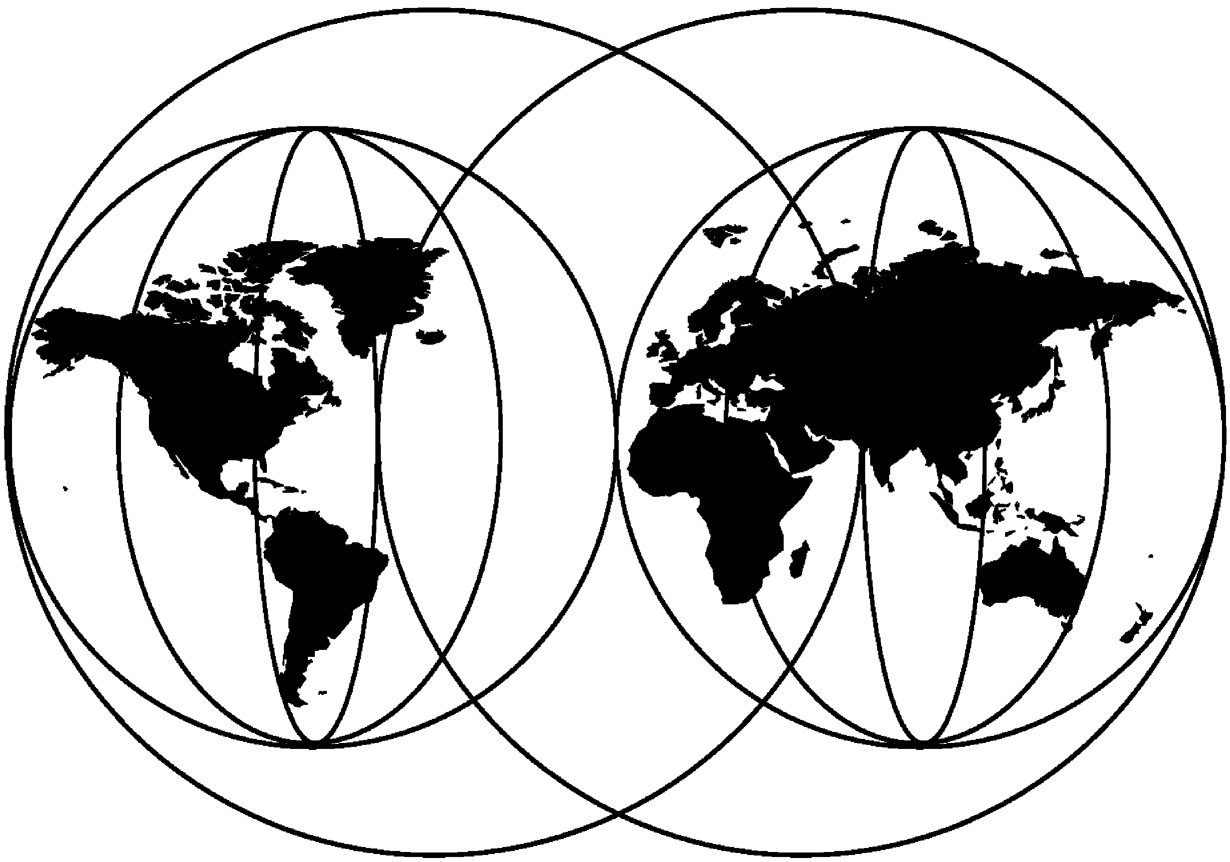
IBM

AS/400 Server Capacity Planning

Gottfried Schimunek   Lloyd Perera
Alan Wallet   Michael Liu Hong Chuan



International Technical Support Organization

http://www.redbooks.ibm.com

This book was printed at 240 dpi (dots per inch). The final production redbook with the RED cover will be printed at 1200 dpi and will provide superior graphics resolution. Please see "How to Get ITSO Redbooks" at the back of this book for ordering instructions.

SG24-2159-00

International Technical Support Organization

SG24-2159-00

AS/400 Server Capacity Planning

February 1998

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix B, “Special Notices” on page 173.

First Edition (February 1998)

This edition applies to Version 4, Release 1 of the Performance Tools, Program Number 5769-PT1 for use with the OS/400

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. JLU Building 107-2
3605 Highway 52N
Rochester, Minnesota 55901-7829

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© **Copyright International Business Machines Corporation 1998. All rights reserved.**
Note to U.S. Government Users — Documentation related to restricted rights — Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract with IBM Corp.

Contents

© Copyright IBM Corp. 1998

Preface

This redbook can help you with capacity planning for server applications running on AS/400 server models. Today there are a wide variety of client/server applications for the AS/400 system on the market place. Unfortunately, there is no easy way to do capacity planning as we used to do with pure "green screen" interactive applications and the BEST/1 feature of AS/400 Performance Tools.

This redbook describes the methodology and the approach to do capacity planning for all server applications and provides examples for some selected server applications such as:

- APPC banking application widely used in China
- Socket based Lawson Software application suite
- TCP/IP based SAP R/3
- HTTP serving with Internet Connection Server

The methodology described here applies equally to applications from companies such as IBS, Intentia, JBA, JD Edwards, Lawson, Marcam, Peoplesoft, SAP, SAS, Software AG, SSA, just to name a few. But also applications based on Lotus Domino, JAVA, ODBC, TCP/IP Sockets, and so on.

How This Book is Organized

Chapter 1 contains an introduction to capacity management and describes briefly the total capacity management approach of which capacity planning is a part of.

Chapter 2 introduces capacity planning specifically on the AS/400 system and gives a short overview of the capabilities of Best/1.

Chapter 3 shows the specifics of capacity planning for server applications. It also discusses the characteristics of AS/400 server models and how to define a unit of measure.

Chapter 4 describes all necessary steps to use BEST/1 for server capacity planning, from measuring and collecting performance data, to selecting time intervals and generating a model over to calibrating the model and extrapolating the measured data.

Chapter 5 is the first example of a server application based on Sockets. All clients are basically thin clients and most of the application is actually running on the AS/400 system. It shows how the volume test has been set up and executed, and shows the differences between real measured data and extrapolated data with the model.

Chapter 6 is another server application based on APPC. All clients are basically connected through a UNIX system to the AS/400 system and the application is executed on the AS/400 system.

Chapter 7 discovers an elegant way of using Internet serving log data and how to use it as a base for capacity planning of HTTP serving. The log data is used not only for capacity planning, but also for getting an insight to what HTML pages users are requesting most and how often.

© Copyright IBM Corp. 1998

Chapter 8 shows the approach to collect and extrapolate performance data with SAP R/3 on AS/400 and allows you to normalize to SAP R/3 Dialog Steps from a variety of different applications within R/3.

The Team That Wrote This Redbook

This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization Rochester Center.

Gottfried Schimunek is a certified I/T Architect at the International Technical Support Organization, Rochester Center. He writes extensively and teaches IBM classes worldwide on all areas of application development and performance. Before joining the ITSO in 1997, Gottfried Schimunek worked in the Technical Support Center in Germany as a Consultant for IBM Business Partners and Customers as well as for the IBM sales force.

