B56

Introduction to z/VM Performance

Brian K. Wade, Ph.D.
bkw@us.ibm.com

IBM System z Expo
September 17-21, 2007
San Antonio, TX

Legal Stuff

Disclaimer

The information contained in this document has not been submitted to any formal IBM test and is distributed on an "as is" basis without any warranty either express or implied. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environment do so at their own risk.

In this document, any references made to an IBM licensed program are not intended to state or imply that only IBM's licensed program may be used; any functionally equivalent program may be used instead.

Any performance data contained in this document was determined in a controlled environment and, therefore, the results which may be obtained in other operating environments may vary significantly.

Users of this document should verify the applicable data for their specific environments.

It is possible that this material may contain references to, or information about, IBM products (machines and programs), programming, or services that are not announced in your country or not yet announced by IBM. Such references or information should not be construed to mean that IBM intends to announce such IBM products, programming, or services.

Should the speaker start getting too silly, IBM will deny any knowledge of his association with the corporation.

Trademarks

The following are trademarks of the IBM Corporation:
IBM, VM/ESA, z/VM
LINUX is a registered trademark of Linus Torvalds

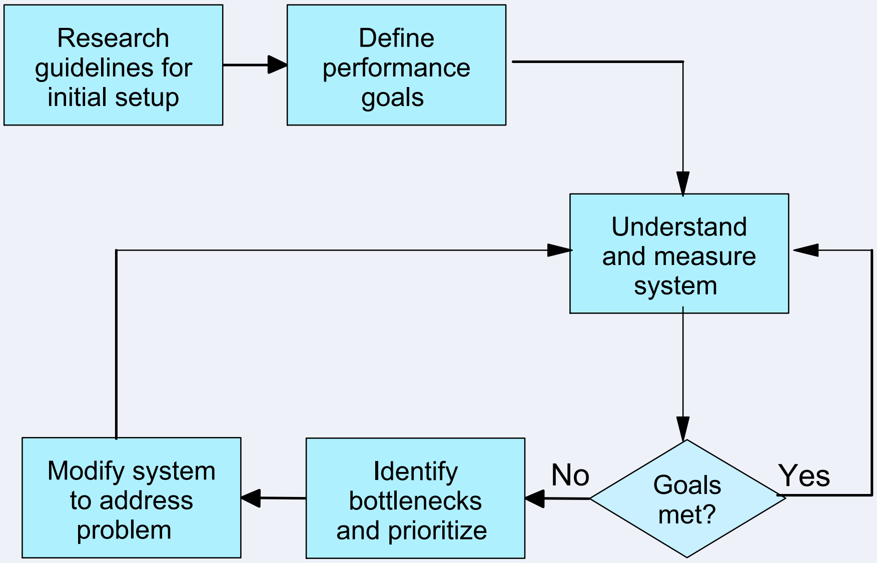
Credits

Thanks to Bill Bitner for letting me present his material.

Overview

- Performance process
- Performance definition
- Guidelines
- Native CP commands
- Other performance tools
- I/O performance concepts
- Case study
- Final thoughts

Performance Process


Research guidelines for initial setup
Define performance goals
Understand and measure system
Goals met?
Yes
No
Identify bottlenecks and prioritize
Modify system to address problem

Definition of Performance

Performance definitions:

- ❑ Response time
- ❑ Batch elapsed time
- ❑ Throughput
- ❑ Resource consumed per unit of work done
- ❑ Utilization
- ❑ Users supported
- ❑ Phone ringing
- ❑ Consistency
- ❑ All of the above

Performance Guidelines

- Processor
- Storage
- Paging
- Minidisk cache
- Server machines

Processor Guidelines

- Dedicated processors - mostly political
  - Absolute share can be almost as effective
  - Gets wait state assist and 500 ms minor time slice
  - Perhaps not a good idea if you are CPU-constrained
  - A virtual machine should have all dedicated or all shared processors
- Share settings
  - Use absolute if you can judge percent of resources required
  - Use relative if difficult to judge and if lower share as system load increases is acceptable
  - Do not use LIMITHARD settings unnecessarily
    - Masks looping users
    - More scheduler overhead
- Use the right number of virtual processors for the guest’s workload
  - Using too many dilutes share and induces unnecessary Diag x’44’ overhead
- Small minor time slice keeps CP reactive.
  - Long minor time slice blocks master-only work
  - Tinkering with this is an experts-only task anyhow

Storage Guidelines

- Virtual-to-real ratio should be <= 3:1 or make sure paging system is robust
- Use SET RESERVE instead of LOCK
  - WSS + 10%, but watch it for runaway
- Give your partition some expanded storage
  - Mitigates cost of a wrong page-out choice
  - Especially important for loads that stress <2 GB storage (z/VM 5.1 or earlier)
  - Rule of thumb: 25% of partition size, up to 2 GB (e.g., 6G/2G)
  - http://www.vm.ibm.com/perf/tips/storconf.html has more guidance
- Exploit shared memory where possible.
  - IPL your Linux guests from a segment
  - Use the Linux XIP (execute-in-place) file system
  - Put commonly-used CMS applications into segments
  - Take advantage of SFS files-in-dataspace technology
- Size guests "just right"
  - Excessively-sized Linux guests consume storage unnecessarily
  - Trim Linux guests until they just barely start to swap

