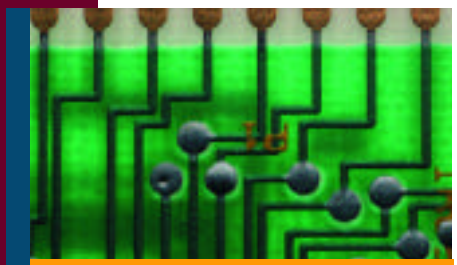

BY CHERYL WATSON

Vendor Performance Ratings: Why Your Results May Differ — Part IV

Your work may vary considerably from the workloads that were used by the vendor to determine the relative capacity and speed of a new model. It's up to you to determine how each of these factors will affect the real performance and capacity you receive.

PART I of this series (***Technical Support***, December 1997) presented my definition of various terms, examined the use of vendor claims for CPU performance, and described how vendors meet their claims. Part II (January 1998) examined IBM's *Large Systems Performance Reference (LSPR)*, Amdahl's and HDS's performance ratings, and why you should use these claims. Part III (February 1998) examined some of the factors that may cause your installation's experience to differ from the vendors' claims. This article will complete this discussion and conclude with some tips on what you can do when your results differ from your vendor's.

Why wouldn't you get the same performance out of a processor model for your workloads? In Part III, I addressed the first nine reasons that your experience may differ from your vendor's claims. This article examines the nine remaining items:

1. Workloads vary.
2. Your workloads don't match the vendor's.
3. You measure different things.
4. Your mix doesn't match the vendors.
5. The workloads vary throughout the day.
6. The volume affects capacity.
7. Constraints in software affect capacity.
8. Constraints in hardware affect capacity.
9. LPAR affects capacity.
10. Dispatch priorities affect capacity.
11. Software levels affect capacity.
12. Levels of PTFs affect capacity.
13. Different facilities invoked.
14. The amount of storage affects capacity.
15. The level of tuning affects capacity.
16. User's behavior changes.
17. The one thing that remains consistent is that you will always have change!
18. All of these reasons.

Dispatch Priorities Affect Capacity

Because you run a mix of workloads, the dispatch priority you have assigned to these workloads will be more important as you get closer to running your system at full capacity. For example, if batch is running at a low dispatch priority, as it is in most sites, the inconsistent CPU load from your higher priority work, such as CICS and TSO, will cause the batch work to receive sporadic, inconsistent access to the CPU. This causes an increase in CPU time that is normally not considered in the vendor's performance claims. That is, if you have all of your batch swapping in and out of storage and moving between multiple CPUs because it doesn't have enough priority to stay on one CPU, you will see increased CPU times in your batch workloads.

Software Levels Affect Capacity

The vendor's benchmarks are run on a level of MVS software that may or may not match yours. Until more installations are running the same level of OS/390, it's highly unlikely that all of your software levels will match the vendor's benchmarks. You need to consider not only the level and release of MVS, but also the release level of VTAM, JES, RACF, TSO, ISPF, CICS, DB2, IMS, and other key products in your installation. Of course, the levels and releases of your monitors, scheduling products, etc., should also be considered.

What this means is similar to the discussion in Part III on what to do if your workloads don't match the vendor's. An example of this is in ISPF. ISPF V4 took a lot more CPU time than ISPF V3. If the vendor is using ISPF V4 for the base and you are running ISPF V3, you will probably see a

difference in how the TSO workload is affected when moving between two models. That is, the vendor did not measure the impact of ISPF V3 – it could have been worse or it could have been better, but only you will know (it won't come out of the benchmarks).

Levels of PTFs Affect Capacity

Just like software levels and releases, the specific PTFs you have on your system will affect the capacity of the machine. As an example, the Catalog Address Space (CAS) takes a lot more CPU time in SP 5. If IBM's benchmarks use SP 5, their ITRRs include the impact on CAS when it's moved to another model. If you are still on SP 4, the CPU time for CAS will be trivial, and won't be affected by a change to a different model.

There have even been cases where IBM has had to apply some PTFs before running their LSPR tests due to some performance improvements that were related to the hardware. You may or may not have these same PTFs applied.

