VM/ESA Performance Update

SHARE 89 - Summer Meeting 1997

Session 9220

Bill Bitner
IBM Corp.
1701 North St.
Endicott, NY 13760
(607) 752-6022
bitner@vnet.ibm.com
USIB1E29 at IBMMAIL

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
- MVS/ESA
- SQL/DS
- OfficeVision

Introduction

- VM/ESA 2.2.0 Performance
  - Changes that affect performance
  - Measurement results
- Other measurements of interest
  - OV/VM 1.3.0
  - VisualGen compared to CSP
  - Development workloads compared to LSPR

VM/ESA 2.2.0 is an exciting release and we'll discuss the changes that make this release special. This presentation will also briefly share some results from measurements we did.

Changes that Affect Performance

- VMCF moved off master processor
- EXEC 2 converted to compiled REXX
- CMSINST moved above 16M line
- Reduced CMS WSS
- GCS Preallocated storage
- CMS GUI Improvements
- Segment changes
- Monitor enhancements

This is a summary list of the changes in VM/ESA 2.2.0 that affect performance. It is a relatively short list compared to some previous releases, but I think there are some very significant items on the list. The VMCF change is particularly noteworthy. Each of these will be discussed in more detail later in this presentation.

Move VMCF off Master

- VMCF no longer serialized by master processor.
- The VMCF pathlength is relatively short by itself.
- Helps some TCP/IP and OV/VM workloads.
- Will not help VMCF traffic that results from SMSG
- Results from pure VMCF workload on 9121-742 (4-way):

VM/ESA Release	2.1.0	2.2.0	Delta	%Delta
VMCF Rate (per sec)	14389	26179	11790	81.94%
CPU/VMCF (msec)	0.245	0.153	-0.092	-37.55%
Master Processor	99.5	100	0.5	0.50%
Alternate Processors	84.5	100	15.5	18.34%

The negative impact of VMCF serialization on the master processor had increased as workloads used more VMCF (in particular parts of TCP/IP and OV/VM). You might recall the infamous University of Illinois at Chicago measurements that showed potential master processor limitations on a 3-way (when the dispatch slice was set for an MVS environment instead of an interactive environment). With VM/ESA 2.2.0, VMCF has full multiprocessor locking capability. Therefore VMCF functions no longer need to run on the master. There is potential for increased throughput and improved response time on systems that are master processor constrained and have applications that use VMCF heavily. The actual pathlength for VMCF is relatively short. However, once we move to the master processor, we tend to stay there until there is a reason to move off. It is important to note that this change does not help VMCF that results from SMSG calls, since SMSG itself is serialized by the master processor.

Measurements were made with a pure VMCF workload on a 9121-742 (4-way processor). The results are shown in the table. By looking at the 2.1.0 column for the processor utilization rows, one can see that there are untapped cycles on the alternate processors, while the master is running at 100%. If this was a pure capacity or ITR question, then we would expect only an improvement of 10 to 15% more VMCFs per second. However, the results showed a VMCF rate increase of almost 82%! This is due to much less system overhead in tasking switching, queuing, and better HSB cache efficiency. This is seen in the 38% decrease in CPU per VMCF.

These results reflect an extreme case, but does prove the change worked as expected.

EXEC 2 Converted to REXX

- Remaining productivity aids converted from EXEC2 to compiled REXX.
- Performance Benefits:
  - CMSINST only contains REXX, so it can be above the 16M line.
  - Reduced processor time.
  - Also made some performance oriented design changes.
- Uncompiled source is still provided on S-disk
- A REXX run time library is provided (as in VM/ESA 2.1.0), but not the compiler

In VM/ESA 2.1.0, we introduced compiled REXX to the base VM. This meant a special REXX run time library (not the actually compiler) was now shipping with VM/ESA. In VM/ESA 2.1.0, we shipped the current REXX productivity aids as compiled REXX. When asked about the EXEC2 programs, I had to smile and say watch this space. Well, here you go. We've rewritten the remaining EXEC2 code to REXX so that it can be compiled and save cycles as well. In the process of rewriting this code, other performance or maintainability enhancements could be made. The source still ships to allow for local mods.