Michael Liu Hong Chuan is an AS/400 advisory sales specialist in South China region. He joined IBM China in 1993 and started to support major finance and security accounts in South China. Michael has a lot of practical experience in AS/400 performance management and system management. After graduating from the Electrical Engineering Department of the University of Electronics Science and Technology, XiAn, China, Michael has been working in the computer area for over 12 years.

Lloyd Perera is an Advisory Systems Specialist in IBM Australia. He has 22 years of experience in the Information Processing fields. He has worked at IBM for 20 years. His areas of expertise include AS/400 performance tuning and capacity planning. He has coauthored ITSO redbooks on Performance Management, Client/Server Performance, and SAP R/3 Implementation for AS/400. He is currently on assignment at the AS/400 Technology Solutions Center, Rochester.

Alan Wallet is a Senior Systems Management Consultant in IBM U.K. He has 19 years of experience gained in various IBM field positions supporting midrange customers; prior to that he worked in the defense industry as an electrical engineer for 6 years. He holds a degree in Pure and Applied Mathematics, a post graduate Diploma in Management Studies, and has obtained the ITIL Service Management qualification. His areas of expertise include Systems Management, Capacity and Performance Management, and AS/400 Volume Testing. He has presented to the British Computer Society on the benefits and approaches to "AS/400 Volume Testing".

Thanks to the following people for their invaluable contributions to this project:

Jim Cook and Neil Willis
International Technical Support Organization, Rochester Center

Dan Frissora
IBM Rochester

R. Mike Smith
IBM Rochester

Comments Welcome

Your comments are important to us!

We want our redbooks to be as helpful as possible. Please send us your comments about this or other redbooks in one of the following ways:

- Fax the evaluation form found in "ITSO Redbook Evaluation" on page 185 to the fax number shown on the form.
- Use the electronic evaluation form found on the Redbooks Web sites:

For Internet users	`http://www.redbooks.ibm.com`
For IBM Intranet users	`http://w3.itso.ibm.com`

- Send us a note at the following address:

  `redbook@vnet.ibm.com`

Chapter 1. Introduction to Capacity Management

The IT industry is going through yet another period of great change, particularly in the AS/400 world. A number of major innovations such as client/server computing, Web based applications, and entirely new programming approaches (in particular, object-based languages such as Java) are making their presence felt.

1.1 Changing Times

The subject of client/server computing has been topical in the AS/400 world for a number of years now. Many AS/400 application vendors have made use of "screen scraping" techniques to give their products a "graphical look and feel", but more recently, several vendors are beginning to ship totally re-engineered AS/400 applications. These re-engineered applications have a stronger claim to being "client/server" than their "screen scraping" predecessors. As such, they deliver more of the benefits that attract businesses to client/server applications, but they bring with them some new and difficult problems. One of these problems is how to size the workload produced by these applications, and how to manage the capacity requirements as the number of users exploiting the new technology grows. It is this set of problems that we address in this publication, but first it is appropriate to consider some of the realities of the client/server environment.

1.2 The Realities of the Client/Server Environment

The client/server world is quite different from the world of "green screens" and structured procedurally-based applications. Some of these differences are discussed in the following list:

- End users are no longer constrained by procedural code. As a result, how any given user chooses to use an application is even more difficult to predict. This, in turn, means that any two users of the same client/server application can have significantly different computer resource requirements to carry out similar amounts of work when measured from the business perspective.
- Many client/server transactions can initiate hidden activity that is difficult to predict response times for. An example is a Domino user launching a transaction from their PC that, as part of its processing on the local server, caused the local system to initiate a Web based enquiry on a remote server. The resultant response time is dependant not only on the utilization levels of the resources of the local server, but is also impacted by such factors as the bandwidth available on the intranet (or Internet) and the utilization levels on the remote server.
- End users demand fast response times. Many end users have a performance expectation based on their experience with PC applications running locally on their own workstation. This experience leads them to expect almost immediate response from any computer system - and to be intolerant of systems that do not meet this expectation. This expectation of almost immediate response is also a key feature of the well-developed and growing demand from businesses to service their customers "directly" (that is, on the telephone with the assistance of responsive and comprehensive

© Copyright IBM Corp. 1998

information systems). Such users must have almost immediate responses from their IT services if they are to successfully deliver the fast efficient service that their customers expect.

- The client/server needs to resolve the skill issue. As with any new technology, most of the application vendors are still building their skills in this area. Most of them are still in the process of building a knowledge base and experience on how to deliver client/server applications that perform well. Most of their effort thus far has been devoted to delivering stable code that meets a functional specification. The net of this skill issue is that there are a number of client/server applications that have heavy resource demands, and the performance tools are still being refined so they can accommodate the client/server workloads.

1.3 An Approach to Client/Server Capacity Management

It is worth highlighting that there is nothing unique to the AS/400 system in the preceding description of the client/server environment. The points can usually be applied to the client/server world in general. A review of this portrayal of the client/server environment prompts the conclusion that "the IT capacity requirements for any client/server workload can be predicted with even less precision than for a legacy workload where a workload can be considered to be a particular application being subjected to a particular rate of usage".

This publication intends to set out a tactical solution to this problem. The solution is simple but is dependent on all parties involved in a client/server situation being realistic, acknowledging the difficulties previously described, and working together to address them.

The tactical approach that we recommend is to:

1. Develop a capacity estimate:

   Use the approach described in later chapters to develop a "capacity estimate" where a capacity estimate, as its name implies, is understood by all parties not to be accurate.

2. Tune the system according to the recommendations found in the AS/400 Work Management Guide SC41-5306.

3. Implement performance monitoring:

   Put in place structured and disciplined system performance monitoring so that the "estimate" can be adjusted in the light of data describing the real workload. If this is done in a methodical fashion, hopefully corrective action can be taken before the service to the business is adversely impacted. It has to be acknowledged that this may require the provision of additional server hardware resources, but this is not always the case as client/server performance problems can be addressed by several approaches including additional bandwidth, more powerful client systems, adjustments to middleware, or tuning of the client/server application code.

   The implementation of performance monitoring may seem so obvious that it is unnecessary to state it here. The effort required is not significant, the tools required are small, and the tools available with the AS/400 system make the automation of this process straight forward. To illustrate how simple this can be, sample documentation to support such a process is included in Appendix A, “Working Paper on AS/400 Performance Management” on page 167.

But what is meant by this widely used term "capacity management"? In the next section, this is discussed and a definition and a breakdown of the process is put forward. If you feel comfortable with this aspect of the subject, you can skip the next section and move to Section 1.5, “Applying the Approach to AS/400 Servers” on page 6.

1.4 A Definition of Capacity Management

The definition used in this publication is as follows: *capacity management is the IT business process that aims to ensure efficient management of IT resources to meet the needs of the business.*

1.4.1 Business Processes

As is often the case, the preceding definition depends on yet another term (in this case, the concept is that of the "business process"). This, is turn, can be defined as:

A business process is the organization of people, equipment, energy, procedures, and material into work activities needed to produce a specified end result (work product). It is a sequence of repeatable activities that have measurable inputs, value added activities, and measurable outputs.

