IBM

Washington Systems Center

Technical Bulletin

Analyzing Guest Machine Performance Under VM/370

By R. J. Knaus
Endicott Development

Published by P. W. VanLeer
Washington Systems Center

DAPS Code 0898
GG22-9026-00
July 1979

Analyzing Guest Machine Performance Under VM/370

DAPS Code 0898
GG22-9026-00
July 1979

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 customer's 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 environments do so at their own risk.

Requests for copies of this and other IBM publications should be made to your IBM representative or to the IBM branch office serving your locality.

A form is provided in the back for any comments or criticisms you may have. If the form has been removed, comments may be addressed to IBM Washington Systems Center, Field Support, 18100 Frederick Pike, Gaithersburg, MD 20760. Comments become the property of IBM.

PREFACE

This bulletin describes the performance analysis for a VM/370 System running a guest SCP (DOS or VS1), using VMAP and standard CP commands. Comments are offered on the effects of the various CP performance options. It was presented at Share 52 (March 1979) by Robert Knaus, Endicott VM/370 Development.

DAPS CODE 0898
GG22-9026-00
JULY, 1979

ANALYZING GUEST MACHINE PERFORMANCE UNDER VM/370

To Be Presented At

SHARE 52

ROBERT J. KNAUS
Department G22
IBM ENDICOTT, N.Y.

INTRODUCTION

I would like to discuss some of the performance considerations to be used when running guest SCP virtual machines, such as DOS or OS/VS1, under VM/370. This presentation will focus upon techniques that can be used to improve the performance of a "guest" machine in an environment where CMS users are also active. I will also try to tie the use of these techniques to information that is produced by the Monitor facility of VM and displayed in reports produced by VMAP.

See Foil 1

The presentation is related to one given by Donna Walker at SHARE 51 in Boston last summer. (Session B115 - Measuring VM/370). In it there is a flowchart describing performance analysis of a VM system. It begins with a decision block labeled "MPL" (multi-programming level), and follows a branch labeled "HIGH". This presentation concentrates on the other branch, LOW, single guest production machine.

Please remember that this presentation is one man's view of the world, namely mine, and that the information presented is derived from performance benchmarks run in a controlled laboratory environment.

ASSUMPTIONS

Before I begin, I want to point out some of the assumptions I will be making throughout the presentation.

See Foil 2

1. I will occasionally divide some of the remarks made into two categories, those that pertain to the base release of VM and those that pertain to the Basic Sytems Extensions Program Product (BSEPP). By the base release I will mean VM Release 5 and by BSEPP I will mean Release 1 of the program product. Most of the information presented will be true for Release 6 of the base and Release 2 of the program product. Where it is important, I will make the distinction. Additionally, most of the things that are true for BSEPP will also be true of the Systems Extensions Program Product (SEPP) although I personally have not run a SEPP system to obtain equivalent information.

2. I am also presuming that the user wishes to be able to tune the guest machine to obtain its maximum throughput given that a certain level of CMS usage (at least 5 users) must also be supported. Maximum

throughput may be defined in terms of Relative Batch Throughput (RBT), if you are looking at batch workloads, or maximum obtainable response time, if running TP applications. I will also assume that if both batch and TP are being run as guest machines, then TP response time is the more important.

3. The presentation references many of the reports that are available through the VM Performance/Monitor Analysis FDP (5798-CPX). If you are not familiar with this program, the manual SB21-2101 describes the reports available and their content.

4. The only recommendations I will make on improving the performance of guest machines will be tuning and setup options. It is possible to improve the performance of any system with added CPU power, memory, or direct access storage. However, for the sake of this presentation I am assuming a fixed set of resources.

5. I will also assume that the guest machine, be it batch or TP or both, would run reasonably well in a "native" environment (without VM) and that, where possible, the operating system has been generated with hand-shaking or linkage enhancement features.

EXPECTATION LEVEL OF PERFORMANCE

See Foil 3

Before getting started, let's look at what one might expect for performance if an OS/VS1 or DOS/VS workload is changed from a native operation to operation under VM/370. Given the same CPU and storage before and after this transformation, performance will depend chiefly on the level of VM assist supported by the CPU. Recognize that these are the upper limits of DOS and VS1 performance under VM/370. MVS performance under VM is not addressed in this presentation.

a. On machines with ECPS (138, 148, 4331, 4341)

Relative Batch Throughput	- .82 to .92	
Relative CPU Seconds	- 1.2 to 1.6	
TP Response Time Change	- 0 to 30%	(native RT < 5 seconds)
	- 50% or more	(native RT > 5 seconds)

b. On machines with VMA (158, 168, 303X)

Relative Batch Throughput - .75 to .85

Relative CPU Seconds	- 1.4 to 2.0	
TP Response Time Change	- 10 to 40%	(native RT < 5 seconds)
	- 50% or more	(native RT > 5 seconds)

c. On unassisted machines

Relative Batch Throughput	- approximately .5	
Relative CPU Seconds	- 3.5 to 4 or more	
TP Response Time Change	- 50% or more	(native RT < 5 seconds)
	- 100% or more	(native RT > 5 seconds)

NOTES:

1. ECPS is the microcode assist that includes virtual machine assist, extended virtual machine assist, virtual interval timer assist, and CP assist. The 4331 VM:ECPS does not include all of the CP Assists.

2. RBT is relative batch throughput - the elapsed time native divided by the elapsed time under VM.

3. Relative CPU Seconds is the result of dividing the product of elapsed time and CPU utilization under VM by the same product from the native environment. The concept is taken from a presentation by P. VanLeer at SHARE 51, session B158 - VM/370 Analysis Methodology - Guest Operating Systems.