Paging Guidelines

- Configure your partition with some XSTORE (paging hierarchy)
  - 25% of partition's storage, up to 2 GB (usually)
- DASD paging allocations less than or equal to 50%.
  - QUERY ALLOC PAGE
- Watch blocks read per paging request (keep >10)
  - Long block runs make paging I/O efficient
  - This comes out in the monitor data (FCX103)
- Multiple volumes and multiple paths
  - Remember, one I/O per real device at a time
- Do not mix PAGE extents with other extents on same volume
  - Change the system (z/VM 4.4.0 and earlier) ... change spool too
- If you have the CPU to spare, consider paging to SCSI
  - z/VM 5.3 can turn 66% more pages/sec to SCSI than to ECKD...
  - ... but each page costs 2.4 times as much CPU

Minidisk Cache Guidelines

- Configure some real storage for MDC.
  - It will use some anyway unless all reads are block-aligned
  - Stop thrashing and take advantage of those intermediate buffers
- In general, enable MDC for everything.
  - It will equilibrate based on page lifetime
- Disable MDC for:
  - Minidisks mapped to VM data spaces
  - Write-mostly or read-once disks (logs, accounting, Linux swap)
  - Target volumes in backup scenarios
  - In large storage environments, may need to bias against MDC.
- Prior to z/VM 5.2, consider disabling XSTORE MDC if constrained below 2 GB
- Better performer than Virtual Disk in Storage (VDISK) for read I/Os
  - Pathlength statement

Server Machine Guidelines

- TCP/IP, RACFVM, SFS, DB/2, Linux router
- QUICKDSP ON to avoid eligible list
  - Overcommits storage, though... be prepared (how?)
- Higher SHARE setting... ABSOLUTE, perhaps
- SET RESERVED to avoid paging
- NOMDCFS option in CP directory... it’s a server
- Routinely collect performance information about these servers
  - MONWRITE and PERFSVM deserve special treatment too

CP INDICATE Command

- LOAD: shows total system load.
  - Processors, XSTORE, paging, MDC, queue lengths
  - STORAGE value not very meaningful and was removed in z/VM 5.2
- USER EXP: more useful than plain USER
  - Shows all address spaces
  - Fields don’t overflow
- QUEUES EXP: great for scheduler problems and quick state sampling
  - Mostly useful for eligible list assessments
- PAGING: lists users in page wait.
- I/O: lists users in I/O wait.
- ACTIVE: displays number of active users over given interval
- Consider using monitor data instead for "serious" examinations

CP INDICATE LOAD Example

```
INDICATE LOAD
AVGPROC-088% 03
XSTORE-000000/SEC MIGRATE-0000/SEC
MDC READS-000035/SEC WRITES-000001/SEC HIT RATIO-099%
~~STORAGE-017%~~ PAGING-0023/SEC STEAL-000%
Q0-00007(00000)                            DORMANT-00410
Q1-00000(00000)           E1-00000(00000)
Q2-00001(00000) EXPAN-002 E2-00000(00000)
Q3-00013(00000) EXPAN-002 E3-00000(00000)

PROC  0000-087%  CP      PROC  0001-088%  CP
PROC  0002-089%  CP

LIMITED-00000
```

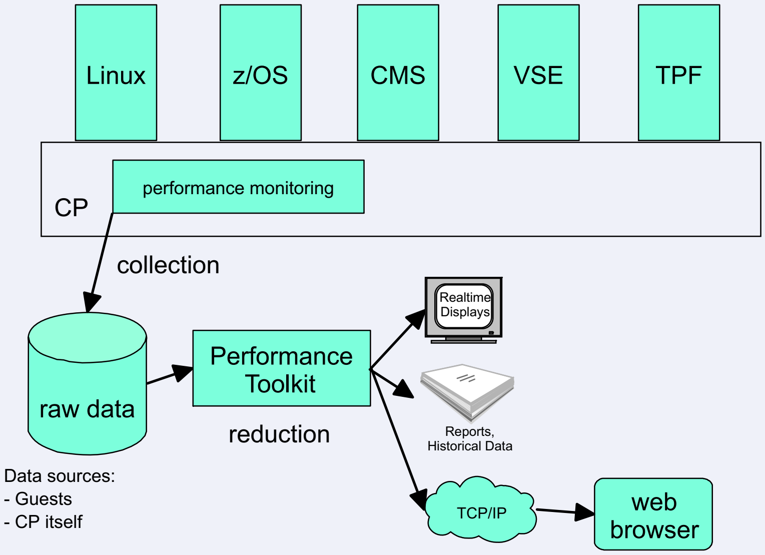
CP INDICATE QUEUES Example

```
INDICATE QUEUE EXP
EDLLIB14      Q3 IO  00002473/00002654 ..D. -.0217 A00
KAZDAKC       Q3 IO  00003964/00003572 .... -.0190 A02
BITNER        Q1 R00 00001073/00001054 .I.. -.0163 A01
LCRAMER       Q3 IO  00003122/00002850 ....  .0259 A00
DSSERV        L0 R   00007290/00007289 ....  .3229 A00
RSCS          Q0 PS  00001638/00001616 .I..  99999 A00
SICIGANO      Q3 PS  00000662/00000662 .I..  99999 A00
VMLINUX1      Q3 PS  00018063/00018063 ....  99999 A02
LNXREGR       Q3 PS  00073326/00073210 ....  99999 A02
VMLINUX       Q3 PS  00031672/00031672 ....  99999 A01
TCPIP         Q0 PS  00018863/00018397 .I..  99999 A02
EDLLNX2       Q3 PS  00032497/00032497 ....  99999 A01
EDLLNX1       Q3 PS  00015939/00015939 ....  99999 A02
```