Business Process Management practitioners usually advocate that a successful business process based on the preceding definition should have the five key elements in the following list:

1. A designated owner of the process:

   In the case of capacity management, the designated process owner clearly depends on the size of the installation, but obvious candidates are (going from small to large) the DP Manager, the Operations Manager, and the Capacity Planning Manager. It may even be a process that an installation believes it has outsourced. This is a dangerous perception and we recommend that the capacity planning "process owner" should still be within the company. With appropriate reporting, the responsibility can be assigned to a non-IT professional such as the chief accountant.

2. Inputs:

   Ideally, these should be measurable, and a good example in the case of capacity management is "user requirements". If these are to be factored into the process, they must first be documented and expressed in measurable terms. For example, it is impossible to successfully carry out capacity management if the user requirements are expressed as "deliver acceptable response times". Acceptable to who? At what time of day? And what is meant by acceptable? This line of reasoning clearly emphasizes that the capacity management process cannot stand alone but needs other processes such as the requirement highlighted previously for a "service level management" process.

   Other "inputs" for a capacity management process are budgetary information, business growth information, application development plans, information on hardware developments, and information on any completely new applications.

3. Outputs:

   Again, these need to be measurable. Examples of typical outputs from a capacity management process are:

   - Reports describing performance delivered
   - A capacity plan
   - Requirements for funds to finance any additional purchases of communications services, software, or hardware

4. Documentation:

   This should describe the process so that it is repeatable, and everyone involved should have access to the the document to promote a shared vision of what the process is and what it is trying to achieve. It is particularly important that such documentation be as brief as possible.

5. Quality Metric:

   The process should have a measurable attribute that allows an easy assessment of how successful the process is in meeting its objective so that quality improvement actions can be put in place if required. In the case of capacity management, the most frequently used metric is "interactive response time". However, in a properly engineered capacity management process, this needs to be qualified by specifying how it is to be measured and when. In the case of client/server applications on the AS/400 system, there is, of course, a particular problem with the Performance Tools reporting in this area, but this is discussed in subsequent chapters and an approach is described.

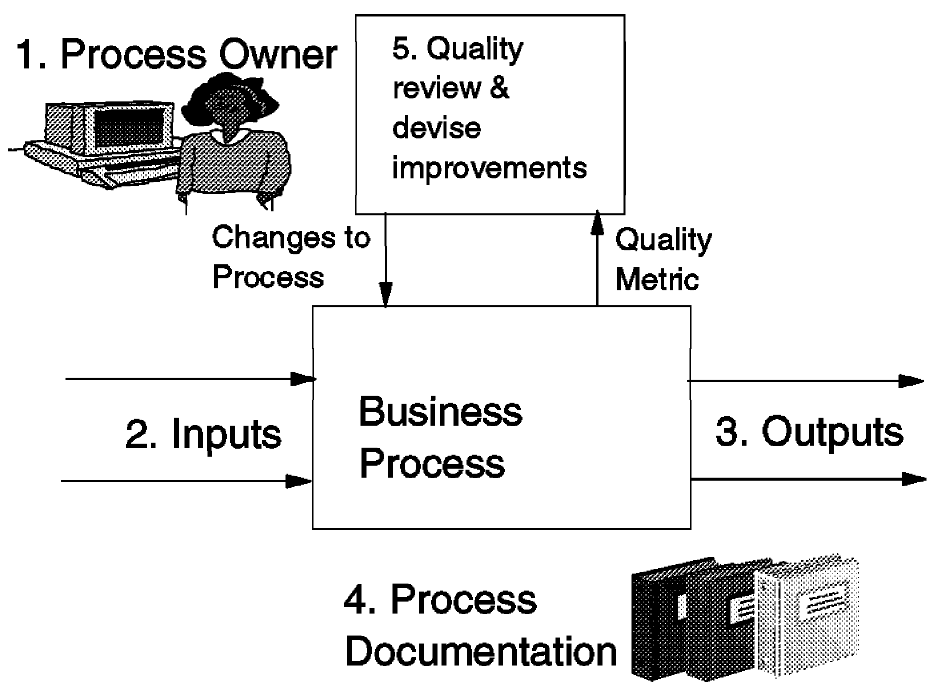


Figure 1. A Business Process

It is hoped that the preceding illustrations of the definition have encouraged the reader to persist with this application of business process management to capacity management. The application can be facilitated by breaking capacity management down into a number of sub-processes, and this is covered in the next section.

1.4.2 Sub-Processes of Capacity Management

It is possible to subdivide capacity management into the following sub-processes:

1. **Workload Management**: Understanding the applications and the workload profile.
2. **Performance Management**: Monitoring and tuning in accordance with a service level agreement.
3. **Capacity Planning**: Forecasting hardware requirements. It includes estimating the requirements of the individual workloads, application sizing, and then assembling them into an overall capacity plan. This is, of course, probably the key output of the overall capacity management process.
4. **Demand Management**: The management of users' demands, both batch and online, for IT resources.

How these sub-processes fit together and some of the key information flows between the components of capacity management are illustrated in the following diagram:

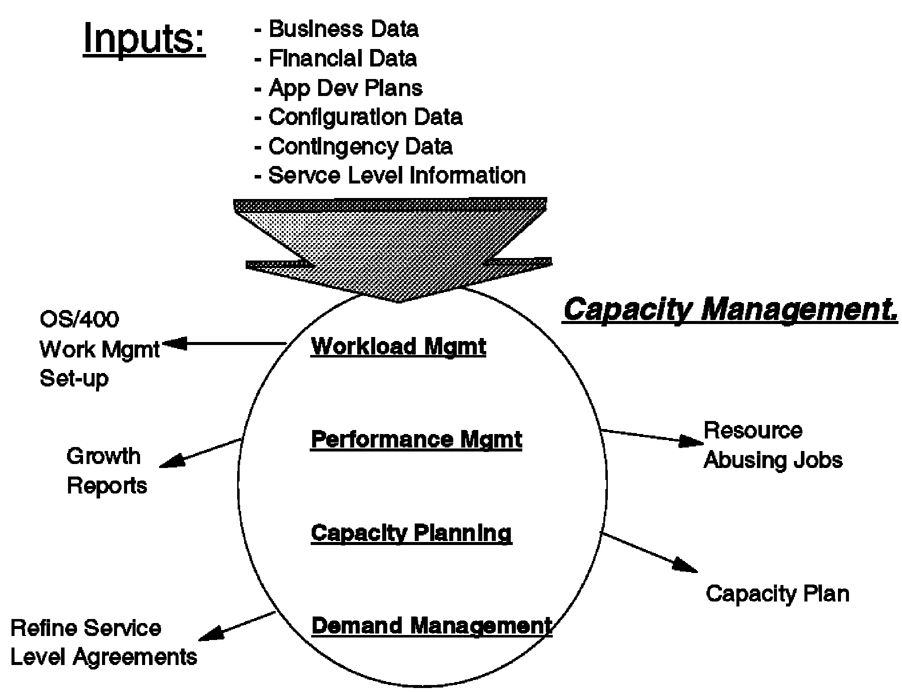


Figure 2. Capacity Management Sub-Processes

Having identified these sub-processes, which may seem a rather theoretical exercise and inappropriate in a redbook, the next section intends to add the practical dimension by discussing the sub-processes in the specific context of AS/400 servers.