EXEC 2 Converted to REXX *(continued)*

- RDRLIST and PEEK measurements showed processor savings of 22 to 27%.
  - Performance of PEEK for spool files created with DISK DUMP command greatly improved (90% for files >8000 records)
  - FILELIST implementation results in improved performance, proportional to the number of files.
    - 10 to 40% processor reduction
  - Over 2 dozen more execs and macros added to CMSINST.

Some of the most significant improvement was seen in RDRLIST and PEEK, where over 20% processor savings were obtained. In particular, the performance of PEEK for DISK DUMP created files was greatly enhanced.
FILELIST also saw significant processor usage reductions. These reductions are proportional to the number of files involved. The more files there are, the greater the savings.
With this release, all the productivity aids programs are now REXX and therefore the CMSINST segment can reside above the 16MB line. In addition to this, we've added several other execs and macros to the segment for added performance.

Other Performance Enhancements

- Reduced CMS Working Set Size
  - Avoids one time storage references when CMS Multitasking is not used.
- GCS Preallocated Storage
  - Eliminates most of OS GETMAIN/FREEMAIN requests by GCS saving processor resources.
- CMS GUI Improvements (also see 9606 RSU for 2.1.0)
  - XEDIT and DtEventNotify

In VM/ESA 2.1.0, there was some increase in the CMS working set size due to the allocation and initialization of control blocks associated with CMS multitasking functions. This storage was used even in the case where multitasking was actually not in use. In VM/ESA 2.2.0, the allocation of these control blocks are avoided. This lowers the virtual storage requirements. For the CMS workloads, the WSS decreased about 3 to 7 pages with the total pages per user decreasing by 24 to 30 pages. Systems with expanded storage for paging tend to see a larger improvement with this enhancement.

Most of the OS GETMAIN/FREEMAIN requests made by GCS have been eliminated through the use of preallocated storage. This lowers the processor resources used for storage management. In our FS8F CMS environment, this resulted in a 0.3% decrease in CPU usage.

CMS GUI saw two areas of improvement for performance. Performance of XEDIT from the CMSDESK GUI application was enhanced to improve response time by over 40% and host processor CPU by over 60% for measured cases. DtEventNotify is one of the functions provided by the distributed GUI toolkit (DT) API. The processing required by DtEventNotify was reduced by up to 60%. Most of the XEDIT enhancements and the DtEventNotify change were shipped as part of the 9606 RSU for VM/ESA 2.1.0.