Different Facilities Invoked

The biggest problem that I see with current performance guarantees is that they consider older, traditional applications and not the newer applications. Since the current benchmarks are run on traditional workloads, how will you be able to determine the impact of a new processor model for your new applications such as IBM's Web Server on MVS, LANServer MVS, object technology with SOM and CORBA, Web applications such as Java, TCP/IP instead of VTAM, DB2 stored procedures, OpenEdition MVS, MQSeries, and similar new applications?

Likewise, consider the applications that are trying to take advantage of some of the facilities that were new as of SP 4 or 5 and still haven't been used, such as SmartBatch, DB2 Sort Assist, CICS storage protection, LPAR automatic recovery, etc.

One of the newest applications, parallel sysplex, is yet to be considered for the hardware benchmarks. In a parallel sysplex configuration, how much does the processor model affect the communication and overhead to and from the coupling facility?

Amount of Storage Affects Capacity

This is an old consideration that people often forget. If you have a large amount of storage available you receive two benefits. First, additional storage reduces the overhead of paging and swapping, which only steals cycles from the CPU. Second, applications that take advantage of storage alternatives such as using a large number of buffers will use less CPU time. The sort program is a good example of this. Incore sorts take less CPU time than DASD sorts, while hiperspace sorts can take more CPU time, but less elapsed time.

If you have a large amount of storage and use it, you will take the least amount of CPU time per transaction. If you are short on storage, you will end up taking more CPU cycles from productive work and spending them on paging activities.

You can do two things to ensure that you get your money's worth. You can obtain a performance guarantee from your vendor before deciding on a processor model, and you can measure (and understand) the relative change in capacity after you've moved to the new model.

The Level of Tuning Affects Capacity

The level of tuning makes a large difference in the effective capacity of a machine. An example of this is good blocking. Yes, you've probably heard for years that good blocksizes (half- or full-track blocks on disk) are the most efficient. While most installations have ensured that production datasets are well blocked, in most sites programmers still tend to use a factor of 10 to get blocksizes (80 x 800, 1600 x 16000) that produce very poorly performing jobs. Good blocking could reduce the CPU by 10 percent to 20 percent. If you have many of these in your batch workload, the programs aren't running very efficiently and may not be getting the maximum benefit out of the new processor. A well tuned system will always get the best performance out of a new configuration.

User's Behavior Changes

One of my favorite true stories is how we improved the system response time for a groups of users from 10 seconds to subsecond. However, within two days, the amount of CPU consumed by those users tripled. When I asked them why their CPU usage increased, they interrupted my question to show me a new trick. They said, "Boy, Cheryl, before that change you made things were really slow! If we wanted to look up your record, we'd have to type in your full name, 'Watson, Cheryl', then wait forever for a response. Now we just type in 'Wa' and start scrolling until your name comes up. It's super fast now!" For those of you that have experienced the crunch caused by a large amount of VSAM browsing in a CICS application from a lot of users, you'll understand how distressed I was. To take advantage of the improved response time, they started using a much less efficient technique that cost us quite a few CPU cycles.

Many sites have gotten burned because an improvement in response times caused users to change their behavior. Another common example is seen when TSO users find that the system is so fast they start doing all of their work in foreground rather than submitting batch jobs. This leads to excessively longer TSO last period response times and CPU consumption.

The One Thing That Remains Consistent is That You Will Always Have Change!

IBM is fortunate in that they can always provide a consistent, unvarying environment in which to run their benchmarks. They are able to obtain consistent results from one run to the next. However, this is seldom the environment that you can expect to see. The only consistency in most production sites is the inconsistency of the workloads. An entire day of processing can be harmed if a batch job from the nightly cycle abended and must be run during the day with the online workloads. TSO users may all come back from a meeting at the same time and hit the system with double the normal TSO load. For example, the CICS group could change a single parameter in their CICS parameters and increase the CICS CPU time by 5 percent, or the DB2 group could add some indexes and reduce DB2 time by 15 percent. Just be assured that you will seldom have two periods of time that are consistent in which to collect your measurements.

All of These Reasons

In many cases, some or all of the 17 documented reasons are interacting with each other at the same time. Very often, there isn't just one reason for a change, rather there are many at the same time. Measurement metrics may appear to report random numbers, frustrating the most senior level measurement expert. That interplay, in itself, may hide the real cause of underlying problems. It is sometimes more difficult to recognize the reasons for poor results than it is to fix the problem.

