18  100 C
 CD03 3390-9 ATP213 PAGE      11793420   3  363.3 374.9  ...   ... 738.2  39.3    0 137.5  1.0  16.6   18  100 C
 C600 3390-9 ATP112 PAGE      11793420   3  362.1 373.5  ...   ... 735.6  41.3    0 133.1   .8  76.8   18  100 C
 C70A 3390-9 ATP136 PAGE      11793420   3  362.2 375.9  ...   ... 738.0  40.6    0 126.6  1.0  33.2   18  100 C
 C20A 3390-9 ATP066 PAGE      11793420   3  361.1 371.2  ...   ... 732.3  38.9    0 125.5   .8  53.3   18  100 C
 CA0A 3390-9 ATP178 PAGE      11793420   3  364.4 372.1  ...   ... 736.5  40.7    0 123.8   .9  53.1   18  100 C
 C80D 3390-9 ATP153 PAGE      11793420   3  364.0 370.3  ...   ... 734.3  39.6    0 122.7   .9  27.7   18  100 C
 CB0A 3390-9 ATP192 PAGE      11793420   3  364.0 375.8  ...   ... 739.8  40.3    0 122.0  1.1  77.5   18  100 C
```

Report FCX292 UPGUTL

```
FCX292  Run 2013/04/10 07:38:36        UPGUTL                                                                 Page  103
                                       User Page Utilization Data
From 2013/04/09 16:02:10                                                                      SYSTEMID
To   2013/04/09 16:13:10                                                                      CPU 2817-744   SN A6D85
For    660 Secs 00:11:00               "This is a performance report for SYSTEM XYZ"          z/VM   V.6.3.0 SLU 0000


_____         .      .     .     .     .     .     .     .         .     .     .     .     .     .     .       .
                <----------------------------------------------- Storage ------------------------------------------------->
                        <----------------------- Resident ---------------------------->
          Data                                     <----- Invalid But Resident ------>                 Base
          Spaces                       <---- Total ----> <-Locked--> <-- UFO --> <-- PNR --> <-AgeList->          Space Nr of
Userid    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
FTPSERVE    .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 concepts: Inst   IBR   UFO   PNR   AgeList
- Amounts are in bytes, suffixed. Not page counts!

Report FCX290 UPGACT

```
FCX290  Run 2013/04/10 07:38:36        UPGACT                                                  Page
102
                                        User Page Activity
From 2013/04/09 16:02:10                                                        SYSTEMID
To   2013/04/09 16:13:10                                                        CPU 2817-744   SN
A6D85
For    660 Secs 00:11:00                "This is a performance report for SYSTEM XYZ"      z/VM   V.6.3.0 SLU
0000
```

```
             .       .      .      .      .      .      .      .      .      .      .
          <---------------------------- 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
DTCVSW1   1.0     0     0  3072  2855     0    0     0     0     0     0    0     0
DTCVSW2   1.0     0     0  3004  2780     0    0     0     0     0     0    0     0
EREP      1.0     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
GCSXA     1.0     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
LCC00002  1.0  657K 24955  487K  486K 54725    0     0     0 44522 18991    0     0
LCC00003  1.0  565K 23012  485K  481K 64065    0     0     0 44783 19859    0     0
LCC00004  1.0  602K 24104  499K  495K 63178    0     0     0 48811 24588    0     0
LCC00005  1.0  717K 25675  500K  499K 65865    0     0     0 66002 28753    0     0
```

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

Zoom in on Report FCX290 UPGACT

```
FCX290  Run 2013/04/10 07:38:36        UPGACT                                          Page 102
                                        User Page Activity
From 2013/04/09 16:02:10                                                    SYSTEMID
To  2013/04/09 16:13:10                                                     CPU 2817-744   SN A6D85
For    660 Secs 00:11:00                "This is a performance report for SYSTEM XYZ"    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   0   0 49866 23670   0   0
LCC00002  1.0  657K 24955  487K  486K 54725   0   0   0 44522 18991   0   0
LCC00003  1.0  565K 23012  485K  481K 64065   0   0   0 44783 19859   0   0
LCC00004  1.0  602K 24104  499K  495K 63178   0   0   0 48811 24588   0   0
LCC00005  1.0  717K 25675  500K  499K 65865   0   0   0 66002 28753   0   0
```