1.5 Applying the Approach to AS/400 Servers

Applying an overall approach to capacity management to AS/400 servers can be done in the following ways for each of the sub-processes:

1.5.1 Workload Management

Workload management requires that a good understanding of the workload is developed. This entails understanding both the **make up of the workload** and the **rate of working** of the respective streams of work that make up the total workload.

1.5.1.1 Source of Workload

The best way to do this is to ask the question, "what are the sources of the work that our AS/400 server is required to carry out?" Typical sources of work are:

- High priority interactive work for critical users such as "tele-sales workers". In AS/400 work management terms, we probably assign these users a priority of 19 to differentiate them from the normal interactive users who are running at the more normal priority of 20.
- Even higher priority high availability software that may be assigned a priority of "say 18" in an effort to ensure that its jobs are keeping a "shadow" database on a backup server as current as possible.
- High priority batch work such as that associated with the production of time critical paperwork. An example is the batch job that printed picking slips in a fast moving distribution business. In this instance, the business may demand that a picking slip was actually printed within a specified number of minutes of the originating order being entered into the AS/400 system. To achieve this, the priority of the high priority batch jobs is possibly set at 30 to ensure that these jobs acquired AS/400 resource ahead of the normal batch work, which typically runs with a priority of 50.

These different categories of work are ordinarily set up to run in separate AS/400 sub-systems with these sub-systems assigning their jobs to appropriately sized main storage pools. This process can be thought of as a "work management" design; it ensures that the AS/400 work management function treats the different elements of work in accordance with the genuine priorities of the **business** rather than in accordance with the IBM standard defaults.

1.5.1.2 Rate of Working

Determining the "rate of work" is often referred to as determining the "workload profile". The AS/400 Performance Tools Graphics facilities provide an easy way to develop this picture. This, of course, assumes that the Performance Tools have been used to build up some performance history, and for the purposes of determining the "workload profile", at least a month of data is required. Once a month's worth of data has been gathered, the daily graphs can be produced of:

- CPU utilization versus time by job priority modified to show the priority bands that are relevant to the installation.
- Interactive response time versus time if there are "green screen" applications on the system.

With these graphs, the following questions can be answered:

- **What is the busiest hour of the working day?**

Typically, the working day means between 08:00 and 18:00, but this can vary. The busiest hour is the hour when the amount of high priority CPU is greatest, and where high priority CPU is taken to be the total CPU required to run all the time-critical work. Clearly this includes all work required to run the interactive and system jobs but can also include high availability work and work required for time critical batch work, which is referred to previously when the production of time critical picking slips was mentioned. In the following graph, 11:00 to 12:00 is the critical hour despite the fact that the total CPU is higher between 16:00 to 17:00. A workload profile with these characteristics results from a situation where the interactive workload is at its peak in the morning due to customers phoning in their orders and making inquiries on deliveries. In the afternoon, this activity subsides but the batch work builds up because more and more batch activity is processed as the mornings orders are processed into picking slips and manufacturing orders, and route planning for the next day is carried out.

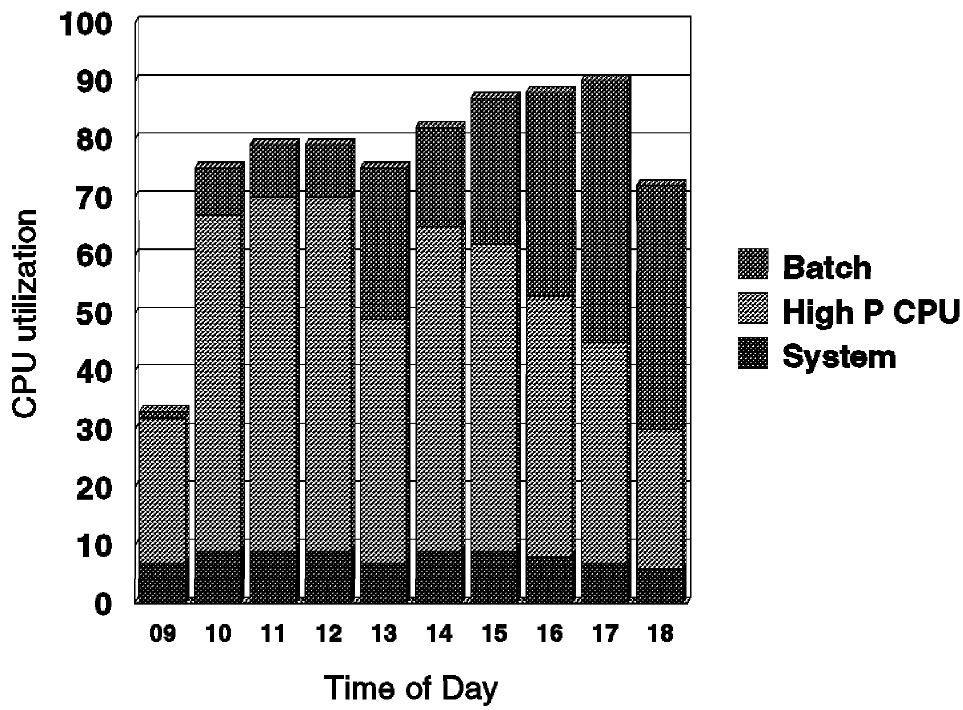


Figure 3. CPU Utilization by Time of Day

- **What is the busiest day of the week?**

  By simply reviewing the actual graphs, this piece of information emerges. Most installations know which day of the week is usually their busiest, so it is a good idea to check the analytical findings with the employees in the installation.

- **What is the busiest time of the month?**

  With only one month of collected data, it is not possible to determine this analytically. However, reference to the employees in the installation under study can easily provide this information.

- **What is the Business Metric?** In discussions with the installation, it is always useful to identify what "business metric" their organization quantifies its level of activity by. It may be "sales volume" (measured financially or in units of output) for a manufacturing company; for a distribution company, it is usually "customer order lines" or "number of picking notes"; for a hospital, it may be "patient admissions"; and for financial organizations, it is usually "number of customer transactions". Whatever the metric is, it is also important to consider it as a "rate" (that is, number done in a unit of time).

Organizations are usually aware of their busiest month in the year and their busiest week in the month (this is usually linked to "financial month end " activity). However, they may not be aware of their busiest hour in the day or their busiest day in the week for their computer systems because these can be affected by factors not visible to non-IT professionals such as a large amount of batch work that has been time shifted from the external events that caused it in the first place. This is why it is important to go through the exercise of producing and reviewing the CPU utilization graphs as previously recommended.

In summary, it is important to identify the following pieces of information for a month of collected performance data:

– The busiest hour in the working day
– The busiest day in the week
– How the month when the data was collected relates to the busiest month of the year in terms of a business metric rate that is in common usage within the installation

So if, for example, in a customer installation:

– The busiest hour of the busiest day in the month has a level of activity of 15 000 transactions per hour, which corresponds to a level of high priority CPU utilization of 60% determined from reviewing the performance data graphs.

  AND

– The month of the performance data collection results in 120 000 order lines, but the busiest month in the year is expected to result in 160 000 order lines.

  THEN

– An estimate of the "peak" level of activity in the "peak" hour of the year is 20 000 transactions per hour, and further work can be done to estimate the system requirements of this level of activity.