Selected CP QUERY Commands

- USERS: number and type of users on system
- SRM: scheduler and dispatcher settings (LDUBUF, etc.)
- SHARE: type and intensity of system share
- FRAMES: real storage allocation
- PATHS: physical paths to device and status
- ALLOC MAP: DASD allocation
- ALLOC PAGE: how full your paging space is
- XSTORE: assignment of expanded storage
- MONITOR: current monitor settings
- MDC: MDC usage
- VDISK: virtual disk in storage usage
- SXSPAGES: System Execution Space (z/VM 5.2)

CP Monitor and Performance Toolkit



State Sampling

- Finds the state of a given user or device
- Sampling interval governed by CP MONITOR SAMPLE RATE
- Consolidation of samples gives useful info
  - User: percent of time in various dispatcher states
  - Device: average length of wait queue
- Findings come out in Performance Toolkit reports
  - FCX108 DEVICE
  - FCX114 USTAT

Sample FCX114 (USTAT) Report

```
FCX114   Run  2007/08/15 09:58:26          USTAT
                                           Wait State Analysis  by User
From 2007/08/14 06:35:05
To   2007/08/14 07:55:02
For     4797 Secs 01:19:57                 Result of JW815 Run
-----------------------------------------------------------------------------------------------------------

 .          ----     .     .     .     .     .     .     .     .     .     .     .     .     .     .     .     .     .     .
                                                       <  -SVM and->                                  <-- %Time spent
 Userid     %ACT  %RUN %CPU %LDG %PGW %IOW %SIM %TIW %CFW %TI %EL %DM %IOA %PGA %LIM %OTH   Q0   Q1   Q2  C
 LXM00002    100    12   20    0    8    0    2   52    0   0   0   0    1    0    0    4   98    0    0
 LX00001     100    52   17    0    6    0   17    5    1   0   0   0    0    0    0    2   99    0    0
 LX00062     100    53   17    0    6    0   16    4    3   0   0   0    0    0    0    1  100    0    0
 LX00063     100    51   17    0    6    0   17    4    3   0   0   0    0    0    0    1  100    0    0
 LX00230     100    46   20    0    4    0   21    4    1   0   0   0    0    0    0    4   99    0    0
 LXM00001     99    10   21    0    8    0   12   43    0   0   0   0    1    0    0    5   99    0    0
 LX00101      99    11   25    0    9    0    4   42    0   0   0   0    1    0    0    7   97    0    0
```

I/O Response Time

Resp Time = Queue Time + Service Time

Service Time = Pending + Connect + Disconnect

- **Queue Time**: from high-frequency sampling of queue in RDEV. Reported in monitor.
- **Function Pending**: time accumulated when a path to device cannot be obtained.
  - < 1 ms, unless contention at channels or control units.
- **Connect**: time device logically connected to channel path
  - proportional to amount of data per I/O
- **Disconnect**: time accumulated when device is logically disconnected from channel while channel subsystem is active.
  - Cache miss
  - Seek on older devices
  - CU management
- **Device Active**: time accumulated between return of channel-end and device-end
  - Often reported as part of disconnect time