Monitor and Performance Products

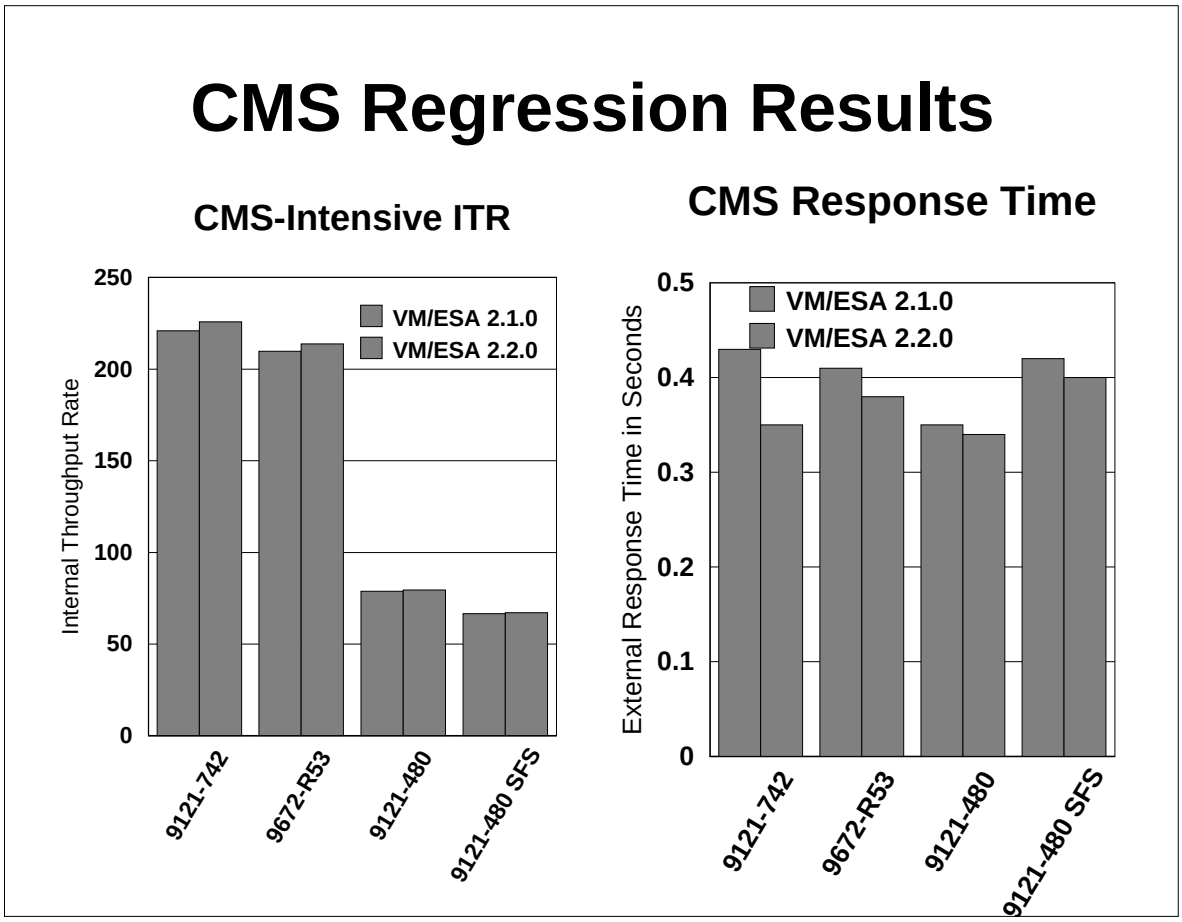
- Monitor Enhancements
  - Scheduler Data
  - Master Processor Data
- VMPRF with VM60882
  - >30,000 users
  - 4 new reports
  - Manual on March 1997 Collection Kit
- RTM/ESA requires APAR GC05398 for VM/ESA 2.2.0
- FCON/ESA enhanced for new monitor data in VM/ESA 2.2.0

Monitor enhancements were made to address user requirements and make system management easier. These include scheduler and master processor data. While scheduler data has been part of the monitor for a long time, key fields such as ATOD and ATOD2 which were missing have been added in VM/ESA 2.2.0. The monitor reports on the number of work items in the Processor Local Dispatch Vector (PLDV). The PLDV is a subset of the dispatch list which holds ready to run work. Each processor has its own PLDV, plus there is an extra PLDV for master only work. In the past, when reporting on the master processor the sum of both the normal and master only PLDVs was used. In VM/ESA 2.2.0, there is a separate count for the master only PLDV. For additional details on the monitor changes, see the MONITOR LIST1403 file.

VMPRF has been enhanced via APARs. With APAR VM60882 or newer service, support is added for dealing with large numbers of users (>30,000), 4 new reports (MINIDISK_CACHE_USAGE_BY_TIME, SYSTEM_FACILITIES_BY_TIME, SFS_BF_REQUESTS_BY_TIME, and USER_CONFIGURATION). In addition, the VMPRF manual has been updated and will be on the March 1997 Collection Kit.

RTM/ESA requires APAR GC05398 for VM/ESA 2.2.0. In addition, there are previous APARs that improve RTM: GC05405 (Dynamic I/O survival) and GC05404 (more data above the 16MB line).

CMS Regression Results



In general, these graphs are uninteresting; but sometimes it is good for performance to be uninteresting. For the CMS regression environment, as measured by the FS8F, workload you see that ITR values are roughly equivalent. Response times follow a similar trend where some improvement is seen in the 9121-742 and 9672-R53 measurements. While several factors are involved here, one key to the difference in the environments is the use of expanded storage for paging in the 9121-742 and 9672-R53 environments. If you look back at the "Reduced CMS Working Set Size" change, you may recall that we discussed the improvements here being more significant for systems paging to expanded storage. This is because there are fewer dead pages that have to flow through expanded storage to paging DASD.