1.5.2 Performance Management

Performance management can be further sub-divided into three main areas as far as AS/400 servers are concerned and these are:

- Performance monitoring
- Performance reporting
- System tuning

Let's look at these in more detail:

1. **Performance monitoring**:

   In "workload management", which was previously described, performance data collection was set up to collect performance data for each day in a working month. As part of **performance monitoring**, which is much more of an on-going activity, it is now possible to set up a more refined performance data collection. For example, the performance data collection can be refined to run only for the busiest hour in the day. However, we recommend that you leave the performance data collection running for the entire prime shift unless disk space and the processor are extremely heavily utilized. To avoid consuming too much disk space with performance data, we suggest that you run CRTHSTDTA every two weeks from "Performance Graphics" to extract "Historic Data". This facility creates the following two summary files:

- QAPGHSTD
- QAPGHSTI

These are small files and, as a result, the space required for roughly a month's worth of data is probably in the area of 0.5MB to 2.0MB. This is a small price to pay to build up the vital picture of how the level of activity on the system is changing over time.

Once the extract has been run, the members that are more than two weeks old can be deleted from the system.

Collecting the data is one thing but this data needs to be processed to provide useful information to allow the server to be properly managed. To do this, the appropriate data should be extracted from the data by running Performance Tools reports such as the System Report (PRTSYSRPT) and then compared with the demands of the business, which ideally is documented in the Service Level Agreement (SLA).

The essential task of performance monitoring is to report on the system performance in such a way that "at a glance", the process owner can see if any of the four key elements of system resource are approaching their threshold values. The threshold values are published in *AS/400 Performance Management V3R6/V3R7*, SG24-4735-00, and in the *Performance Tools/400* manual. It is usually accepted that there are four main aspects that need to be monitored from a performance point of view for any computer system. These are:

- CPU resource
- Memory
- Disk arms
- Communication utilization

Hence, to monitor an AS/400 server, a simple table of the type shown in the following figure can be used:

Peak Hour	Wk 15	Wk 16	Wk 17	Wk 18	Threshold
High Priority CPU Utilization	62.5%	64.3%	61.6%		70%
Total Faulting - all pools	67.9	89.7	75.9		200
Disk Arm Utilisation & Space Utilisation.	32.4% & 67.9%	33.6% & 69.0%	31.2% & 71.2%		50% & 85%
Highest Comms IO Utilization	5.6%	5.4%	5.5%		35%

Figure 4. Performance Monitoring Table

From this simple table, the process owner can easily see if any of the four key aspects of system performance are approaching their "threshold values". Of equal importance, the process owner can also assess how quickly the value is approaching the critical level. The preceding table suggests that the

"CPU capacity" will become the first "critical" factor, but there is not enough data in the table to estimate when this will happen, which makes the point that it is important to run the performance data collection to build up the historic basis for sensible forecasts to be developed.

The information for such a table can easily be obtained from the AS/400 Performance Tool's System Report. In fact, it is suggested that slightly more data is extracted from the System Report and a spread sheet of the type shown in Appendix A should be constructed. This type of spread sheet is not needed by the the capacity management process owner, but it should be easily available to the process owner if required.

Performance monitoring in addition to routinely and regularly reporting on performance as previously shown to review the performance of the key aspects of the system should also monitor for the existence of "rogue jobs". A "rogue job" is one that uses large amounts of system resource; a classic example is a "looping job". These jobs are sometimes referred to as "performance hogs"; programs have been written to monitor for these jobs and such programs are sometimes referred to as "hog hunters". It is a responsibility of performance monitoring to detect such jobs. However, once notified, it is the "problem management" process that should assume responsibility for eliminating the occurrences. If the application code in use in an installation is stable, this is unlikely to be a concern; however, in installations where there is a significant amount of ongoing development, it is prudent to put an automated facility in place to detect "rogue jobs".

2. **Performance reporting**:

   A vital aspect of capacity management is to report on how well the IT systems are meeting the needs of the business. Ideally, this entails reporting achievement against a set of published criteria in a Service Level Agreement (SLA). In the AS/400 world, SLAs have not been extensively used, but as AS/400 installations become larger and organizations become more demanding of their IT functions, it is an unwise IT department that does not have some form of SLA!

   An SLA should be a small and simple document to encourage the staff to read it and ensure that it is understood. An SLA for a medium sized AS/400 installation can be written such that it fits on two typed pages (that is, both sides of one sheet of paper).

   In Appendix A, "Working Paper on AS/400 Performance Management" on page 167, there is a spread sheet that can be used to provide not only the information required for "performance monitoring" (see the preceding table), but also the performance information that is required to provide simple SLA reporting.

3. **System tuning**:

   The objective of system tuning is to ensure that the server is operating as efficiently as possible. This can be done manually by reviewing the tuning guidelines in the AS/400 manuals, or "Performance Advisor" can be used. Performance Advisor is part of Performance Tools/400 and reviews collected performance data to produce a report that highlights any anomalies that it identifies, and it even makes recommendations as to how the anomaly can be resolved.

1.5.3 Capacity Planning

The capacity planning sub-process is the major topic that is addressed in this publication. A brief summary is, therefore, all that is covered in this chapter. Capacity planning is the process of estimating the system requirements based on forecasts of growth of existing workloads and the addition of any new workloads.

This estimating can be done in one of two ways:

- **Manually**: This involves reviewing historic data, obtaining forecasts of business growth, and making a judgement based on simple mathematics to predict future resource requirements. This approach is described in Chapter 2, “AS/400 Capacity Planning” on page 13.
- **Using a Modelling Tool**: This is similar to the manual approach in that it is based on collected performance data and information on anticipated growth, but instead of simple mathematics, it requires the use of the Best/1 modelling tool to invoke the mathematical queuing theory to predict future system requirements. This approach for AS/400 servers is the subject of Chapter 3, “The AS/400 System in a Client/Server Environment” on page 23.

1.5.4 Demand Management

This may sound rather old fashioned, especially in the era when the user is meant to be "king" and the IT department is meant to give its customers (that is, the business users) everything they want. However, experience has shown that often if users are simply asked to avoid running a large query at a particular time of day, they are only too happy to oblige (especially if the IT department offers to automate the running of that query by adding it to an overnight batch run). Clearly, this is a simplistic example, but hopefully it illustrates the point.

Another aspect of demand management with regard to AS/400 servers is to consider "offloading" work to other servers to avoid the file conflicts that can arise if users are using the same data libraries for their online processing as for their queries and other report generation jobs. It may be possible to migrate the "decision support" workload to a dedicated server. Clearly, some form of overnight data replication needs to be put in place to keep the decision support data libraries up to date.