Sample FCX108 DEVICE Report

```
FCX108  Run 2007/08/15 09:58:15          INTERIM DEVICE
                                         General I/O Device Load and Performance
From 2007/08/14 06:35:09
To   2007/08/14 06:40:08
For    300 Secs 00:05:00                 Result of JW815 Run
------------------------------------------------------------------------------------------------------------
 .     .                     .          __     .    .    .    .    .    .    .    .    .    .    .     .
<-- Device Descr. -->      Mdsk  Pa-  <-Rate/s-> <------ Time(msec) -------> Req.  <Percent>   SEEK
Addr Type  Label/ID        Links ths   I/O Avoid Pend Disc Conn Serv Resp CUWt Qued  Busy READ  Cyls
>>All DASD<<                ....        4.2   .8   .3  3.7  1.7  5.7  5.7   .0   .0     2    0   406
DB16 3390  C FDB16 CP          1    2  35.6   .0   .3   .0  4.0  4.3  4.3   .0   .0    15  100  1482
DB10 3390  C FDB10 CP          1    2  33.4   .0   .3   .0  3.9  4.2  4.2   .0   .0    14  100   933
C098 3390  P C098  CP          1    2  32.2   .0   .3   .1   .8  1.2  1.2   .0   .0     4  100   283
C097 3390  P C097  CP          1    2  32.2   .0   .3   .1   .8  1.2  1.2   .0   .0     4  100  2661
C09E 3390  P C09B  CP          1    2  31.6   .0   .3   .0   .8  1.1  1.1   .0   .0     4  100   221
C09B 3390  P C09A  CP          1    2  31.5   .0   .3   .2   .8  1.3  1.3   .0   .0     4  100    43
9B0D 3390  J SPG15 CP          0    2  29.8   .0   .5   .7  1.3  2.5  2.5   .0   .0     8  100    54
9F23 3390  X PG2   CP          0    3  27.5   .0   .3  4.8  1.4  6.5  6.5   .0   .0    18  100  2032
9F25 3390  X PG4   CP          0    3  27.5   .0   .3  4.7  1.3  6.3  6.3   .0   .0    18  100  2400
9F30 3390  X PG7   CP          0    3  27.1   .0   .3  4.4  1.4  6.1  6.1   .0   .0    17  100  1574
9F2F 3390  X PG6   CP          0    3  27.0   .0   .3  4.8  1.3  6.4  6.4   .0   .0    18  100   112
9F32 3390  X PG9   CP          0    3  26.7   .0   .3  4.9  1.4  6.6  6.6   .0   .0    18  100  2402
9F26 3390  X PG5   CP          0    3  26.5   .0   .3  3.9  1.6  5.8  5.8   .0   .0    16  100  2074
9F24 3390  X PG3   CP          0    3  26.0   .0   .3  3.7  1.4  5.4  5.4   .0   .0    14  100     0
9F22 3390  X PG1   CP          0    3  25.8   .0   .3  4.0  1.2  5.5  5.5   .0   .0    14  100  1844
9F31 3390  X PG8   CP          0    3  25.4   .0   .3  4.8  1.7  6.8  6.8   .0   .0    18  100  1348
DD43 3390  J SPG1A CP          0    2  23.9   .0   .3   .0  4.4  4.7  4.7   .0   .0    11  100   431
DD42 3390  J SPG19 CP          0    2  23.7   .0   .3   .0  4.4  4.7  4.7   .0   .0    11  100   157
DD40 3390  J SPG17 CP          0    2  23.2   .0   .3   .0  4.5  4.8  4.8   .0   .0    11  100    80
D00D 3390  J SPG04 CP          0    2  22.8   .0   .3   .0  2.4  2.7  2.7   .0   .0     6  100  2158
```

On Transactions

- Formal definition: a unit of work performed by the application for which you bought the computer.
  - One HTTP request and response
  - One database update

- VM scheduler definition: a continuous period in which a guest is ready to use CPU, ended by (for example...):
  - Guest becoming not dispatchable (e.g., it loads an enabled wait PSW)

- Performance is the relationship between work done and resources consumed
  - Throughput rates
    - "External" rate: transactions per wall clock second
    - "Internal" rate: transactions per CPU second
  - Processor consumption: CPU-seconds used per transaction performed

- Use the definition (and do the calculation) that makes sense for your situation

Other Sources

- z/VM Performance manual
  - SC24-5999-02 -- z/VM 4.4.0
  - SC24-6109-00 -- z/VM 5.1.0
  - Part of the z/VM Library
- http://www.vm.ibm.com/perf/
  - Links to documents, tools, reference material
- http://www.vm.ibm.com/perf/tips/
  - Common problems and solutions
  - Guidelines
- http://www.vm.ibm.com/devpages/bitner/
  - Presentations with speaker notes

A Case Study

- Customer calls in
- My system isn’t running fast, but it isn’t paging either
- My application formats lots of VDISKs... aren’t they in memory? Shouldn’t this be fast?
- I have raw monitor data... will you take a look?
- Customer sent raw monitor file 20070501 MD111606
- He says his workload uses disk volumes 1240-59 and 16C0-E3
- I took a look-see