Other VM/ESA 2.2.0 Results

- OV/VM 1.3.0 on 9121-480 Regression
  - No significant change
  - VMCF change insignificant here on 2-way without dominating VMCF workload.
  - EXEC2 conversion to REXX is not as beneficial for OV/VM.
- VSE/ESA 2.1.0 on 9121-320 Regression
  - V=R and V=V measurements
  - Equivalent to VM/ESA 2.1.0

We measured two other regression environments and both of them were also fairly uninteresting. OV/VM 1.3.0 was measured with the IOB workload. There was no significant change in performance here. Note that this was a 2-way configuration and that VMCF, while used, does not dominate the workload. Also, OV/VM replaces many of the standard CMS productivity aids that were rewritten in REXX, so that benefit is not seen here.

The VSE/ESA regression environment showed no changes to performance with VM/ESA 2.2.0 for VSE guest environments.

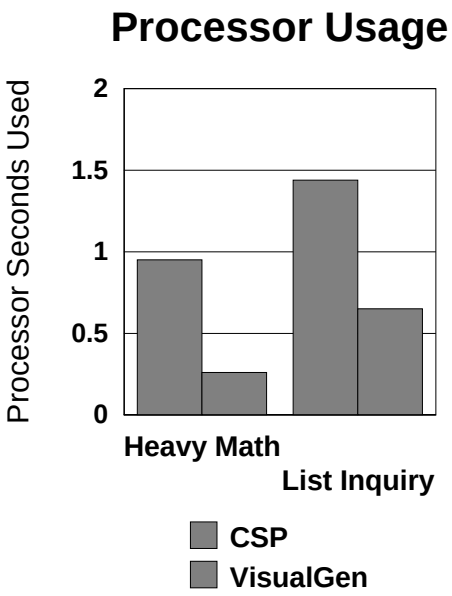
OV/VM Release 1.3.0

- OV/VM 1.3.0 compared to 1.2.0
- 9121-480
- Equivalent
- Did not use the Calendar Feature in either case

You might have noticed that the previous foil talked about OV/VM measurements, and that it has been a few releases since we (VM performance) have measured OV/VM in Endicott. Well we didn't do it just for the release regression. The workload was dusted off to check on the performance of OV/VM itself. We measured releases 1.2.0 and 1.3.0 and saw equivalent performance for the IOB workload. Note that the calendar feature was not used in these measurements.

VisualGen compared to CSP

- Comparison of CSP/AE 3.3 and VisualGen Host Services 1.1.0
- Processor savings due to CSP being interpretive vs. compiled VisualGen
- Linkage tradeoffs
  - type dynamic is quicker but does not release storage
  - type noncsp is slower and releases storage

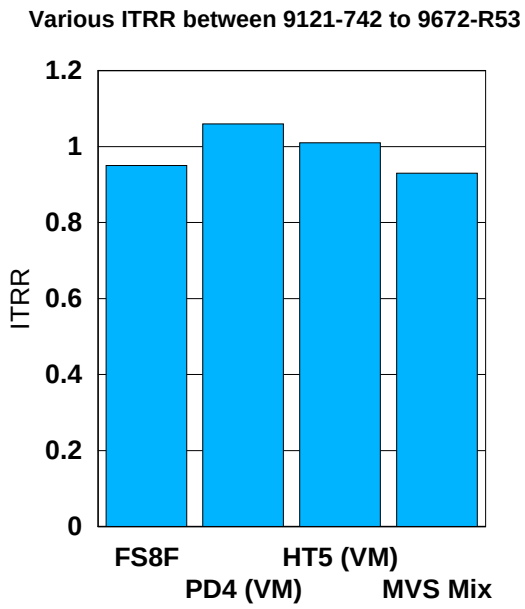


Some measurements were made to compare VisualGen Host Services 1.1.0 to CSP/AE 3.3. While the performance differences are very workload dependent, VisualGen should generally use less processor resources. This is mostly due to the fact that CSP is interpretive while VisualGen is compiled. Therefore compute intensive applications will benefit the most from VisualGen. The graph shows the results of two tests where equivalent function applications were measured in both environments. The Heavy Math test case does complex mathematical calculations with little I/O. The List Inquiry test involves retrieving and displaying data from a VSAM database, and is more I/O intensive.