THE BEGINNING

See Foil 4

Even before collecting VM Monitor data for reduction by VMAP, a number of facilities exist to begin looking at the performance of the system. The commands INDICATE LOAD, INDICATE QUEUES, INDICATE PAGING, and INDICATE I/O can be issued by the system programmer or operator. VM Release 6 supports the collection of monitor data to disk with the file being closed every 'n' intervals. This data can be reduced by VMAP as it is gathered. The VM/370 Graphic Monitor IUP (TVMON) can be used to look at performance data "on-line".

With these facilities one can check utilization, paging, storage contention and get a general idea of where bottlenecks may be occurring.

It will also be extremely helpful if you can establish your own "benchmark", in terms of native performance or in terms of how the guest machine(s) run under VM with no other users active. If you can, it will give you a yardstick to measure against when tuning.

Next, begin to gather VM Monitor data. The monitor classes to use are USER, SCHEDULE, PERFORM and DASTAP. The SEEKS and RESPONSE classes can also come in handy and I will mention SEEKS class later. RESPONSE class is useful to measure CMS response times though it may be necessary to write your own analysis program to look at the data, depending on whether you can get what you want from the VMAP reports.

The response time data produced in the VMAP User Response-Time Analysis Report can be used to check CMS Response Times. However, the classification of CMS commands into trivial, minor and major can be misleading. A trivial command on this report is recorded when a console output line is written before the Q1-drop record is found on the monitor tape. Some long running commands such as LOAD or COBOL write a line to the console before processing begins. This output is counted as a trivial command. The net effect can be an understated response time for the average command in the system. In our performance work we use a reduction program that calculates a response time in a way similar to VMAP, but without considering queue drop, as well as Total Time. Total Time is measured from the console input to the next console read sent to the terminal (the next thing after the CMS READY message). Analyzing CMS performance using Total Time can more easily show effects of VM tuning.

Another way to derive response time data when using BSEPP or SEPP, is from the Resource Manager variables printed by VMAP in the statistical summary. Q1SEC and Q2SEC, which are average seconds on Q1/E1 or Q2/E2 between drops correlates closely with response time even though it does not include "terminal time" (the time to transmit data to and from the terminal). With this you do not need to use the RESPONSE class of monitor.

See Foil 5

Upon getting your VMAP reports look at the summary pages (OUTSTAT LISTING) for

PCTCPUQ-percent of users waiting for CPU
PCTSTGQ-percent of users waiting for storage
PCTPAGEQ-percent of users waiting for page
PCTIOQ-percent of users waiting for I/O

This data, which has been summarized for the entire run is based on snapshots of user's VM block data (namely VMRSTAT and VMDSTAT). These variables are good bottleneck indicators. If any values of these four variables are greater than 10, they are worth investigating. If they are ALL less than 5 you have a system that is running very well and you are to be congratulated.

Some may wonder why not look at PAGEWAIT AND IOWAIT as an indication of a bottleneck. A system Pagewait condition is set if the sum of the working sets for in-queue users in page wait is greater than half of the avaliable pages. A system IOwait condition is set if the pagewait criteria is not met and ANY user is in I/O wait. If neither of the above conditions is met, and there is at least one user in queue, then a pagewait condition is set. Therefore the condition reported by monitor and printed by VMAP is somewhat sensitive to the mixture of users in queue.

See Foil 6

A high PCTCPUQ is not very common in guest machine environments. This percentage indicates that CMS users and, perhaps, the guest user are waiting for CP services, primarily priv op simulation and the start of I/O (both virtual Start I/O's and DIAGNOSE I/O's). Look at the USER RESOURCE SUMMARY REPORT of VMAP and see which users are in CPU wait. If it is obvious that the guest machine user is not normally in CPU wait and that the CMS users are, then you may have overtuned the guest machine and must "back off" one of your tuning parameters, probably Set Favor Percent.

See Foil 7

PCTSTGQ, if high, shows that a substantial number of users are on the eligible list, awaiting enough storage for their working sets. There are some tuning parameters that can be used which can either cause this problem or help it go away. This problem can occur when running large guest operating systems and may be caused by SET FAVOR, SET RESERVE, or LOCK commands. Unless this value gets very high (over 20), it is probably most noticeable in the CMS user's response time. Once again you can determine which users seem to be the ones in storage wait by looking at the USER RESOURCE SUMMARY REPORT of VMAP. If it is spread fairly evenly, or it is most noticeable for CMS users then the guest machine is dominating storage. If it is the guest machine, then reserving pages for it or using the Set Favored command should be considered. When trying to optimize guest performance, CMS users in storage wait is very likely to happen and should not be considered "bad" if their response time is not impacted severely.

See Foil 8

PCTPAGEQ, if higher than 10 or so, indicates that paging is causing a bottleneck. This will have a significant effect on performance. If you have this situation, and you use preferred paging, that is, some CP owned volumes have been specified as paging-only devices (see DMKSYS, SYSOWN macro), check the DASD TAPE REPORT for I/O balance across channels and devices. If you have very high rates to these devices or to the channels, your performance is suffering because of both paging and waiting for the page device. (Note that the definition of very high is somewhat subjective. The type of device used for paging is the determining factor. Experiences on 3330 drives on a 370/148 leads to me to a conclusion that 25 I/O's a second is the definition of high for it.)

If you have more system packs than page packs on line normally, and these packs are the same device type, then consider spreading paging across all of the system packs and change DMKSYS so that all relevant system packs are TEMP. It is also recommended that TEMP space be put on the center of the pack in one extent. When paging is spread across devices and centered on the packs, VM will "round robin" the paging and arm movement on the devices will be minimized by ordered seek queueing.

If this is not possible, then put paging on the lowest utilized device and keep the TEMP space near the part of the device with the heaviest usage.

If your paging problem is not obvious from the DASD TAPE REPORT then collect SEEKS class data with VM Monitor and reduce it to determine how the I/O's are spread by cylinder on the devices. Use this information to group the more heavily referenced extents together and put them near to the center cylinder of the device.