Basic System Summary

```
FCX225   Run  2007/05/02 12:56:34              SYSSUMLG
                                               System Performa   nce Summary by Time
From 2007/05/01 11:16:08
To     2007/05/01 12:37:10
For     4861   Secs 01:21:01                   Result of     20070501 Run
--------  --------------------    --------  --------  --------  --------  ---------------    -------

               <------- CPU        --------      > <Vec> <-  -Users--   > <---I/    O---> <Stg>     <-Paging-->
                     <--Ratio-->                                         SSCH   DASD Users <-Rat    e/s-->
 Interval       Pct       Cap-           On-      Pct     Log-           +RSCH   Resp     in PGIN+ Read+
 End Time       Busy  T/V  ture          line     Busy     ged Activ        /s   msec  Elist PGOUT Write
 >>Mean>>       10.3 106.3 .7577 27       .0      ....     280     263 122.  7   11.1       .0  5418  1445
 11:17:08        8.8  5.43 .8412        27.0      ....     280     261    88.5    1.1      .0    .0   4.0
 11:18:39        9.7 12.49 .8054 27       .0      ....     280     261    44.6     .7      .0  2195    .2
 11:19:11        8.7 114.1 .7863 27       .0      ....     280     257    47.5     .8      .0  3852    .0
 11:22:40       10.3 163.6 .8114 27       .0      ....     280     267    39.9     .7      .0  3252    .0
 11:23:41        9.9 180.7 .8232 27       .0      ....     280     263    25.8     .8      .0  2645    .0
 11:24:40       10.3 193.5 .8051 27       .0      ....     280     263    23.8     .7      .0  2707    .0
 11:25:39       10.5 196.8 .8218 27       .0      ....     280     262    23.6     .8      .0  2825    .0
 11:27:10        9.7 159.5 .8232 27       .0      ....     280     262    29.9     .7      .0  3714    .0
 11:28:09        9.8 108.2 .8015 27       .0      ....     280     266    48.4     .8      .0  8942    .1
 11:29:40        9.8 119.2 .8134 27       .0      ....     280     264    33.2     .9       .0  8602   2.8
 11:36:10       10.3 119.6 .8048 27       .0      ....     280     263    45.7     .6      .0  9327    .0
 11:37:40       10.5 136.8 .8028 27       .0      ....     280     262    30.3     .6      .0  9213    .0
 11:39:10       10.8 144.2 .8158 27       .0      ....     280     264    30.7     .7      .0  9189    .0
 11:40:40       10.5 135.6 .8093        27.0      ....     280     264    32.5     .7     .0 10083          .0
 11:41:39       10.7 166.5 .8124 27       .0      ....     280     262    25.2     .8      .0  8942    .0
 11:42:41       10.2 167.6 .8070 27       .0      ....     280     262    23.0     .7      .0  9311    .0
```

Look at those T/V ratios! What is CP doing?

Think About the Application

- Customer says he is formatting VDISKs
- VDISKs are address spaces
- We page them when storage gets tight
- We do seem to be spending a lot of time in CP
- Let’s see if DEVICE CPOWNED shows us anything

DEVICE CPOWNED

```
FCX109   Run 2007/05/02 12:56:34          DEVICE   CPOWNED
                                          Load and Performance of  CP Owned Disks
From 2007/05/01 11:16:08                                                                  20070501
To   2007/05/01 12:37:10                                                                  CPU 2094
For     4861  Secs 01:21:01               Result of   20070501 Run                        z/VM    V
--------------------------------------------------------------------------------------------------

Page / SPOOL Allocation Summary
 PAGE slots available        34745k        SPOOL slots available        3656598
PAGE slot utilization            3%        SPOOL slot utilization            9%
T-Disk cylinders avail.    .. .. ..        DUMP slots available               0
 T-Disk space utilization      ... %       DUMP slot utilization            ..%

__ __ .           .                      .      .      .      .      .      .     .      .     .     .      .     .
<Device Descr. ->                        <-- -- -- --- -- -- Rate/s -- -- -- --- -- --> User         Serv  MLOAD Block %Used
          Volume Area    Area      Used  <-- Page-- ><- -Spool- ->          SSCH Inter Queue   Time  Resp  Page    for
Addr Devtyp Serial Type  Extent       %  P-Rds  P-Wrt S-Rds S-Wrt Total     +RSCH feres Lngth  /Page  Time  Size Alloc
1240 3390  JSPG20 PAGE    0-3338      3    1.2   17.6   ...   ...  18.8      1.4      1     0   3.8    3.8   14    44
1241 3390  JSPG21 PAGE    0-3338      3    1.3   16.8   ...   ...  18.1      1.3      1     0   7.8    7.8   14    42
1242 3390  JSPG22 PAGE    0-3338      3    1.3   17.4   ...   ...  18.6      1.3      1   .57   6.7    9.0   14    43
1243 3390  JSPG23 PAGE    0-3338      2    1.3   16.2   ...   ...  17.5      1.3      1  1.08   5.2   11.0   14    40
1244 3390  JSPG24 PAGE    0-3338      2    1.3   16.4   ...   ...  17.7      1.3      1  1.16   5.0   11.5   14    41
1245 3390  JSPG25 PAGE    0-3338      2    1.2   15.9   ...   ...  17.1      1.3      1   .57   5.6    8.6   14    40
1246 3390  JSPG26 PAGE    0-3338      2    1.3   15.7   ...   ...  17.0      1.2      1     0  12.5   12.5   14    39
1247 3390  JSPG27 PAGE    0-3338      2    1.3   15.4   ...   ...  16.7      1.2      1  1.49   5.9    9.3   14    38
1248 3390  JSPG28 PAGE    0-3338      2    1.3   15.5   ...   ...  16.8      1.2      1  1.08   9.9   14.7   14    39
1249 3390  JSPG29 PAGE    0-3338      2    1.2   14.9   ...   ...  16.1      1.2      1   .49   9.1    9.3   14    37
124A 3390  JSPG2A PAGE    0-3338      3    1.3   17.3   ...   ...  18.6      1.3      1  1.19  13.1   19.3   14    43
124B 3390  JSPG2B PAGE    0-3338      2    1.2   16.2   ...   ...  17.3      1.3      1     0   7.9    7.9   14    40
124C 3390  JSPG2C PAGE    0-3338      2    1.1   15.7   ...   ...  16.9      1.2      1     0   8.2    8.2   14    39
124D 3390  JSPG2D PAGE    0-3338      3    1.1   16.6   ...   ...  17.7      1.3      1   .54   4.4    4.5   14    41
16D5 3390  JSPG0A PAGE    0-3338      2    1.2   15.9   ...   ...  17.1      1.2      1  1.38   8.5   14.8   14    39
16D6 3390  JSPG0B PAGE    0-3338      3    1.2   16.5   ...   ...  17.7      1.3      1  2.62  16.7   20.9   14    41
```

From 11:16 to 12:37 the paging devices have queues *on average*?
Let's look at some INTERIM reports and see what we see...