1.6 Interaction with Other IT Management Processes

A review of the capacity management sub-processes should begin to make it clear that capacity management has to be but one component in an IT organization's drive to deliver service to its customers. How capacity management should interact with a number of the other key service delivery business processes is illustrated in the following diagram:

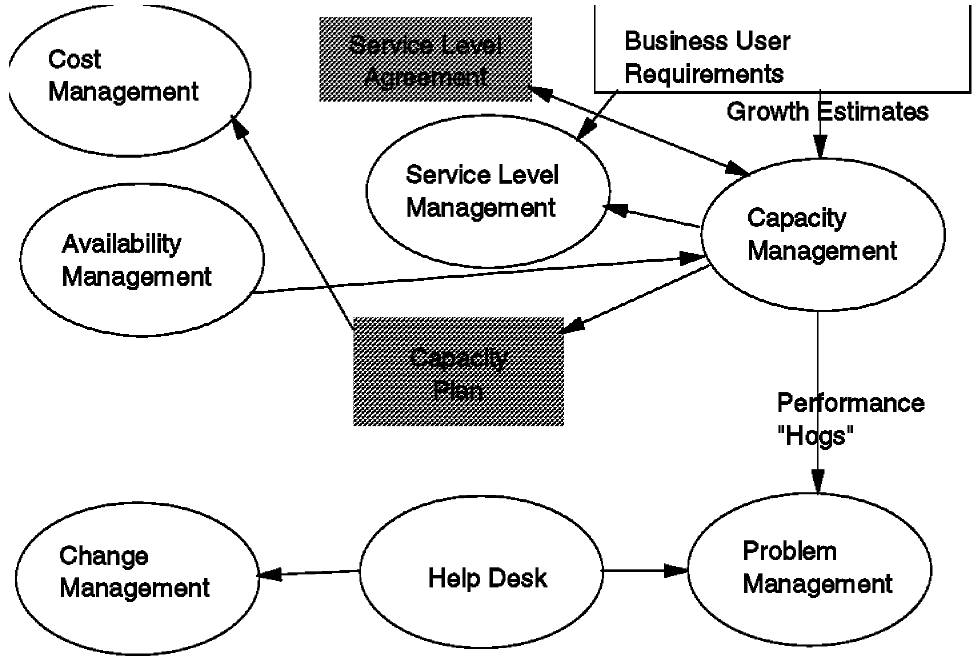


Figure 5. Capacity Management's Interaction with Other Service Delivery Processes

1.7 Summary

In this chapter, we have tried to set capacity planning in the overall context of capacity management and illustrate some of the interactions that exist between capacity management and other IT service delivery processes. We have also attempted to describe a simple approach to capacity management on AS/400 servers and illustrate this approach with sample documentation that is included in Appendix A, “Working Paper on AS/400 Performance Management” on page 167. We have also tried to explain that successful capacity planning requires more than simply running a modelling tool such as Best/1. In the rest of this publication,we put forth two approaches to capacity planning for AS/400 servers, but we emphasize that these approaches must be accompanied by on-going performance monitoring, especially because of the particular challenges of a client/server.

Chapter 2. AS/400 Capacity Planning

In Chapter 1, "Introduction to Capacity Management" on page 1, we considered the entire capacity management process and explained that capacity planning is one of the four sub-processes of capacity management. In this chapter, we look at capacity planning in more detail and put forward a fairly straightforward approach to capacity planning for AS/400 Systems in general. In subsequent chapters, we provide information about capacity planning for AS/400 **server** models, but these more specialized techniques are based on the material in this chapter. If you are familiar with the basics of capacity planning, it may be appropriate to quickly review the section headings in this chapter and move on to Chapter 3, "The AS/400 System in a Client/Server Environment" on page 23.

2.1 Introduction

The generalized approach covered in this chapter is simple, but it relies heavily on having good performance monitoring in place. Essentially, it consists of using historic data on the utilization of key aspects of the system to use straightforward mathematics and graphing to predict when a particular aspect of the system's performance will exceed the relevant guideline figure. The approach does *not* use the sophisticated mathematical modelling techniques that are found in tools such as BEST/1 and, consequently, can only generate rough estimates, which may not be applicable in all situations. However, the poor quality of the growth information that is usually available for capacity planning means that this lack of precision in the method is not often a major issue. In fact, it can be a benefit because it emphasizes the approximate nature of the output from the capacity planning exercise.

This chapter discusses the objectives of capacity planning, considers the assumptions that capacity planning exercises should make, and then describes the steps that have to be gone through to carry out a capacity planning exercise.

2.2 Objective of Capacity Planning and Associated Terminology

The objective of any capacity planning is usually accepted as being: *to develop an estimate of the system resource required to deliver performance in accordance with a documented Service Level Agreement (SLA) and based on a forecast level of business activity at some specified date in the future.*

Another term that is often used in this area is *capacity sizing* or *application sizing*. This document works on the basis that *capacity sizing* is the process of estimating the system resource required to deliver performance in accordance with a documented SLA for a forecast level of business activity, a specific application, or a particular area of the business. In other words, the aggregation of the capacity or application sizings for all of the applications used in a business or for all of the sections of a business should create the overall capacity plan for that business.

An important point that must be emphasized with regard to this addition of capacity sizings or application sizings is that this addition is **not** necessarily linear. In other words. if two applications running independently and exclusively on a particular system have CPU utilizations of A% and B% respectively, then if these two workloads run on the same system concurrently, it is unlikely that the

© Copyright IBM Corp. 1998

CPU is only (A+B)%. This is due to the increased level of contention on the system, not only for hardware resources, but also possibly for data resources (that is, possibly higher levels of "file locking"). However, at low levels of utilization on a system, reasonably accurate answers may be obtained by simply adding utilization figures.

2.3 Capacity Planning Prerequisites

This section presents an overview of the factors that need to be considered prior to launching a capacity planning exercise. The main prerequisites are as follows:

- **Work Management must be set up properly**: The requirements of the business users should be taken into account when the work management setup is put in place. For example, if a particular set of users need especially good response times, they should be put into a high priority subsystem designed to deliver these faster response times.
- **Workload profile must be established**: Most important, the time of day of the "peak" level of high priority activity on the system must be established, and ideally some appreciation should be developed of the degree of peaking in the workload. This can help when assessing the "sensitivity" of a particular capacity estimate.
- **IT environment must be tuned**: We strongly recommend that every effort be made to make the application, its implementation, and use as efficient as possible in the business environment. If this is not done as a prerequisite, any problems associated with the inefficiencies are carried through with the increased workload, causing increased resource overheads.
- **All components must be below utilization guidelines**: None of the components in the IT environment should be operating above their published utilization guideline figures for good performance; see Appendix A in the *AS/400 Performance Management V3R6/V3R7*, SG24-4735.
- **History of performance data should be available**: Ideally, a year of data is required for existing workloads, but certainly not less than three months worth of data must be available as an absolute minimum.
- **Details of new applications should be known**: If any new applications are to be added to the system under study, it is important that both a workload profile and details on the likely intensity of usage are available.
- **SLA should have been published**: Documented specification of required performance should be available.
- **Growth data needs to be available**: Ideally, this data should be expressed in units of measure that can be turned into meaningful information for capacity planning purposes. For example, if the forecast rate of working in the "peak" hour is expressed in "computer transactions per hour", this is ideal. It is extremely likely in the real world that the data will be available in this form. It is more likely that the growth forecast is expressed as a rather vague overall percentage figure derived from a sales target or other high level business objective. One of the key skills of a good capacity planner lies in deriving individual growth figures for each of the main applications on the system.