Another thing to remember is that the use of PAGEX ON for DOS or VS1 will "camouflage" a page wait condition as virtual PSW wait. Since VM will return control to the guest machine on a page fault, the guest, if it has nothing else to do, will load an enabled wait PSW. This no longer looks like page wait to the system even though that is what it really is.

See Foil 9

PCTIOQ indicates users waiting for their own I/O. The problem is similar to that of paging but is not as easy to solve. The DASD TAPE SUMMARY REPORT will show how many I/O's are being done to each DASD or Tape device. Given that you know how your guest machine's data sets are laid out on its packs, you can balance I/O on the system to achieve minimum contention for the guest. Further analysis of DASD seeks can be obtained with the use of Monitor SEEKS class data and the facilities of VMAP. In VM Release 6 and BSEPP Release 2 you can select SEEKS class data by device type and cut down on the overhead caused by the Monitor Calls commensurately. Severe I/O wait problems will probably take a bit of experimentation with different setups to determine optimum placement.

In guest SCP's where the Start I/O Fast instruction is implemented, such as, VS1 Release 7, MVS and the Airline Control Program, I/O wait can be hidden since control is returned to the SCP as soon as the channel program is translated but before the actual I/O is scheduled. This condition will look like virtual PSW wait when it occurs and the guest SCP has nothing further to execute.

SOME OTHER MONITOR VARIABLES TO REVIEW

PGBLPGS - number of pageable pages

RESPGS - number of reserved pages

SHRPGS - number of shared pages

See Foil 10

PGBLPGS should be checked to see if there are any large changes in the number of pages available during the monitoring period. If there is it is an indication of the system needing to extend FREE storage into the page pool which is quite inefficient. This phenomena is due to the number of CMS users logging on and off and the sizes of their virtual machines (including number of virtual devices, use of ECMODE, etc.) It may also be caused by a "sudden" need by CP for pages in which to build control blocks, load non-resident CP pages, or for I/O areas. If this occurs frequently, it is an indication that not enough FREE space has been defined in DMKSYS (SYSCOR macro) and it should be made larger.

To approximate the number of pageable pages in the system use this formula.

$$\#\ \text{Pageable Pages} = \frac{\text{Real Storage} - (R + 64K + T + (\#\ \text{of users} * 3K))}{4K}$$

where:

R - is the size of the VM resident nucleus in K bytes,
64K - is the FREE space needed for the system,
T - is the size of VM's Trace Table,
# of users - the average number of logged on users

BSEPP Release 2 has a new function that improves the efficiency of the use of extended pages for CP subpools. Subpools of doublewords for which CP has done a FRET are put on the subpool chain of DMKFRE, even though they were originally obtained from extended storage. This allows them to be reused without searching through the extended FREE pages. The subpool search is also done with CP assist on the 138, 148 and 4341. When any user logs off, in BSEPP Rel 2, the system will scan extended Free pages and attempt to return them to the page pool if they are empty. DMKFRENP, the address of which may be found in the nucleus map and displayed with the DCP command, points to two words of storage. The first is the count of times the system has extended free space into the page pool and the second is the number of these pages that have been returned to the page pool. The difference, therefore, is the current number of extended pages.

RESPGS - allows you to check the number of reserved pages you allocated to the batch user.

SHRPGS - while the shared pages benefit a number of users they are effectively removed from the pageable page pool.

THE USER RESOURCE UTILIZATION SUMMARY REPORT

See Foil 11

The User Resource Utilization Summary Report of VMAP also offers some information about the guest machine.

Find the report line for the user id that is running the guest SCP. Look at Relative % CPU utilization, TV Ratio and WKSET.

Relative CPU Utilization should be quite high for maximum throughput. The rationale for this statement is that the desired Relative CPU for a guest should be equal to or greater than the total CPU utilization that the guest would use if it were running native.

If relative CPU is not high enough, SET FAVOR and SET FAVOR % can be used to help raise it. I will discuss the use of the SET FAVOR command a little later.

If TV Ratio is high, the machine is using a lot of CP services (for PRIV OP translation, I/O and page faults). DOS and VS1 systems generated with handshaking have most of these "extra" overhead items already removed. Consideration should be given to eliminating unneeded privileged instructions in the guest machine, which are usually a source of this overhead.

Average and maximum working set size gives you information on the batch guest storage requirements. This data will come in handy if you elect to use the SET RESERVE command.

Another analysis tool in this area is the VMAP TRACE report which may be produced for the guest machine. This report shows elapsed and CPU times between queue drops along with the records showing when the user is made eligible, added to queue and dropped from queue. If RESPONSE class is enabled, the console output is shown allowing you to relate the operation of the guest machine to what has been seen on the console log. Paging and steal data is also printed. From the report one can see the queue and eligible activity of the machine which can be used to guide the selection of tuning parameters as well as help measure their effects.

At this point, we have looked at some of the more obvious indicators of performance but have said very little about using VM commands that change the performance characteristics of the system.

PERFORMANCE TUNING

The best time to begin tuning performance is when you have some of the more "fixable" performance bottlenecks solved. Most of the steps that come under the category of tuning will tend to increase the effect of current bottlenecks in the system. This is especially true when using BSEPP or SEPP where the resource manager is attempting to distribute resources fairly equally while also trying to keep paging overhead and storage contention from degrading system performance.

Let's look at some of the tuning commands.

See Foil 12

1. USER PRIORITY - This is probably the easiest and safest area to look at first. In the base release of VM, setting a "preferred" machine's priority to 1 with all other machine's having the default of 50 has a positive effect in improving the performance of the guest machine. However, in BSEPP and SEPP it's effect is much more noticeable. Using a high priority (such as 0 of 1) for a guest machine will not "penalize" it for using more than a fair share of the CPU or for causing paging overhead. At priority of 0 the guest machine will be allowed to use 64 times its "fair share". I usually prefer to go with extremes in guest machine priority, that is use a priority of 1 or 0, and leave the other users at the default of 64. It is also possible to lower the priority of other virtual machines in the system below their default of 64. However, this usually has a more drastic effect on interactive response time than is preferable, especially for CMS users.