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

Report FCX295 AVLA2GLG

```
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 End Time	<----------- Storage ------------->						<--Times-->		<-Frame Thresh-->		
	<Available>		<Requests/s>		<Returns/s>		<-Empty/s->		Sing	<-Contigs->	
	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

MDC Spaces: FCX134 DSPACESH (newer one with new PTRM layout)

Owning		Users									
				<--Resid-->		<-Locked-->		<-Aliases->			
Userid	Data Space Name	Permt	Total	Resid	R<2GB	Lock	L<2GB	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

(Header: <----------------Number of Pages---------------->)

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

I/O Devices

- Number of subchannels in a partition (device numbers): **65,536**

- Number of devices per virtual machine: **24576 (24K)**

- GDPS environments can have secondary DASD devices defined in an alternate subchannel set with the Multiple Subchannel Set Support

- Concurrency
  - ECKD without PAV or HyperPAV: **1**
  - ECKD with PAV or HyperPAV: **8**

I/O Disk Sizes

Type	CMS	Minidisk	Dedicated	CP Use
ECKD 3390	~45GB 65,520 cylinders (practical 22 GB)[3]	~812 GB[5] 1,182,006 cylinders	~812 GB 1,182,006 cylinders	~812 GB[6] 1,182,006 cylinders ~45 GB non-paging
SCSI EDEV	381 GB (practical 22 GB[3])	1023 GB[2]	1023 GB[2]	64 GB[4]
SCSI Dedicated	n/a	n/a	?????	n/a

[1] – Sizes listed above are in powers of 2
[2] – Exact value is 1024 GB minus 4 KB
[3] – Due to file system structure under 16MB, unless there are very few files
[4] – CP can use, but PAGE, SPOL, DRCT must be below 64 GB on the volume
[5] – Requires z/VM 6.4 VM65943 or z/VM 7.1, otherwise limit is ~45GB, 65,520 cylinders
[6] – Requires z/VM 7.1 VM66263, otherwise limit is as for other device types ~45 GB at 65,520 cylinders

I/O – Other Limits

- Virtual Disk in Storage (VDISK) size z/VM design: **2 GB** (minus eight 512-byte blocks)

- Total VDISK z/VM design: **1 TB**
  - "Infinite" = 2,147,483,648 512-byte blocks

Single Virtual Switch 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                                              CPU 2094-700   SN
For   3181 Secs 00:53:01                                              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    0   ...    .0
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 12.6  ...   .3   .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 Limits – Spool and CMS Files

- Number of spool files (z/VM design):
    - Limit **9999** per virtual machine (2499 in SSI)
    - Limit **1.6 million** spool files per system

- 1024 files per warm start block * (180 blocks * 9 cylinders)

- Number of logged-on virtual machines (design point): **about 100,000**

- CMS Files
    - Maximum Records: 2,147,483,647 ($2^{31}-1$) records, each of which consist of from one to $2^{31}-1$ bytes of data (a record in a file with variable-length records is further restricted to 65,535 bytes of data).

Other Limits

- 255 CP-owned slots

- 16 ISFC links between a pair of systems
  - No limit on total number of ISFC links

- 1 GB Distributed IUCV maximum message size

- 8 Alternate Operators

- Password length: 8 characters, 100 characters with RACF

- 1000 System Environment Variables
  - Up to 63 character named
  - Up to 255 character values

- HyperPAV aliases:
  - 254 per pool
  - 160,000 pools per system

No Hard Limits, but Potential Soft Limits

- **Virtual Switch**
    - Users into hundreds, broadcast group limited to 1000
    - Better performance when users spread over multiple virtual switches

- ISFC
    - Network topology important if network is large
    - Propagation effects in large, sparse network
    - Internal structure stresses in large, dense network

- Guest levels
    - **7th level** z/VM system is impractically slow
    - Diagnose x'00' returns up to **5 levels** of information

LOCKACT report example

```
1FCX326  Run 2019/06/13 09:30:42        LOCKACT                                      Page   58
                                        Spin Lock Activity