Your work may vary considerably from the workloads that were used by the vendor to determine the relative capacity and speed of a new model. It's up to you to determine how each of these factors will affect the real performance and capacity you receive.

WHAT CAN YOU DO?

You can do two things to ensure that you get your money's worth. You can obtain a performance guarantee from your vendor before deciding on a processor model, and you can measure (and understand) the relative change in capacity after you've moved to the new model.

Performance Guarantee

Each vendor can (optionally) provide a performance guarantee, but they will almost always qualify the performance as it applies to what they think your workloads will experience. Vendors have a lot of experience with a variety of workloads on their own models, but seldom publish this information. In my experience, they do a very good job of sizing when they know that a performance guarantee will be used.

Part of the performance guarantee is an agreement on the methodology that will be used to confirm that you receive the performance you expect. Generally, this consists of itemizing your important workloads and specifying their current performance with expected performance from the vendor.

Most performance guarantees require that the analysis be done between two environments where all changes have been frozen. That is, new workloads, changes in operating system, parameter changes, etc., are not allowed between the two during the period of analysis. Be sure that you can handle this period of time without any system or application changes.

Don't forget to check with your vendor about tools they might have available to help estimate the effect on your workloads. IBM has a marketing tool called SoftCap that is available for planning and analysis. Amdahl has a marketing tool called MP-CALC that also takes LPAR configurations into account. HDS has similar tools available.

Measure Your Own System

To understand whether a new processor model is meeting its expectations, you need to measure what you are actually experiencing. IBM provides one solution for this in "The Complete View" section of Chapter 5 of their *LSPR* manual. As an introduction to their solution, IBM states that "For a validation to work, there must be a commitment that the workload run on the new processor be the same as that on the old processor. In other words, there should be no shifting of workloads until after the validation is complete." Their technique is to use the logical EXCPs related to the total processor busy over a week of prime shift data.

I've tried this technique and have found that it only worked in a few cases because users could not make the commitment that the workload not change during the week directly after a processor upgrade. The workload will almost certainly change after a processor upgrade, and changes will be made by the data center personnel.

I've found much more success with the technique of identifying stable job steps and online transactions before the change was made and then seeing how these jobs and transactions were affected after the change. This technique was first introduced by Joseph B. Major. Though this technique doesn't take operating system or sub-system differences into account (such as the effect on JES or RACF), it will definitely show the effect on the application. If you don't have time to write your own programs to find these stable jobs and collect this data, take a look at our latest product, BoxScore. BoxScore identifies and quantifies the effect of any change, such as tuning, Year 2000 conversions, processor upgrades, etc., on stable job steps and transactions. This software, as well as this article series, is based on research that I've been doing in this area for the past 10 years.

CONCLUSION

The performance estimates for new processor models from IBM, Amdahl, and HDS provide valuable data to help you understand how much capacity you can expect to see if you move to those new models. This is especially true of the IBM LSPR ratings by workload. We hope to see workload performance ratings from the other vendors at some point in the future.

Your workloads may not see exactly the same effects because of several factors, including the facts that your workloads don't mimic the vendor's and most installations run in an LPAR environment, which may not be considered in performance claims.

To ensure that the vendor will help you if you don't get the performance you expect, I recommend that you have the vendor provide a performance guarantee before delivery of a new model.

You should define a technique to identify the relative effect of any processor model change based on your own workloads, not on estimates from artificial workloads. Remember that the vendor's claims are almost always provided for the optimum environment – one running with no constraints in a well-tuned non-LPAR environment. If you are running in an LPAR, have any constraints, or are not well-tuned, you can't expect to achieve the same performance results. ts

NaSPA member Cheryl Watson is an internationally-known consultant and author who has been working with IBM mainframes since 1965. She is the president of Watson & Walker, Inc., and is the principal author of Cheryl Watson's TUNING Letter, a journal of OS/390 advice published since 1991. She will soon be releasing BoxScore, the first in her ADVISOR Series of low-cost performance analysis software. Visit her web page at www.watsonwalker.com.