INTERIM DEVICE, 11:47

```
1FCX108 Run 2007/05/02 12:56:29                 INTERIM DEVICE
                                                General I/O Device  Load and Performance
 From 2007/05/01  11:45:39
 To   2007/05/01  11:47:37
 For     118 Secs  00:01:58                     Result  of 20070501  Run

---- --- ---- ------ ------ ------ --- --- --- --- --- ---- ------ ------ --- --- --- --- --- ---- ------ ----

 .      .       .          .   __ _     .    .    .    .     .     .    .    .    .    .
 <- -Device Descr. -->   Mdisk Pa-  <-Rate/s-> <-- ----  Time (msec) -- --- --> Req.  <Percent>
 Addr Type  Label/ID     Links ths  I/O  Avoid  Pend Disc Conn Serv  Resp CUWt Qued  Busy READ
 1240 3390  JSP G20 CP      0   2   1.3   .0    47.3  .9   5.4  53.6  53.6  .0   .0    14     0
 16DE 3390  JSP G0E CP      0   2   1.3   .0    48.9  .1   6.5  55.5  55.5  .0   .0    12   100
 16E0 3390  JSP G0F CP      0   2   1.3   .0    53.8  .6   7.0  61.4  61.4  .0   .0    12     0
 16D9 3390  JSP G0D CP      0   2   1.3   .0    53.3  .9   6.1  60.3  60.3  .0   .0    14   100
 16DF 3390  JSP G09 CP      0   2   1.3   .0    49.9  .0   7.1  57.0  57.0  .0   .0    11   100
 16DC 3390  JSP G07 CP      0   2   1.2   .0    50.7  .0   6.5  57.2  57.2  .0   .0    12   100
 1247 3390  JSP G27 CP      0   2   1.2   .0    52.2  .7   6.4  59.3  75.0  .0   .0    15     0
 16DB 3390  JSP G06 CP      0   2   1.2   .0    51.6  .0   7.0  58.6  58.6  .0   .0    12     0
 16DD 3390  JSP G08 CP      0   2   1.2   .0    54.6  .4   7.2  62.2  62.2  .0   .0    13     0
 16D8 3390  JSP G0C CP      0   2   1.2   .0    54.7  .0   6.6  61.3  61.3  .0   .0    13   100
 1241 3390  JSP G21 CP      0   2   1.2   .0    48.9  .8   7.0  56.7  56.7  .0   .0    13     0
 16D6 3390  JSP G0B CP      0   2   1.1   .0    55.7  .5   6.9  63.1  63.1  .0   .0    13     0
 1242 3390  JSP G22 CP      0   2   1.1   .0    45.5  .0   7.3  52.8  52.8  .0   .0    12     0
 1245 3390  JSP G25 CP      0   2   1.1   .0    54.8  .1   6.9  61.8  61.8  .0   .0    13     0
 16D5 3390  JSP G0A CP      0   2   1.1   .0    59.1  .0   6.6  65.7  65.7  .0   .0    13     0
 124C 3390  JSP G2C CP      0   2   1.1   .0    55.4  .0   7.3  62.7  62.7  .0   .0    14     0
 1246 3390  JSP G26 CP      0   2   1.0   .0    60.3  .0   7.0  67.3  67.3  .0   .0    13     0
 124B 3390  JSP G2B CP      0   2   1.0   .0    53.9  .0   7.2  61.1  61.1  .0   .0    13     0
 124D 3390  JSP G2D CP      0   2   1.0   .0    53.1  .3   5.9  59.3  59.3  .0   .0    12    50
 124A 3390  JSP G2A CP      0   2   1.0   .0    46.7  .0   7.5  54.2  54.2  .0   .0    12     0
 1243 3390  JSP G23 CP      0   2   1.0   .0    61.6  .0   6.6  68.2  68.2  .0   .0    13     0
 1244 3390  JSP G24 CP      0   2    .9   .0    58.8  .0   7.1  65.9  65.9  .0   .0    13     0
 1249 3390  JSP G29 CP      0   2    .9   .0    62.0  .0   7.1  69.1  69.1  .0   .0    13     0
 1248 3390  JSP G28 CP      0   2    .8   .0    70.3  .5   7.6  78.4  78.4  .0   .0    13     0
```

Look at that pending time on the paging volumes!
High pending time usually means channel contention...

Configuration

From FCX131 DEVCONF:

```
1240-1259  0008-0021     3390-3 (E)  67 69  . . .              . . .  2105        -E8   Online
16C0-16E3  0050-0073     3390-3 (E)  67 69  . . .              . . .  2105        -E8   Online
```

Two ESCON chpids for all this paging DASD?
I don’t think so...