 From 2019/06/13 03:51:22                                              A10Z6040
 To   2019/06/13 04:01:52                                              CPU 3906-M05   SN 146E7
 For    630 Secs 00:10:30               Result of A10Z6040 Run         z/VM   V.7.1.0 SLU 0000
 _____

          .     .    _____      .      .     .     .      .      .     .     .      .     .
          <------- Combined --------->  <------- Exclusive -------->  <-------- Shared ---------->
 LockName  CCol/s  CAvSpn C%Busy CCAD/s  ECol/s  EAvSpn E%Busy ECAD/s  SCol/s  SAvSpn S%Busy SCAD/s
 >>Total>  3805.5   .923   .351   .000  2151.7   .602   .129   .000  1653.8  1.340   .222   .000
 SRMSLOCK  1952.1  1.211   .236   .000   298.29   .498   .015   .000  1653.8  1.340   .222   .000
 HCPPGDAL  105.08  4.315   .045   .000   105.08  4.315   .045   .000    .000   ...    .000   .000
 FSDVMLK   151.01  1.198   .018   .000   151.01  1.198   .018   .000    .000   ...    .000   .000
 HCPPGDPL  27.429  1.570   .004   .000   27.429  1.570   .004   .000    .000   ...    .000   .000
 DSV_0000  125.64   .332   .004   .000   125.64   .332   .004   .000    .000   ...    .000   .000
 DSV_0005  119.19   .311   .004   .000   119.19   .311   .004   .000    .000   ...    .000   .000
 DSV_0002  116.97   .301   .004   .000   116.97   .301   .004   .000    .000   ...    .000   .000
 DSV_0007  112.18   .314   .004   .000   112.18   .314   .004   .000    .000   ...    .000   .000
 DSV_0004  111.59   .310   .003   .000   111.59   .310   .003   .000    .000   ...    .000   .000
 DSV_0003  113.27   .304   .003   .000   113.27   .304   .003   .000    .000   ...    .000   .000
 DSV_0006  108.51   .316   .003   .000   108.51   .316   .003   .000    .000   ...    .000   .000
 DSV_0001  110.84   .301   .003   .000   110.84   .301   .003   .000    .000   ...    .000   .000
 SRMATDLK  115.12   .271   .003   .000   115.12   .271   .003   .000    .000   ...    .000   .000
 HCPTRQLK  31.689   .551   .002   .000   31.689   .551   .002   .000    .000   ...    .000   .000
```

Changes in Limits with Single System Image Clusters

- Horizontal scaling through four z/VM members (systems) in a cluster.

- Balance that with whitespace that might be required for Live Guest Relocation (LGR)

- If n-way or scaling effects for one very 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 Processor Limits

- Real processors:
  - 80 x 4 = **320 processors** (if all members running z14 and newer with z/VM 7.1 + VM66265)
  - Consider white space
  - Low processor requirements for cross-member communication as long as system resource (device) access is stable
  - Greater efficiency in cases with smaller n-way
  - In a sense, gives 4 master processors

- Virtual processors:
  - If splitting z/VM system into smaller systems, remember to ensure no virtual machine has more virtual CPUs than the z/VM member (logical partition) has logical processors.

SSI Effect on Memory Limits

7.2+

- Real Memory:

    - 4 TB x 4 = ~~16 TB~~ (If all members running z/VM 7.2 + VM66173)

    - Consider white space, cannot share like processors

    - Low memory costs to duplicate z/VM kernel and most control structures