   The listing of DMKSTP, at the label PRITBL, lists priority values and their resultant fair share allocation.

   It is also worth noting that in BSEPP and SEPP priority is overridden by use of the SET FAVOR Percentage command for the guest machine.

See Foil 13

2. SET FAVOR - Use of this command keeps the userid specified on the runlist, it will never be put on the eligible list. This is true for both the base release and BSEPP or SEPP. The favored user has a much better opportunity to keep its pages in storage and to keep from losing pages to CMS users. However, the system is more likely to show more storage contention (PCTSTGQ) and larger eligible lists. Depending on the level of CMS service you intend to provide, this may not be a negative factor.

   In BSEPP and SEPP the Set Favor command may be used for more than one virtual machine. In the base release it may not.

   In conjunction with a high priority setting (i.e. 0 or 1), use of SET FAVOR will usually provide the most dramatic increase in performance for a guest machine.

See Foil 14 and 15

3. SET FAVOR percent attempts to "guarantee" a user a certain amount of CPU during its time slice. Its implementation is different in the base release and BSEPP or SEPP.

   In the base release SET FAVOR percent implies SET FAVOR, whereas in BSEPP and SEPP it does not. Also, in the base release SET FAVOR percent can only be used for one virtual machine but in BSEPP and SEPP it can be user for more than one.

   In the base releases of VM, the use of a set favor percentage alters the position of the favored user in the run list. The userid will be kept at the top of the run list until it has gotten the percent of available CPU designated in the SET FAVOR percent command. After the percentage is obtained the userid will "drop" to its normal position in the run list (probably the bottom of the list). The "problem" with this can be the effect on other user's response time. A user's response time will be impacted by the position of the favored virtual machine on the runlist at any given point in time. However, if you are willing to live with this, it is a very effective tuning parameter.

   In trying to determine the percentage to use I would suggest either estimating the CPU utilization of the

guest running native and use it as the number, or starting fairly high and experimenting. By "high" I mean 50 or 60 and going all the way up to 99. A cross check on the effect of changing this percentage may be made by looking at the USER RESOURCE UTILIZATION SUMMARY REPORT, specifically the relative percent utilization of the guest machine. It is usually possible to get this utilization moving up to a point where you will see very little change no matter what you try to do. For those users who are bold, you can always start with 99 and work your way back down. SET FAVOR 99 will usually run a guest machine as well as it will ever run but can create eligible lists that are quite long if a guest machine with a large working set size is able to obtain a large portion of the available pages.

In BSEPP, the favor percent is used to adjust dispatching priority not absolute position on the run list. The favored user's deadline priority will be adjusted to its time slice plus one minus the favor percent times the time slice. For example, with Set Favor 99 and a 2 second Q2 time slice the user's deadline priority will be 2 seconds whereas without a favor percent, in a system that is running with an expansion factor of 5 for example, the deadline priority would have been 10 seconds.

As a result when using Set Favor Percent with BSEPP or SEPP interactive users will still be on top of the run list with the guest machine following close behind. (I am assuming that the guest machine is usually a Q2 or Q3 user.) Any bottleneck the CMS users are able to create (paging or I/O primarily) will affect the performance of the batch guest more than it would by using Set Favor Percent in the base release.

With BSEPP, I would suggest using SET FAVOR and SET FAVOR percent, with the percentage being quite high, that is, over 80.

BSEPP Release 2 has a new SET FAVOR userid 100 command that causes one virtual machine to be placed at the top of the runlist. The logic is similar to that in the base release of VM, with approximately the same effect.

See Foil 16

4. SET RESERVE is a command that will reserve a specified number of pages for a specific user id. It gives the

guest machine a page pool of its own and tends to keep CMS users stealing each other pages not the reserved pages. In trying to arrive at the number of pages to reserve, one should consult the USER RESOURCE UTILIZATION SUMMARY REPORT and look at average and maximum working set size for the guest machine. Also look at the number of pageable pages in the system (PGBLPGS in the Summary Report) which will certainly give you an upper limit. The working set size in the User Report is in K bytes while the pageable pages are stated in 4K pages, which can be embarassing if you forget it. If the working set of the guest machine is greater than half of the available pages and if the system has more of a paging bottleneck than a storage bottleneck, then the use of reserved pages can be useful if you watch out for a couple of things.

The number of pages to reserve on the first try should be equal to 75 percent of the average working set though not more than 67 percent of average available pages.

In the base release watch out for storage contention and eligible lists getting very large. If PCTSTGQ goes over 15 or 20 percent or the average eligible list size is greater than one-half the active users, back off on the reserved pages by at least 10 percent and try again.

I will mention a couple of warnings here. First, if you have reserved pages for a user then the average working set shown in the USER RESOURCE UTILIZATION SUMMARY REPORT will NOT show the reserved pages. As a result you will see a fairly small working set for your guest machine.

Second, in BSEPP or SEPP after PLC 7 the rules for qualifying for Q3 were changed slightly. Not only must the user have used six consecutive Q2 time slices but its working set must be at least 12% of available storage. It is possible with reserved pages to make a user appear to have a very small working set, which may keep him out of Q3. This is more important for batch guest machines and for guest machines that do very little I/O to terminals.

Lastly, if you have been using Set Favor and, perhaps, Set Favor Percent and have not experienced any "queue blocking" of CMS users, as seen by spikes in the number of users on the eligible lists, the additional use of Set Reserve may cause them. While the probability of this is not very high, it is worth watching out for. (In BSEPP, you can use the SET SRM

MAXWSS command to place a limit on the guest machine's working set size should you begin experiencing this problem. Try starting with MAXWSS equal to 90% of available storage).

5. LOCK - Besides locking page 0 which is always valid, it is difficult to determine what other pages could be locked that would not remain in storage due to the frequency of their reference.

