CAPACITY PLANNING USING CAPTURE/MVS™, BEST/1™ AND SAS

Robert Ashton, Jeffrey Buzen and Leonard Lipner

BGS Systems, Inc., Waltham, Mass.

In this paper, we present a method for capacity planning which illustrates the mutually complementary roles of CAPTURE/MVS, BEST/1 and SAS. The first two of these are software products developed by BGS Systems, Inc.; they provide the ability to extract and analyze MVS measurement data, and to project the performance impacts of system or workload changes. In this study, SAS is used in conjunction with output from the BEST/1 iteration facility to generate graphs that depict the performance characteristics of several proposed upgrade strategies.

The capacity planning methodology in this paper has been used in a large number of studies, carried out by BGS Systems consultants and by our clients. It has received widespread approval, and is routinely considered both credible and relevant in the minds of senior management. The example shown is a stylized version of a specific study which employed this methodology.

The focus of this illustration, from the banking industry, is an IBM 3033 8 megabyte system, which supports a variety of online and batch workloads under MVS/SP1. During prime shift the principal work comes from TSO and IMS, with some background batch work using slack capacity when it exists. At peak times on a normal day there are 30 TSO terminals logged in, almost all of them used for program development. IMS consists of 8 message regions, which collectively support both inquiry and update transactions. Strict response time objectives have been set for the three online workloads (TSO, IMS Inquiry, IMS Update), and none have been set for daytime batch. Standards do exist for overnight batch; however experience has shown that a system capable of meeting daytime objectives is usually adequate to handle the overnight work. This simply means that when growth occurs, the principal questions (for this client) deal with ensuring that there is enough capacity to handle the daytime work.

Relatively rapid growth is in fact expected to occur. Projections call for approximately 5 additional TSO terminals per quarter over the next two years due, interestingly enough, to the requirement to keep programmer turnover low. The intent is to move closer to the industry standard of 1-2 programmers per terminal, but this also implies the need for additional mainframe capacity to keep response time at a level acceptable to programmers. At the same time, in the IMS area, vigorous advertising and introduction of new banking services are expected to produce transaction volume increases amounting to 25% each calendar quarter, again extended over a 2-year planning horizon. The questions of course center on ways to manage this growth. To make this more specific, management has requested answers to several specific hardware strategies: adding 4 megabytes of memory, paging on 3380's, and paging on 3350's attached to 3880-11 disk controllers with cache buffers. To state it simply, the questions revolve around when some upgrade will be needed, how well each upgrade will perform, and for how long.

As in any planning study, the starting point is a thorough understanding of the current (or baseline) environment. To accomplish this, several peak hour intervals were chosen, and measurement data from each was collected (RMF, SMF, the IPS in effect at the time, and various IMS measurements). CAPTURE/MVS analyzed this data, together with built-in information regarding the structure and flow of control of MVS, to produce a BEST/1 model of each interval. Successive refinements over the past several years have enabled CAPTURE/MVS to generate BEST/1 models that are as correct (or valid) as they can be, given the quality of the measurement data; however, in some cases, adjustments to the model must be made to reflect shortcomings in the measurements. Model validity is checked by having BEST/1 perform its queueing analysis on the model input, and then comparing its outputs (i.e. calculated response times, throughputs, queue lengths) against their measured counterparts. To the extent that calculated and measured performance match, the model will be considered valid. In contrast, where there is a mismatch, one must re-study the measurement data (or lack of it) to account for the discrepancy.

CAPTURE/MVS plays a principal role in finding such flaws in measurement data, by means of the reports it produces as a by-product of model generation. For example, CAPTURE/MVS can indicate when EXCP's constitute only a small fraction of the I/O's to a disk, requiring further analysis of other I/O measurements, or when, because of missing termination records, one workload's EXCP's are uncharacteristically low. Another example, which appears from time to time, is unusually high unaccounted CPU time (such as spin locks for the dispatcher queue) for which no specifie SMF measurements exist. These phenomena and others like them, often affect model validation in a significant way. In addition to its principal role of preparing BEST/1 models, CAPTURE/MVS has a number of special facilities that help to detect and/or correct the model in order to reflect these measurement problems.

The product of our analysis was the selection of a truly representative interval, and a valid BEST/1 model of that interval. The next step consisted of using the model to determine when the current system configuration would no longer meet the installation's response time objective. Prior to runnung the model repeatedly in this way, a BEST/1 VARIABLES file was allocated , to which BEST/1 writes a single record for each iteration of the model. The VARIABLES which the user may select for inclusion in these records can be any performance information calculated by BEST/1 or any value input to BEST/1 as part of the model. In the present illustration we have chosen to write the three major components of each workload's response time, namely memory queue time, CPU residency time, and total residency time for TSO, IMS Inquiry and IMS Update. (SAS calculates I/O residency from the selected variables.) This data is written to the VARIABLES file in fixed length format, with spaces for delimiters, periods representing omitted values, and names for variables in standard SAS format. The BEST/1 SAS command also writes a small SAS program to read the data in the VARIABLES file. We used PROC GCHART to create a BLOCK diagram for each alternative strategy considered in this study.