Recommendation

- I told customer he need a lot more channel capacity to his paging DASD
- Customer added four ESCON chpids
- (Why didn’t he add FICON? Who knows...)
- He was quiet for a while, and then...

He’s Baa-aaack

```
FCX109  Run 2007/08/15 09:58:19            INTERIM DEVICE  CPOWNED
                                           Load and Performance of  CP Owned Disks
From 2007/08/14 07:15:03                                                                       JW815
To   2007/08/14 07:20:02                                                                       CPU 209
For    299 Secs 00:04:59                   Result of  JW815 Run                                z/VM
____________________________________________________________________________________________________

 Page/SPOOL Allocation Summary
 PAGE slots available        51540k         SPOOL slots available        4257606
 PAGE slot utilization           53%        SPOOL slot utilization           24%
 T-Disk cylinders avail.     .. .. ...      DUMP slots available               0
  T-Disk space utilization       ...%       DUMP slot utilization           ..%

 __ __ .                 .                        .     .     .     .     .     .     .     .     .     .     .
 <Device Descr.->                            <-------- Rate/s --------------> User          Serv  MLOAD Block %Used
                Volume Area    Area     Used <--Page---><--Spool-->      SSCH Inter Queue   Time  Resp  Page   for
 Addr Devtyp Serial Type    Extent       %   P-Rds P-Wrt S-Rds S-Wrt Total +RSCH feres Lngth /Page  Time  Size  Alloc
 16D5 3390   JS PG0A PAGE    0-3338   88    21.7  19.1  ...   ...   40.8  15.5     1 33.00   1.0    2.9     4     49
 16D6 3390   JS PG0B PAGE    0-3338   88    20.5  17.2  ...   ...   37.7  15.1     1 19.00   2.2   42.5     4     44
 16D8 3390   JS PG0C PAGE    0-3338   88    22.7  18.1  ...   ...   40.7  15.8     1 22.00   1.2   28.7     4     45
 16D9 3390   JS PG0D PAGE    0-3338   87    21.1  18.5  ...   ...   39.6  15.2     1 29.00    .8   25.0     4     48
 16DB 3390   JS PG06 PAGE    0-3338   87    22.3  20.0  ...   ...   42.3  15.6     1 20.00    .8   17.1     4     51
 16DC 3390   JS PG07 PAGE    0-3338   86    21.9  17.7  ...   ...   39.6  15.7     1 10.00    .9   10.4     3     45
 16DD 3390   JS PG08 PAGE    0-3338   86    22.0  18.2  ...   ...   40.3  15.5     1 106.0    .8    5.9     4     47
 16DE 3390   JS PG0E PAGE    0-3338   86    21.4  19.6  ...   ...   41.0  15.0     1     0    .6     .6     4     48
 16DF 3390   JS PG09 PAGE    0-3338   84    22.1  19.6  ...   ...   41.7  14.2     1 17.00   1.0   18.4     5     50
 16E0 3390   JS PG0F PAGE    0-3338   83    20.4  17.6  ...   ...   38.1  12.4     1 63.00   2.2  139.3     5     44
 5805 3390   CF 5805 PAGE    810000   12    46.5  41.9  ...   ...   88.4  21.3    10     0    .1     .1    11    100
 9F23 3390   XP G2   PAGE    0-3338   99    18.6  18.1  ...   ...   36.7  25.9     1 23.00    .7   16.9     2     47
 9F24 3390   XP G3   PAGE    0-3338   99    19.2  17.5  ...   ...   36.6  25.8     1 29.00    .6   19.2     2     46
 9F25 3390   XP G4   PAGE    0-3338   99    18.6  17.4  ...   ...   36.0  26.9     1     0    .6     .6     1     46
 9F2F 3390   XP G6   PAGE    0-3338   99    20.9  17.9  ...   ...   38.8  27.1     1 35.00    .6   20.6     2     47
 C09E 3390   PC 09B  PAGE    0-3338  100    22.4  19.2  ...   ...   41.6  30.2     1     0    .6     .6     1     98
 D007 3390   CF D007 PAGE    896800   17    46.1  40.7  ...   ...   86.8  19.9     1 30.00    .1     .1    11     99
 D008 3390   CF D008 PAGE    896800   17    42.2  39.7  ...   ...   81.9  18.1     1 32.00    .2     .2    11     99
 D00D 3390   JS PG04 PAGE    896800   20    42.9  39.0  ...   ...   81.9  18.5     1     0    .3     .3    12    100
```

I removed 25 100%-full 3990-3 volumes from this excerpt!

So What’s His Problem Now?

- 40 3390-3 paging volumes nearly full
- 4 3390-9 paging volumes have the free space
- We can do only one I/O at a time to those gigantic model 9’s
- Get rid of those mod 9’s and add a lot of mod 3’s
- He’s working on it

Some Final Thoughts

- Routinely collect data. This records "good performance".
- Implement a change management process.
- Make as few changes as possible at a time.
- Performance is often only as good as the weakest component.
- Relieving one bottleneck will reveal another. As attributes of one resource change, expect at least one other to change as well.

Old Charts

Gone but not forgotten!