In a VisualGen environment, there is a trade off to be made in regards to linkage. Linkage type "noncsp" will tear down the old environment and in the process release the storage associated with it. This processing would require extra processor time to accomplish this. Linkage type "dynamic" does not release the old environment and therefore avoids the associated overhead. A VisualGen environment that only uses "dynamic" linkage runs the risk of greatly increasing virtual storage requirements. There should be some "noncsp" linkages in the tree somewhere. One suggestion is in linkages off the main menu. For linkages in hot paths, you might want to use "dynamic" linkages to lower processor overhead. In addition, use of shared segments for LE/370 is recommended.

LSPR Workloads on 9672

- LSPR ITRR going to 9672 from certain machines appear better for VM than MVS
- FS8F workload showed results closer to MVS trends.
- Check MVS numbers for worse case when migrating from 3090, 9121, 9021 to 9672 and 2003 processors.

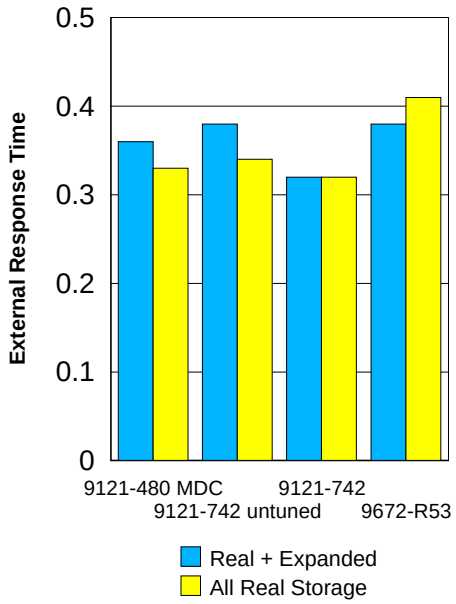


When the Large Systems Performance Reference (LSPR) VM measurements were made on the first 9672s, the results showed that VM did better than MVS on these new processors. However, after some real field experience we found that some customers were seeing ITR ratios more in line with MVS. We have recently measured the VM development workload (FS8F) on a 9672-R53 and a 9121-742 and compared the results to the LSPR workloads (HT5 and PD4). We have found that FS8F is more in line with the MVS results.

It is our recommendation that when migrating from a 3090, 9121, or 9021 to a 9672 or 2003 processor, that you check the MVS ITR ratios as well as the VM workloads.

Storage configured as Real vs. Expanded

- It Depends
- If expanded storage just for MDC, then use all real storage
- expanded storage for paging allows high speed paging for short periods of idleness
- Watch for transaction skews (broken transaction when DASD Paging.
- In untuned environment (no BIAS), then real storage can be better.



This is a question on which I am reversing previous thoughts and recommendations. In the past, I strongly recommended that if you had a processor where you defined how much of the processor storage was real or expanded storage, that you should define as much as possible as real storage. There are cases where that is appropriate, but we have learned of situations where it is not.

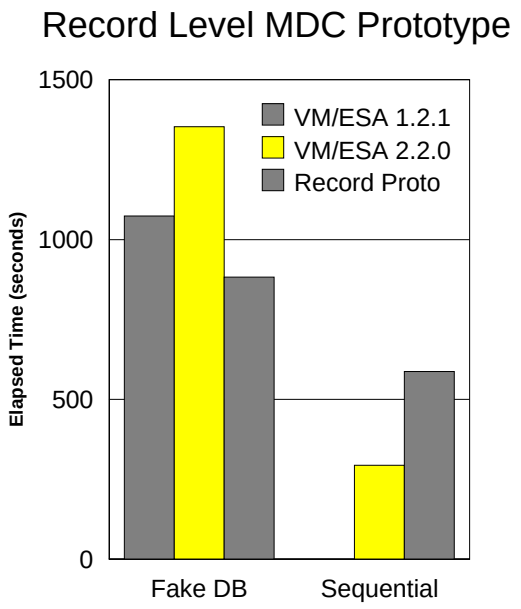
If you had expanded storage configured and were only using it for MDC, then having that configured as real storage is beneficial. This is seen in the first series of bars in the graph above. On that 9121-480, expanded storage was used only for MDC. This configuration was often the case on older 9221s which in addition had less efficient paths for simulation of expanded storage than current technology. Also, in the case where limits or bias settings were not used where they could, the configuration of storage as real storage was more flexible. This is seen on the second set of bars marked "9121-742 untuned".