In the "real world", it is rare that all of these prerequisites for a capacity planning exercise are in place. However, if they are not in place, the accuracy of the outcome of the exercise is adversely affected. It is, therefore, important that any shortcomings in these areas are fully documented in the final deliverable from the capacity planning exercise so the readers of the plan can make the appropriate allowances.

2.4 Simple Capacity Planning Process

The steps to follow in this simple version of a capacity planning exercise can be summarized as follows:

1. **Build a spread sheet to summarize collected data**. Review the performance data that has been collected and build a spread sheet to summarize the data for the "peak" hour from each day.

Peak Hour (11-12):	Mon 18/8	Tues 19/8	Weds 20/8	Thurs 21/8	Fri 22/8		Ave Wk 34
System Level:							
Ave Resp Time	1.00	0.63	0.67	0.65	0.59		0.71
Total CPU Utilisation %	96.9	89.3	96.5	87.6	86.6		91.4
Applied Workload:							
Txs / Hr - excl DDM	75,136	81,348	75,529	76,519	78,264		77,359
Number of Active Jobs - Inter	1,139	1,164	1,206	1,203	1,101		1,163
Batch Log I/O per sec	1,517.0	762.1	1,896.0	467.3	409.8		1,010.4
Number of Active Jobs - "Batch"	216	227	232	216	237		226
Interactive Profile:							
Ave Resp Time	1.00	0.63	0.67	0.65	0.59		0.71
Ave Client Acc CPU Secs / Tx	0.10	0.09	0.09	0.10	0.09		0.09
Ave Client Acc SDIO / Tx	9.9	9.5	10.8	9.4	8.9		9.7
Ave Client Acc DBIO / Tx	144.5	144.0	110.2	140.0	142.7		136.3
CPU Breakdown:							
Machine (P<10)	9.8	9.4	9.3	8.9	8.8		9.2
Interactive	56.8	54.7	48.3	53.2	54.3		53.5
The rest	30.3	25.2	38.9	25.5	23.5		28.7
Total High (P <26) - GL<81	92.5	76.5	77.8	71.2	73.0		78.2
Batch at P>50	3.1	10.6	17.3	14.2	11.3		11.3
Main Storage:							
Tot Faults all pools - GL<400	264.8	223.5	282.0	176.5	163.2		222.0
Non DB in *MACHINE- GL<10	3.3	3.5	4.3	3.6	3.2		3.6
Disk Performance:							
Space Utilisation %	89.2	90.3	91.5	91.8	86.3		89.8
Arm Utilisation %	8.5	6.7	8.5	5.4	5.2		6.9
Response Time (milli secs)	8.6	7.9	9.5	7.7	7.2		8.2

Figure 6. Spread Sheet of Performance Summary for Peak Hour

All of the data in the preceding spread sheet can be obtained from the Performance Tools/400 System Report. The data has been organized to allow easy assimilation of the key aspects of workload and performance:

- A **system level** appreciation of overall performance given by "total CPU utilization %" and interactive response time. However, interactive response times are only easily available for any non-client/server interactive workloads. This gives a quick assessment of the performance of the system.
- The **applied workload** that is tailored to the particular environment. In the following sample, DDM transactions have been judged not to be important and have been excluded from the metric of interactive workload. However, these figures give a simple view of how the applied workload is behaving (is it increasing or decreasing, and so on).

- **Interactive profile**: For systems with significant non-client/server interactive workloads, this is valuable information. It should remain relatively unchanged over time, but if it does change, this should alert the capacity planner to revisit the capacity plan as the application or how it is used has changed significantly and this should be investigated.
- **CPU breakdown**: This is the first of three utilization figures that should be monitored carefully. The high priority CPU figure is usually the figure that should be watched most closely because if it exceeds its threshold figure, high priority work begins to suffer.
- **Main storage**: The guideline figures here are obviously dependent on the model but they provide again a quick check, and if the figures approached the guidelines, then further investigation should be instigated.
- **Disk performance**: Both arm utilization and space utilization need to be monitored as they can adversely affect performance if they exceed their guideline figures (50% and 85% respectively).

Because we are only considering capacity planning for the AS/400 system, and are not taking into account any network issues, we have ignored any communications-related utilization figures.

2. **Calculate weekly averages and plot "best fit" lines**: For each week, determine the average set of figures for the week. Review the figures to see which of the three crucial figures (CPU, main storage, or disk drives) is closest to its threshold figure and extract that figure into a separate worksheet. In the following example, Figure 7 on page 17 it is CPU that is closest to its threshold figure. Since the system in question is a 4-way, the CPU threshold figure is 81%. The sample worksheet also contains the interactive transactions per hour as this was found to be the key driver behind the growth in CPU. It should be noted that the high priority CPU in the example has been set at 26 rather than the more normal figure of 20. This is because there is high availability software in use on this system and this work was considered important; hence, was included in the high priority category.

   In the worksheet, the mathematics is carried out to calculate the coordinates of a simple "best fit" line for transactions per hour and for high priority CPU percentage. The graphs included on the following page illustrate both the actual figures and the "best fit" lines. The graphs show that the best fit lines are reasonably good approximations; hence, they give a reasonable indication of future events. The "forecasts" included at the bottom of the worksheet suggest that the system will exceed the CPU utilization threshold in four to five weeks.

Weekly Averages:												
		Actual Inter				Best Fit			Actual High			Best Fit
	Wk #	Txs / Hr				Txs / Hr		Wk #	P CPU %			CPU %
N = 17	X	Y	X.X	X.Y	Y.Y	Y(X)		X	Z	X.Z	Z.Z	Z(X)
Wk 22	22	63,469	484	1396313	4028282227	63,131		22	60.4	1328	3645	63.6
Wk 23	23	66,049	529	1519136	4362523240	64,186		23	72.6	1670	5271	64.5
Wk 24	24	66,047	576	1585128	4362206209	65,242		24	66.2	1590	4388	65.4
Wk 25	25	67,925	625	1698125	4613805625	66,298		25	69.3	1732	4800	66.4
Wk 26	26	67,885	676	1765010	4608373225	67,353		26	68.9	1791	4747	67.3
Wk 27	27	68,068	729	1837836	4633252624	68,409		27	64.2	1734	4127	68.2
Wk 28	28	64,409	784	1803452	4148519281	69,464		28	64.5	1805	4158	69.2
Wk 29	29	69,600	841	2018394	4844132160	70,520		29	69.9	2027	4886	70.1
Wk 30	30	73,429	900	2202858	5391759298	71,576		30	74.5	2236	5553	71.0
Wk 31	31	74,219	961	2300795	5508489649	72,631		31	70.6	2190	4990	72.0
Wk 32	32	71,403	1024	2284902	5098416970	73,687		32	67.6	2163	4570	72.9
Wk 33	33	69,367	1089	2289124	4811836183	74,742		33	66.6	2198	4436	73.8
Wk 34	34	77,359	1156	2630213	5984445825	75,798		34	78.2	2659	6115	74.8
Wk 35	35	80,305	1225	2810658	6448812720	76,854		35	71.6	2507	5130	75.7
Wk 36	36	71,809	1296	2585131	5156561205	77,909		36	76.7	2760	5880	76.6
Wk 37	37	79,637	1369	2946562	6342019914	78,965		37	78.2	2892	6109	77.5
Wk 38	38	85,805	1444	3260598	7362532347	80,020		38	87.5	3323	7649	78.5
Totals	510	1,216,785	15708	36934235	87705968702				1207.5	36605	86454	
17	X	Y	X.X	X.Y	Y.Y	Y(x)			Z	X.Z	Z.Z	Z(X)
	Y(X) = A + B * X									Z(X) = C + D * X		
	where:	A =	39908							where:	C =	43.08824
	and	B=	1055.6							and	D=	0.931373
Forecasts:												
Wk 41	41					83,187						81.3
Wk 45	45					87,410						85.0
Wk 52	52					94,799						91.5