See Foil 17

6. VIRTUAL=REAL OPERATION

   This requires using VM's virtual=real area to execute guest macines. If RBT's are compared between running VS1 or DOS with handshaking and running VS1 or DOS in V=R, V=R will be marginally better (this assumes no other virtual machines are active). In environments with CMS, handshaking often outperforms V=R. There are two points to consider. One is using the SEPP package when running V=R to get the shadow table bypass function. The other is using V=R when running large TP systems (e.g. CICS) under VM for maximum performance with minimum interference from CMS. Performance is still sensitive to the paging rate of the guest machine since it will be doing its own paging in V=R, and VM often does a better job of paging for the guest machine that the guest machine itself.

   Another interesting thing about Virtual=Real Operation under BSEPP is that the guest machine does not get the benefit of Q3 operation because of the working set restriction. That is, to be eligible to be placed on Q3 the machine must have used six consecutive Q2 time slices and have a working set that is 12 percent of available storage. Since the V=R guest machine's working set is considered to be 0 by the system, it does not make it into Q3.

DIVIDING WORK AMONG MULTIPLE GUEST MACHINES

See Foil 18

If the preferred guest contains both TP and batch partitions, and it is functionally possible to split the workload, then two virtual machines can offer some performance advantages. Some things to consider are:

1. The need to duplicate system packs unless you can guarantee that only one virtual machine will write to the pack. If this is true the pack can be shared via MDISK statements in the directory.

2. The possibility of creating a storage bottleneck with two copies of the SCP in storage (one for each guest) may outweigh any possible benefits. If a storage constraint already exists in the system with one guest then I would discourage the use of the second guest.

3. In the base release of VM you can favor one virtual machine, in BSEPP you can favor multiple machines.

4. Only one virtual machine can have reserved pages.

It is possible to get better performance from two virtual machines than from one if you can take advantage of tuning. By splitting TP and batch work into two guests, it is easier to indicate to the system which is to get the largest share of resource.

MISCELLANEOUS TUNING ITEMS

See Foil 19

SETTING PAGEX OFF OR ON

Setting PAGEX ON allows VM to return control to a guest machine when page faults occur. This will allow the guest to dispatch another partition while awaiting the page. The PAGEX capability is used with DOS and VS1 systems generated with handshaking.

Setting PAGEX OFF often improves the TP response times of a guest system that is running both TP and batch. In a multiple partition batch only guest, throughput will be dependent on how jobs are assigned to partitions. With PAGEX OFF the highest priority partition may get twice the throughput of a PAGEX ON system, but the lower priority partitions will receive commensurately less service.

RESOURCE MANAGER TUNING OPTIONS

When using BSEPP there are three additional tuning options available. These are interactive bias and paging bias which can be changed with the SET SRM command, and the SET PAGING variable, the ideal CPU per page read.

Interactive bias comes into play in adjusting the eligible list priority of a Q1 user that is not getting a fair share of the CPU. In a limited amount of experimentation, I have not seen this variable change performance by being changed from its default of 2 to 0, which would have the "worsened' the priority of some CMS users.

Paging bias is a dynamically calculated variable that weights a user's paging versus it's CPU utilization when calculating a queue priority. The paging bias has a default of 40. Raising it will "penalize" users that do more paging than average and lowering it will "penalize" users that use more CPU than average. VMAP, Version 3, now lists the Resource Manager variables that are used in Scheduler calculations. The current CPU-Paging bias (CURRBIAS) that the system is calculating periodically, and the limiting CPU-Paging bias (LIMBIAS - the default of 40) are shown in the RESOURCE MANAGER REPORTS of VMAP. CURRBIAS will only have a value when eligible lists exist.

While you will rarely see the value of CURRBIAS close to the value of LIMBIAS, raising LIMBIAS from its default will cause higher values of CURRBIAS. This will affect priorities of large working set users. If running the guest machine with SET FAVOR %, then this should only penalize large CMS users, which is not a bad idea.

The SET PAGING variable defines an "ideal" overhead for CP paging. If the system is exceeding the ideal then the scheduler attempts to lower the multiprogramming level of the system by inflating each user's working set size. If an eligible list exists, this has the effect of slowing down the rate at which users are added to the run list. By using the INDICATE LOAD command in BSEPP, you can check the current paging overhead. If it is consistently higher than 4% (its default) consider raising the value. The reason for this is that if the system begins increasing all working set sizes then it may do more harm than good to the guest machine, particularly if it is not FAVORED. Additionally, when the paging overhead tolerance is exceeded, users that are dropped from queue have their pages immediately put on the flushlist which may play havoc with CMS response times. This can be seen by inspecting the first byte of DMKPTRXX. If a bit is on in this byte the system is attempting to dampen paging overhead. In limited experimentation I have found that raising this value (e.g. from 4 to 8) has improved the RBT of some benchmarks and lowering it (from 4 to 1) has worsened the RBT. It is worthwhile to check the VMAP statistics which can pinpoint the actual effects of this tuning. I would suggest looking for changes in PAGERATE, E1, E2, CMS users working set sizes, and the batch user's Relative CPU %.

EVALUATING THE EFFECTS OF TUNING

In trying to evaluate the effects of using the Resource Manager tuning parameters, a popular method is to change a value, wait five or ten minutes for it to take effect, and collect monitor data for ten or fifteen minutes. Use VMAP to reduce the data and check for variables that seem to be influenced by the tuning. Doing this over time for different values of the same variable may allow you to spot a trend that you wish to take advantage of. I would say that, in general, you will not correct any severe problems through the use of resource manager tuning, but it may allow you to squeeze the last ounce of performance out of your system.