The last two sets of bars show cases where running without expanded storage made no difference or could potentially hurt performance. Much of this is gated by how much paging to DASD results from the configuration. Since we only page due to a lack of real storage, it would make sense that you configure everything as real storage. That's what I thought in 1995. However, think of the times you page data into storage. It can be after being idle for long (minutes) or short (between or within a transaction) periods of times. You would like the delays between short periods to be short, where as a delay when coming back from coffee is more acceptable. Paging to/from expanded storage provides these two levels of paging performance.

One other thing to watch out for is that in an environment that pages to DASD, the potential exists for transactions to break up with the paging I/O. This could cause a real storage only configuration to look like the throughput rate is higher.

Minidisk Cache - Record Level

- MDC with track caching can result in worse performance for large database jobs.
- Prototype measured which provides record level MDC for CMS data.
- Development APAR VM61045 for 2.1.0 and 2.2.0.
- SET MDC MDISK option in APAR (and QUERY MDC)
- Directory options in future



When VM/ESA 1.2.2 enhanced minidisk cache, we knew there was potential for extreme workloads to run worse under the new MDC. Since then a number of customers have run into problems when running large databases that are built off of CMS flat files. "Large" is an important term. In these case, large means files that are 100s of cylinders in size.

Currently customers have found relief by increasing processor and or CU storage sizes, tuning the system, or tuning the application. We know this is not acceptable as a long term solution.

Some prototype measurements are shown of an enhancement that became APAR VM61045. (on RSU 9703 for 2.1.0 and 9706 for 2.2.0). This code adds a flavor of record-level MDC back into the system. It applies only to CMS data accessed through certain CMS I/O interfaces.

The measurements made included an application that modeled a customers database access pattern. You can see that not only is performance improved over the track MDC, but it is also better than the original record level MDC seen in the VM/ESA 1.2.1 measurement. A second workload was run that accessed a file sequentially. The results show that the track level MDC remains the better performer in this case.

Performance Futures

- The release to release performance has been good for the past several releases
- The regression improvement list is getting rather short
- In future releases, expect some more specific changes to address particular environments
  - Mitigate the extra page required for segment tables to address greater than 32MB

For the measured regression environments, VM/ESA performance has improved or remained equivalent for several releases now. However, the degree of improvement has been decreasing. The list of performance improvements for the regression environment has become fairly short. The new near term strategy is to look for performance enhancements that are particular to certain environments or that help address capacity issues. For example, we are currently exploring how to provide relief to the 32MB line. This is related to the fact that when storage addressability is extended above the 32MB, the segment table no longer fits inside the VMDBK and we must then use an entire page of storage for the segment table (non-pageable) or 2 pages for greater than 1GB.

If you know of other areas you'd like to see improved, please let us know. We are always looking for candidates. What gets into a release will depend on several factors, so we can not make promises. Thanks.

Summary

- VM/ESA 2.2.0 Performance
  - Potential for improvement for master constrained heavy VMCF scenarios
  - Potential for improvement in some CMS environments
  - Other regression environments equivalent
- VM Performance Information
  - VM Home Page
  - Performance Manual

There were a number of enhancements to VM/ESA 2.2.0 performance, some of which could provide noticeable improvement in performance, particularly in master processor constrained environments. Otherwise, the regression performance remains equivalent to VM/ESA 2.1.0.

For addition performance information, check out the VM Home Page at http://www.vm.ibm.com . The home page includes the complete VM/ESA 2.2.0 Performance Report and lots of other goodies. We made a few updates to the VM/ESA Performance manual, particularly in the area of MDC.

As always, I appreciate all you customers who work with VM because you are the reason I get to do what I do. Thanks.

References

1. VM/ESA 2.2.0 Performance SC24-5782.
2. "VM/ESA Data IN Memory Techniques" page, adapted from a paper by Kris Buelens and Guy De Ceulaer -- IBM Belgium. See http://www.vm.ibm.com/perf/tips/dim.html
3. VM/ESA 2.2.0 Planning and Administration SC24-5750.
4. VM/ESA 2.2.0 CP Programming Services SC24-5760.
5. ITSC Redbook "VM/ESA Storage Management with Tuning Guidelines", GG24-3934-01.
6. "Minidisk Cache Guidelines" page. See http://www.vm.ibm.com/perf/tips/prgmdcar.html