- Virtual Memory:

    - No change for individual virtual machine

    - 64TB x 4 = ~~256 TB~~ (aggregate)

- Paging Space:

    - Some CP-owned 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 limit on DASD and FICON CTC in the Cluster is 100km with repeater technology

- Distance limit on Network on SSI cluster from OSA to switches is 10km with repeater technology.

- For virtual machines using Virtual Switch and being relocated, those virtual switches need to be in the 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..."
    - Contention for certain locks
    - Search algorithms with scaling issues

- Because of the above, we often publish supported limits that are less than the designed or architected limits.

Other Notes on z/VM Limits

Limits we've tested, tend to have two distinct shapes

– Performance drops off slowly

– Performance drops off rapidly when a wall is hit.
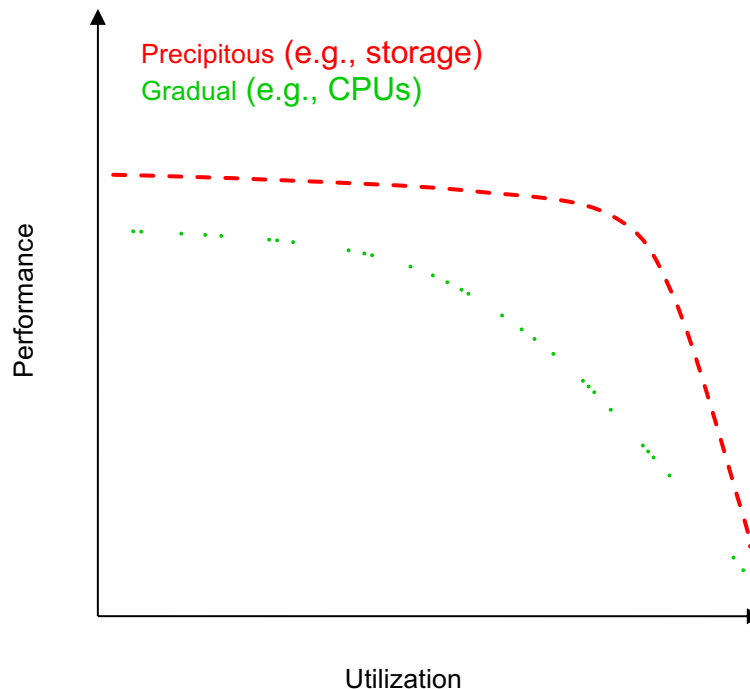
What Consumption Limits Will We Hit First?

`7.2+`

- Depends on workload
  - Memory-intensive:
    - 1:1 overcommit gated by real storage limit (4TB on z/VM 7.2 with VM66173)
    - Larger overcommit ratios gated by your paging subsystem
    - Mitigation by application tuning or by using CMM
  - CPU-intensive:
    - FCX100 CPU and FCX 114 USTAT will reveal CPU limitations
    - 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

APPENDIX

Older Limits from non-supported releases / hardware

Maximum LPAR size z9: 512 GB minus your HAS

Maximum LPAR size z10: 1 TB

Total maximum memory z900: 256 GB

Total maximum memory z990: 256 GB

Total maximum memory z9: 1 TB

Total maximum memory z10: 8 TB

Expanded storage architected limit: 16 TB

Expanded storage z/VM limit supported: 128 GB

Expanded storage z/VM design unsupported: ~660 GB dependent on other factors

z/VM 6.3 and older RoT: Keep paging space under 50% allocated for best performance.

V:R Ratio: FCX113 UPAGE (move to appendix)

```
          <--------- Paging Activity/s ---------> <---------- Number of Pages ---------------->
          <Page Rate>   Page  <-Page Migration-->      <-Resident-> <--Locked-->         Stor
  Nr of
Userid  Reads Write Steals >2GB> X>MS MS>X X>DS    WSS  R<2GB  R>2GB L<2GB  L>2GB XSTOR   DASD  Size
  Users
>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

- V:R = (1310 MB * 212) / 91 GB = 2.98

- For z/VM 6.2 and older

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                                                             CPU 2094-700
For  3180 Secs 00:53:00                                                              z/VM   V.5.3.0 SLU 0701
_____

         <--------------------------------------- Available List Management --------------------------------------->
         <---- Thresholds ---->  <--------- Page Frames --------->  <-Times->  <----------- Replenishment -----------> Perct
Interval <---Low--->  <--High--->  <Available> <Obtains/s> <Returns/s>  <-Empty->  <--Scan1--> <--Scan2--> <-Em-Scan-> Scan Emerg
End Time  <2GB  >2GB  <2GB  >2GB   <2GB  >2GB  <2GB  >2GB  <2GB  >2GB   <2GB >2GB  Compl Pages Compl Pages Compl Pages Fail  Scan
>>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.