At this point, we had determined the point at which an upgrade seemed to be required, and the BLOCK chart illustrating this became one of several direct inserts in our report to management. As is normally done in our studies, several alternative strategies were then identified for avoiding performance degradation, and "played out" by appropriate modification of the model. Not all of these will be described here, but a few sample ones will illustrate the range of alternatives considered, and the general method for studying them:

- Memory Upgrades: Several cases must be investigated since additional memory can be used in several ways. At one extreme, multiprogramming levels (MPL's) can be increased, assuming that paging will stay about the same. At the other end, we may assume that the extra memory is used entirely to reduce paging. Between these extremes are a large number of intermediate steps - using some of the memory to increase MPL's (but how much and for whom?) and using the rest to reduce paging. Several such analyses were carried out; in each case, reflection within BEST/1 of the scenario considered was accomplished in a few commands.
- 3350 Paging with 3880-11 Cached Controller: Here again, the reflection in BEST/1 is immediate, given the baseline description of the I/O subsystem. The effects of the 3380-11 fall into a few catagories: lower effective page access time, reduction in actual demand for the 3350 paging devices, and reduction of the number of RPS reconnects per page I/O, since those which are satisfied out of the CACHE require no reconnects (and hence no reconnect "misses") at all. All these effects are based on the cache hit ratio, and for this value we turned to published IBM studies, since at that time no other information was available.

Once again, BEST/1 command files were used to automatically study the client's anticipated growth in the context of these upgrades, and SAS/GRAPH used with the BEST/1 VARIABLES file to show the results in a form readily accessible to management. We were able to identify in detail the workloads whose performance was most likely to degrade at a certain time, and the reasons for this degradation (i.e., where the "bottlenecks" would appear). Finally, we noted that the workload growth projections were themselves open to question, and the BEST/1 - SAS/GRAPH combination was used to perform and present sensitivity analyses in each case. These consisted of identifying in each case the high-risk areas, that is, the "good" performance which could become "bad" (unacceptable) if the actual workloads turned out to be even a little larger than forecasted. And once again, management found this to be of particular value. In fact, it is our experience that sensitivity analyses, properly based and graphically presented, are one of the main factors in establishing the credibility of this approach to capacity planning.

BIBLIOGRAPHY

1. BEST/1™ User's Guide, BGS Systems, Inc., BE78-020-1. 1 University Office Park, Waltham, MA 02254, February, 1982.

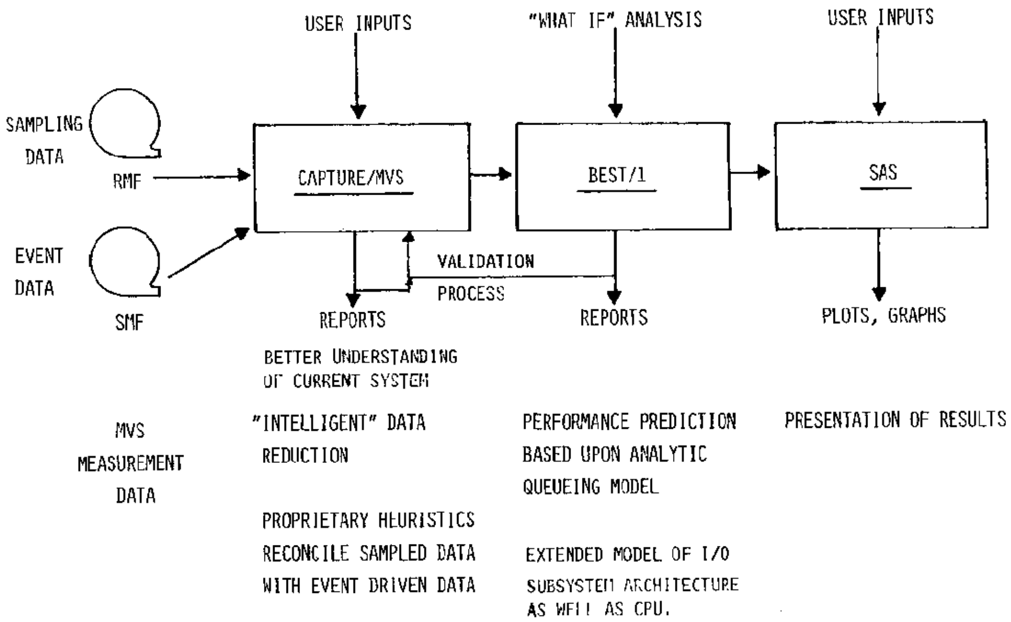
2. Buzen, J.P., et al., "BEST/1-Design of a Tool for Computer System Capacity Planning," Proceedings of the NCC, Vol. 47, AFIPS Press, Montvale, NJ, pp. 447-455.