A PRIORITIZED TUNING CHECKLIST

See Foil 20

1. Generate guest machines with handshake or linkage enhancements.
2. Insure guest machine runs reasonably well "native".
3. LOCK page 0.
4. Decide whether PAGEX should be ON or OFF.
5. Run guest machine alone (if possible).
6. Check for bottlenecks. Look at USER RESOURCE UTILIZATION SUMMARY REPORT for Relative % CPU, total seconds and working set. Take steps to improve TV Ratio, if necessary.
7. Run "normal" system load (i.e. with CMS users, etc.)
8. Look for causes of Page Bottlenecks, if any, and fix, if possible.
9. Look for causes of I/O Bottlenecks, if any, and fix, if possible.
10. If Storage Bottleneck, determine which users in Storage Wait.
11. Check if adequate number of system FREE pages allocated.
12. Change guest machine priority (from default to 1 or 0).
13. Use Set Favor for guest machine.
14. Use Set Favor Percentage, based on Relative % CPU in step 6.
15. If guest machine in Page or Storage Wait a significant amount of time, use SET RESERVE.

IF STILL HAVING PERFORMANCE PROBLEMS ESPECIALLY WITH REGARDS TO POOR OR UNEVEN TP OR CMS RESPONSE TIMES:

16. Install BSEPP or SEPP.
17. Re-do steps 8 to 14.
18. Divide workload into multiple guest machines, if no storage constraint.
19. Experiment with SET PAGING based on Load Percentage in INDICATE command.
20. Evaluate other Resource Manager tuning options.
21. Cut back workload or look into more CPU, memory or DASD.

ANALYZING GUEST MACHINE PERFORMANCE

UNDER VM/370

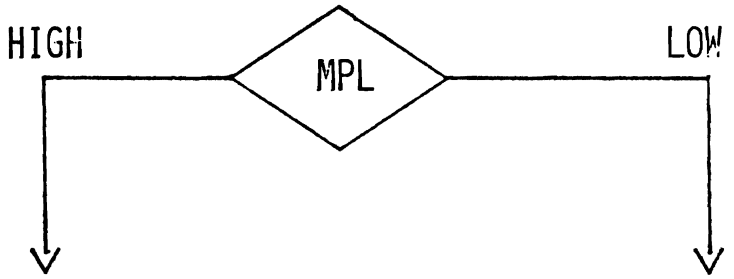
SHARE 52

MARCH 18-23,1979

R. J. KNAUS

IBM GLENDALE LAB

ENDICOTT, NY



SEE D. WALKER'S
PRESENTATION
SHARE 51
MEASURING VM/370

SINGLE PROD
VIRTUAL MACHINE
SHARE 52

FOIL 1

ASSUMPTIONS

1. LEVELS OF VM/370

   RELEASE 5

   BASE RELEASE

   RELEASE 6

   BASIC SYSTEMS EXTENSION PP

   RELEASE 1

   BSEPP

   RELEASE 2

   SYSTEMS EXTENSION PP - SEPP

2. WANT TO TUNE GUEST FOR MAXIMUM THROUGHPUT

   RBT
   TP RESPONSE TIME

3. VMAP FAMILIARITY

4. NO ADDED RESOURCES

   CPU, MEMORY, DASD

5. GUEST RUNS WELL NATIVE

   HANDSHAKING USED

EXPECTATION LEVEL OF PERFORMANCE

1. On Machines With Full ECPS (138, 148, 4331, 4341)

RELATIVE BATCH THROUGHPUT	- .82 to .92
RELATIVE CPU SECONDS	- 1.2 to 1.6
TP RESPONSE TIME CHANGE	- 0 to 30% (native RT < 5 seconds)
	- 50% or more (native RT > 5 seconds)

2. On Machines with VMA (158, 168, 303X)

RELATIVE BATCH THROUGHPUT	- .75 to .85
RELATIVE CPU SECONDS	- 1.4 to 2.0
TP RESPONSE TIME CHANGE	- 10 to 40% (native RT < 5 seconds)
	- 50% or more (native RT > 5 seconds)

3. On Unassisted Machines

RELATIVE BATCH THROUGHPUT	- APPROXIMATELY .5
RELATIVE CPU SECONDS	- 3.5 to 4 or more
TP RESPONSE TIME CHANGE	- 50% or more (native RT < 5 seconds)
	- 100% or more (native RT > 5 seconds)

LOOK AT PERFORMANCE

INDICATE COMMANDS

"REAL TIME" MONITOR DATA COLLECTION AND VMAP (VM REL6)

VM GRAPHIC MONITOR IUP

LOOK FOR PAGING, UTILIZATION, STORAGE CONTENTION, LOAD %

MONITOR DATA COLLECTION

MONITOR CLASSES

USER SCHEDULE PERFORM DASTAP

RESPONSE - USEFUL FOR CMS RESPONSE TIMES AND ANALYZING VMAP TRACE

MONITOR OVERHEAD

USING A MONITOR TAPE - 1-2% CPU AND 3-4 PAGES

WITH SEEKS CLASS - 4-5% CPU

BOTTLENECK INDICATORS

VMAP OUTSTAT LISTING - STATISTICAL SUMMARY REPORT

PCTCPUQ - PERCENT OF USERS WAITING FOR CPU
PCTSTGQ - PERCENT OF USERS WAITING FOR STORAGE
PCTPAGEQ - PERCENT OF USERS WAITING FOR PAGING

PCTIOQ - PERCENT OF USERS WAITING FOR I/O

PERCENTAGES GREATER THAN 10 ARE PROBLEMS THAT MAY BE ABLE TO BE SOLVED WITH TUNING.