IBM System z Expo
September 17-21, 2007
San Antonio, TX

The Grinch That Stole Performance

From VMPRF USER_STATES_BY_TIME PRF007 Report January 5:

```
<----Percent of True Non-Dormant Time Waiting on--------------->
                                        <---SVM and---->  I/O
    Load-                 Inst  Test  Cons  Test Elig-  Dor-   Ac-
CPU   ing  Page   I/O     Sim  Idle  Func  Idle  ible  mant  tive

0.1   0.1   0.1  18.8    2.3  10.0   0.4   3.4     0  50.8   8.4
0.1     0   0.1  16.0    1.9   9.9   0.4   3.1     0  53.8   9.9
```

From VMPRF DASD_BY_ACTIVITY PRF012 Report January 5:

```
      SSCH   Pct    <-----------Time------------> <--Queue-->
Dev.  Rate   Busy   Pend   Disc   Conn   Serv   Resp   Mean    Max

1742  26.7   65.4    1.3   18.4    4.7   24.5   69.0    1.2    8.5
```

Went to check VMPRF DASD_BY_ACTIVITY_EF PRF095 for control unit cache stats, but it didn't exist!

It is a good thing I keep historical data -- let's go back and see what's going on...

When Did We Last See Cache?

From VMPRF DASD_BY_ACTIVITY PRF012 Report from December 8:

```
      SSCH   Pct    <-----------Time------------> <--Queue-->
Dev.  Rate   Busy   Pend   Disc   Conn   Serv   Resp   Mean    Max
1742  41.0   10.5    0.3    0.2    2.0    2.6    2.9    0.0    0.3
Jan5: 26.7   65.4    1.3   18.4    4.7   24.5   69.0    1.2    8.5
```

VMPRF DASD_BY_ACTIVITY_EF PRF095 Report for 1742 on Dec 8:

```
<---------Rate--------> <-----Percent----------->
Total  Read  Read Write      <---------Hits----->
  I/O NonSq   Seq    FW Read   Tot Read  Wrt  DFW

 53.0  52.3     0   0.6   99    99   99   96   96
```

No _EF report at all now? This means there is no cache now.

Performance Toolkit Device Report

```
FCX110      CPU 2003   GDLVM7    Interval INITIAL. - 13:08:47     Remote
Data

Detailed Analysis for Device 1742 ( SYSTEM )
Device type :  3390-2       Function pend.:     .8ms      Device busy   :
27%
VOLSER      :  USE001       Disconnected  :   20.3ms      I/O contention:
0%
Nr. of LINKs:     404       Connected     :    5.4ms      Reserved      :
0%
Last SEEK   :    1726       Service time  :   26.5ms      SENSE SSCH    :
...
SSCH rate/s :    10.5       Response time :   26.5ms      Recovery SSCH :
...
Avoided/s   :    ....       CU queue time :     .0ms      Throttle del/s:
...
Status: SHARABLE

Path(s) to device 1742:     0A     2A     4A
Channel path status    :    ON     ON     ON

Device                 Overall CU-Cache Performance           Split
DIR ADDR VOLSER        IO/S %READ %RDHIT %WRHIT ICL/S BYP/S   IO/S %READ %RDHIT
```

Down for the 3-Count

```
q dasd details 1742
1742 CUTYPE = 3990-EC, DEVTYPE = 3390-06, VOLSER= USE001
       CACHE DETAILS:  CACHE NVS CFW DFW PINNED CONCOPY
            -SUBSYSTEM    F     Y   Y   -     Y
            -DEVICE       Y     -   -   Y     N        N
       DEVICE DETAILS: CCA = 02, DDC = 02
       DUPLEX DETAILS: SIMPLEX
```

Pinned data! Yikes! I had never seen that before!

Performance Toolkit Device Report

What volumes are on rdev 1742?

```
 MDISK Extent     Userid   Addr IO/s VSEEK Status    LINK MDIO/s
+-----------------------------------------------------------------+
|   101 -  200     EDLSFS   0310   .0      0 WR           1      .0 |
|   201 -  500     EDLSFS   0300   .0      0 WR           1      .0 |
|   501 -  600     EDLSFS   0420   .0      0 WR           1      .0 |
|   601 - 1200     EDLSFS   0486   .0      0 WR           1      .0 |
|  1206 - 1210     RAID     0199   .0        owner                  |
|                  BRIANKT  0199   .0      0 RR           5      .0 |
|  1226 - 1525     DATABASE 0465   .0        owner                  |
|                  K007641  03A0   .0      0 RR           3      .0 |
|  1526 - 1625     DATABASE 0269   .0        owner                  |
|                  BASILEMM 0124   .0      0 RR          25      .0 |
|  1626 - 1725     DATABASE 0475   .0        owner                  |
|                  SUSANF7  0475   .0      0 RR           1      .0 |
|  1726 - 2225     DATABASE 0233   .0      0 owner      366    10.5 |
+-----------------------------------------------------------------+
```

DATABASE 233 is key to our source code library.

Solution

- Use **Q PINNED** CP command to check for what data is pinned.
- Discussion with DASD Management team.
- Moved data off string until corrected.

Pinned data is very rare, but when it happens it is serious.