3. Buzen, J.P., "Operational Analysis: The Key to the New Generation of Performance Prediction Tools," Proceedings of the IEEE COMPCON, 1976, New York.

4. Buzen, J.P., "A Queueing Network Model of MVS," Computing Surveys 10.3, September 1978, pp. 319-332.

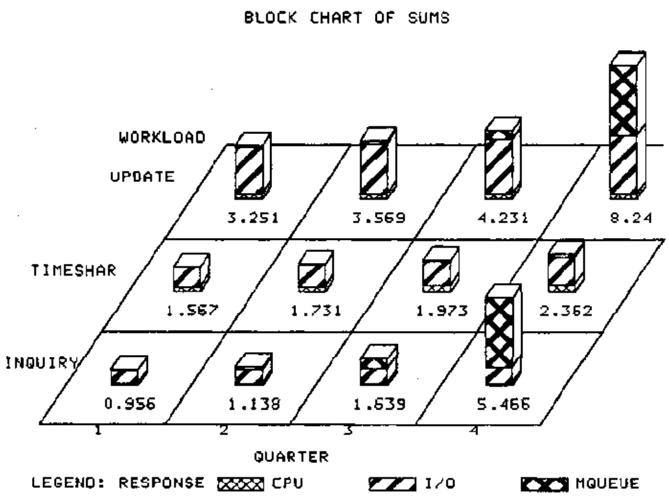
5. Lipner, L.D., "DP Capacity Planning, Ensuring Service to End Users," American Bankers Association - Operation Automation Division, 1979.

6. CAPTURE/MVS™ User's Guide, BGS Systems, Inc. CM81-030-2, 1 University Office Park, Waltham, MA 02254, February, 1982.



ONLINE RESPONSE TIMES

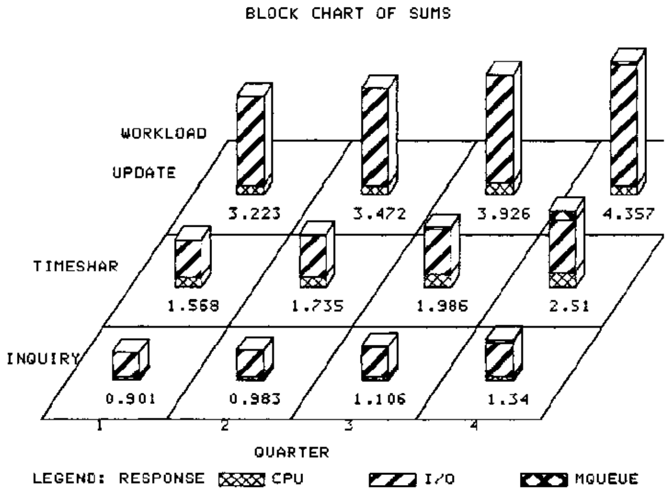
1982 BY QUARTER



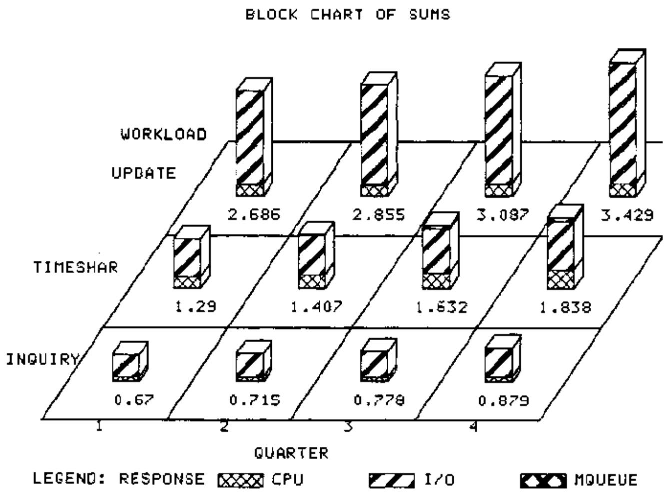
ORIGINAL SYSTEM

ONLINE RESPONSE TIMES

1982 BY QUARTER



2 MEG MEMORY UPGRADE



3880 MOD-11