SYSTEM PAGEWAIT

SUM OF WORKING SETS FOR IN-Q USERS GREATER THAN
ONE-HALF AVAILABLE STORAGE

SYSTEM IOWAIT

NOT PAGEWAIT AND ANY USER IN IOWAIT

OTHERWISE

IF ONE USER IN-Q, THEN PAGEWAIT

PCTCPUQ

NOT COMMON WITH GUEST MACHINES

CHECK USER RESOURCE UTILIZATION SUMMARY

REPORT TO SEE WHICH USERS IN CPU WAIT

"OVERTUNING" MAY RESULT IN CPU WAIT FOR CMS

PCTSTGQ

CHECK WHICH USERS

GUEST CAN TIE UP STORAGE BECAUSE OF

SET RESERVE

SET FAVOR

CMS USERS CAN MAKE IT TOUGH FOR GUEST - IF SO, SET RESERVE, SET FAVOR

PCTPAGEQ

AVOID PREFERRED PAGING

BALANCE I/O

TEMP SPACE NEAR MIDDLE, ONE EXTENT

PAGEX ON CAN CAMOUFLAGE

PCTIOQ

DASD TAPE SUMMARY REPORT

CHECK I/O DISTRIBUTION

USE SEEKS CLASS

SIOF CAN CAMOUFLAGE

PGBLPGS - # OF PAGEABLE PAGES

RESPGS - # OF RESERVED PAGES

SHRPGS - # OF SHARED PAGES

PAGEABLE PAGE CALCULATION

$$\text{PAGEABLE PAGES} = \frac{\text{REAL STORAGE} - (R + 64K + T + (\text{USERS} * 3K))}{4K}$$

R - SIZE OF RESIDENT NUCLEUS IN K BYTES

64K - FREE SPACE FOR SYSTEM

T - TRACE TABLE SIZE

USERS * 3K - 3K OF FREE SPACE FOR EACH LOGGED USER

BSEPP RELEASE 2 IMPROVES EFFICIENCY OF USE FOR EXTENDED PAGES

DMKFRENP - FIRST WORD IS # OF EXTENDS
SECOND WORD IS # OF RELEASES

USER RESOURCE UTILIZATION SUMMARY REPORT

FIND USER ID RUNNING GUEST MACHINE

LOOK AT RELATIVE % CPU, TV RATIO, CPU SECONDS, WORKING SET

LOOK AT RUNNING STATUS FOR EACH USER

VMAP TRACE REPORT

**** USER RESOURCE UTILIZATION SUMMARY REPORT **** PAGE

		<----------CPU---------->					<STORAGE		<------RUNNING STATUS----					
		RELATIVE				TCT:	WKSET		<PCT DELAYS DUE TO>					PCT
		CUM	<--SECONDS-->			VIRT	K BYTES		WAITING FOR				AP	VOL
RANK	USERID	PCT	PCT	TOTAL	VIRT	RATIO	AVG	MAX	CPU	STG	PAG	I/O	LOK	WAIT
1	U11	80	80	975	848	1.2	343	580	22	0	6	0	0	4
2	U4	4	85	52	38	1.4	54	200	2	0	8	4	0	78
3	U2	4	88	43	31	1.4	48	260	0	0	6	4	0	86
4	U6	3	91	36	18	2.1	82	200	2	4	4	8	0	82
5	U38	2	94	27	18	1.5	67	180	2	0	8	0	0	88
6	U37	1	95	16	8	1.8	46	150	2	0	6	0	0	92
7	U3	1	96	11	6	1.8	22	120	0	0	6	2	0	92
8	U7	1	96	9	3	2.5	31	180	2	0	4	0	0	94
9	U5	1	97	8	3	2.4	29	80	2	0	0	0	0	96
10	U10	1	98	8	1	8.4	22	40	2	2	4	0	0	92
11	U35	1	98	7	2	4.6	35	100	2	0	6	0	0	92
12	U36	1	99	7	2	4.6	39	120	2	0	4	0	0	94
13	U1	0	99	5	1	3.9	20	110	0	0	6	0	0	94
14	U9	0	100	4	1	5.0	15	140	0	0	4	0	0	96
15	U8	0	100	4	1	5.0	14	140	0	0	2	0	0	98
16	OPERATOR	0	100	0	0	0.0	9	62	0	0	0	0	0	100
	TOTALS:	100	100	1,213	981	1.2	55	586	3	0	5	1	0	86

FOIL 11A

SET PRIORITY

USER PRIORITY

BASE RELEASE 1 TO 99 - DEFAULT 50

BSEPP 0 TO 255 - DEFAULT 64

0 IS GOOD; 255 IS BAD

MAKE GUEST 0, LEAVE CMS 64

SET FAVOR

KEEP USER ON RUN LIST

MAYBE MORE STORAGE CONTENTION, LONGER ELIGIBLE LISTS

BSEPP/SEPP MORE THAN ONE USER

THIS COMMAND AND PRIORITY SHOULD GIVE GOOD PERFORMANCE FOR MOST ENVIRONMENTS

SET FAVOR PERCENT

"GUARANTEE" CPU TO GUEST

BASE RELEASE IMPLIES SET FAVOR: BSEPP DOES NOT

BASE RELEASE: ALTERS USER'S POSITION ON RUNLIST

BSEPP: ALTERS PRIORITY

ESTIMATING PERCENTAGE

1. NATIVE CPU UTILIZATION

2. CHECK USER RESOURCE RPT

IN BASE, START 50 OR 60

IN BSEPP, START HIGHER, GO TO 99

SET FAVOR 100 IN BSEPP RELEASE 2

SET RESERVE

RESERVES PAGES FOR GUEST

GET WORKING SET FROM USER RESOURCE UTILIZATION REPORT

USE 3/4 OF AVERAGE WORKING SET OR 2/3 OF STORAGE

KEEP AN EYE ON PCTSTGQ AND E1, E2

Q3 CRITERIA AFFECTED

SET SRM MAXWSS CAN SOLVE SOME PROBLEMS

LOCK

PAGE 0 IS NICE

VIRTUAL = REAL

HANDSHAKING OFTEN BETTER
NO Q3

DIVIDE ONE GUEST INTO TWO

CONSIDER

PACK DUPLICATION

STORAGE FOR 2 SCP's

MULTIPLE FAVORING WITH BSEPP

ONLY ONE RESERVED PAGE USER

EASIER TO TELL SYSTEM WHERE TO PUT RESOURCE

CAN BE A BETTER PERFORMER

MISCELLANEOUS TUNING

PAGEX - OFF BETTER FOR MIXED BATCH AND TP

RESOURCE MGR TUNING

- INTERACTIVE BIAS
- PAGE BIAS
- PAGING OVERHEAD
- DATA ON VMAP REPORTS

VARIABLE	AVERAGE	MINIMUM	MAXIMU	DESCRIPTION
DEADLINE	3.86	3.53	4.8	SYSTEM-WIDE TIME-SLICE DEADLINE (SEC)
ELTIME	0.05	0.00	0.1	SYSTEM-WIDE SECONDS IN ELIGIBLE LISTS
FAIRCPU	7.27	1.27	9.4	FAIR-SHARE OF CPU PER USER (SECS)
FAIRPAGES	88.25	88.00	89.0	FAIR-SHARE # PAGES PER USER
PROJCPU	12.78	5.09	24.7	PROJECTED CPU MS OVERHEAD/PAGE READ
AVHDCPU	7.34	6.80	7.7	ACTUAL CPU MS OVERHEAD/PAGE READ
CURRBIAS	2.21	1.00	5.0	CURRENT CPU-PAGING BIAS
LIMBIAS	40.00	40.00	40.0	MAXIMUM ALLOWED CPU-PAGING BIAS
INTBIAS	4.00	4.00	4.0	CURRENT INTERACTIVE BIAS SETTING
Q1TIME	13.36	7.01	23.8	SECONDS/MINUTE SPENT IN QUEUE 1
E1TIME	0.59	0.00	6.7	SECS/MIN SPENT IN Q1 ELIGIBLE LIST
Q1CPU	1.59	0.77	4.3	CPU SECONDS/MIN USED BY Q1 USERS
Q1PGSEC	596.65	123.81	1,421.1	Q1SEC * Q1PAGES (OCCUPANCY FACTOR)
Q1DROPS	0.23	0.19	0.3	DROPS PER SECOND FROM QUEUE 1
Q1SEC	1.00	0.49	1.8	AVG SECS IN Q1+E1 BETWEEN DROPS
Q1CPUUSE	0.11	0.05	0.2	AVG CPU MS USED PER Q1 CYCLE
Q1PGREADS	11.45	5.50	17.4	PAGE READS/SEC FOR QUEUE 1 USERS
Q1PGSTEALS	0.92	0.00	3.6	PAGES STOLEN/SEC BY QUEUE 1 USERS
Q2TIME	18.68	0.30	40.3	SECONDS/MINUTE SPENT IN QUEUE 2
E2TIME	0.93	0.00	10.1	SECS/MIN SPENT IN Q2 ELIGIBLE LIST
Q2CPU	10.26	5.20	14.8	CPU SECONDS/MIN USED BY Q2 USERS
Q2PGSEC	2,865.61	1.82	9,371.9	Q2SEC * Q2PAGES (OCCUPANCY FACTOR)
Q2DROPS	0.19	0.11	0.2	DROPS PER SECOND FROM QUEUE 2
Q2SEC	1.74	0.04	4.7	AVG SECS IN Q2+E2 BETWEEN DROPS
Q2CPUUSE	0.89	0.41	1.1	AVG CPU SEC USED PER Q2 CYCLE
Q2PGREADS	6.08	0.03	16.6	PAGE READS/SEC FOR QUEUE 2 USERS
Q2PGSTEALS	6.32	0.03	18.5	PAGES STOLEN/SEC BY QUEUE 2 USERS

RESOURCE MGR - VMQBLOCK VARIABLES

FOIL 19A

ANALYZING BATCH PERFORMANCE

A PRIORITIZED TUNING CHECKLIST

1. GENERATE GUEST MACHINES WITH HANDSHAKE OR LINKAGE ENHANCEMENTS.
2. INSURE GUEST MACHINE RUNS REASONABLY WELL "NATIVE".
3. LOCK PAGE 0.
4. DECIDE WHETHER PAGEX SHOULD BE ON OR OFF.
5. RUN GUEST MACHINE ALONE (IF POSSIBLE).
6. CHECK FOR BOTTLENECKS. LOOK AT USER RESOURCE UTILIZATION SUMMARY REPORT FOR RELATIVE % CPU AND TV RATIO. TAKE STEPS TO IMPROVE TV RATIO, IF NECESSARY.
7. RUN "NORMAL" SYSTEM LOAD (I.E. WITH CMS USERS, ETC.)
8. FIX PAGING BOTTLENECKS, IF ANY.
9. FIX I/O BOTTLENECIS, IF ANY.
10. IF STORAGE BOTTLENECK, DETERMINE WHICH USERS IN STORAGE WAIT.

11. Check if adequate number of system FREE pages allocated.

12. Change guest machine priority (from default to 0 or 1).

13. Use Set Favor for guest machine.

14. Use Set Favor Percentage, based on Relative % CPU in Step 6.

15. If guest machine in Page or Storage Wait a significant amount of time, use SET RESERVE.

IF STILL HAVING PERFORMANCE PROBLEMS ESPECIALLY WITH REGARDS TO POOR OR UNEVEN TP OR CMS RESPONSE TIMES:

16. Install BSEPP or SEPP.

17. Re-do steps 8 to 14.

18. Divide workload into multiple guest machines.

19. Experiment with SET PAGING based on Load Percentage in INDICATE command.

20. Evaluate other Resource Manager tuning options

21. Cut back workload or look into more CPU, memory or DASD.

ANALYZING GUEST MACHINE PERFORMANCE UNDER VM/370

READER'S COMMENTS

Title: Analyzing Guest Machine Performance Under VM/370
Washington Systems Center
Technical Bulletin GG22-9026-00

Please state your occupation: _____

Comments:

Please mail to: P. W. VANLEER
IBM Corporation
18100 Frederick Pike
Gaithersburg, Md. 20760