Figure 7. Spread Sheet and Best Fit Line for Weekly Averages

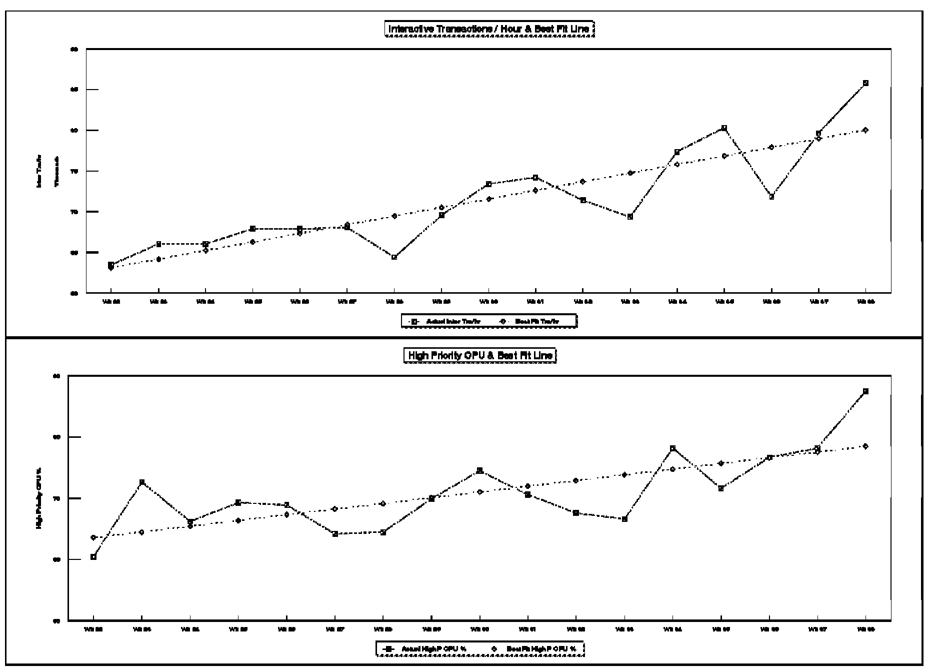


Figure 8. Transactions/Hour and High Priority CPU (Actual and "Best-Fit" Lines)

3. **Forecast when an upgrade is required**: In the preceding sample data, it is clear that the high priority CPU is increasing at an average rate of 0.93 percentage points per week. The figures show that this increase is being

driven by a growing rate of interactive transactions per hour. The average figure does not adequately represent the urgency of the situation because examining the graphs reveals that there is a "monthly cycle" in addition to the more general growth rate. This monthly cycle is causing the high priority CPU to be particularly high every fourth week due to heavy month end processing. It is important, therefore, to take both of these effects into account. If this is done, there must be a particular concern about the system performance in weeks 42 and 46 (that is, the upcoming weeks with month end processing). We, therefore, conclude that upgrade action should be taken before week 46, and every effort should be made to manage the workload up to that time.

These three simple steps illustrate simple capacity planning, but they do not answer the questions that any business usually asks with regard to upgrades to their system. These questions are:

- What upgrade is required and what will it cost?
- How long will the upgrade last?

To answer these questions, we have to leave behind our simple graphs and "best-fit lines", and use the Best/1 modelling tool.

2.5 Capacity Planning with BEST/1

In summary, the main steps in a BEST/1 modelling exercise are as follows:

1. Select a suitable data collection to base the modelling on.
2. Build the model in BEST/1.
3. Calibrate the model, if required or necessary.
4. Use the model to carry out the "what if" exercises to answer whatever questions are being posed by the business and build a Capacity Plan.

The *BEST/1 Capacity Planning Tool*, SC41-3341, covers the use of the modelling tool. However, the following additional notes should be considered:

- **Selecting performance data**: Selecting the particular performance data that the modelling is based on is one of the most important steps in the entire capacity planning process. However, it is seldom recognized as such and often an individual is asked to carry out a capacity planning exercise on a single "snap shot" of performance data. It is extremely unlikely that a good capacity planning exercise can be carried out if this is the situation.

  To allow a good capacity planning exercise to be carried out, at least three months of data needs to be available as we previously stated. Then from a spread sheet of the type shown in Figure 6 on page 15, a suitable "peak hour" can be selected where the following criteria has been met:

  – None of the performance thresholds have been exceeded; if this is the case, additional queuing is present.

  – It is one of the peak hours that has one of the highest figures for high priority CPU utilization.

  – It is one of the peak hours that has one of the highest figures for external workload metrics (that is, the level of interactive transactions per hour is one of the highest).

  – The amount of non-high priority work being processed is average. This can be assessed by reviewing the figures in the spread sheet for CPU allocated to priority 50 work.

– The interactive transaction profile is an average. The key values to consider include "DBIO/Tx" and CPU Secs/Tx.

The goal of these criteria is to ensure that a data member is selected that is one of the highest collected but is also representative of the typical workload.

If these criteria were applied to the data in Figure 6 on page 15, the decision making is as follows:

1. Exclude Monday's figures because the high priority CPU utilization at 92.5 clearly exceeds the threshold figure.
2. Of the remaining days, Thursday and Friday are excluded because they both exhibit lower high priority CPU, which leaves the choice between Tuesday and Wednesday.
3. Between Tuesday and Wednesday, Tuesday is the best choice because its data is more representative. This is borne out by the fact that Tuesday's DBIO/Tx and amount of priority 50 work are closer to the average figures than Thursday's.

Clearly, it is more difficult to select a particular member when there is more data to be considered than the five members in the preceding work example. However, even if months worth of data are available, it is relatively easy to make a valid and considered choice if the data has been summarized into a spread sheet of the type shown.

- **Model Creation**: It is important at this stage to think ahead to the "what if" stage that takes place towards the end of the process. This is because to enable the "what if" stage to answer the required questions, the "job classification" needs to have the appropriate flexibility built into it. For example, if one of the questions that is being asked is: "What size of system do we need to support an increase in our tele-sales department of 100% while the activity among our other users only increases by 15%?" Such a question can only be answered if the tele-sales users can be allocated to a specific "workload" so that it can be inflated by the 100% figure. The definition of this tele-sales workload is achieved through creating a suitable "job classification". The means that are used to achieve this are dependent on the work management setup. If the tele-sales users have been assigned to a specific subsystem, then the job classification is based on subsystem. If this is not the case and there are not too many users in the tele-sales department, then USERIDs are a possibility. On similar lines, if a functional group has been defined over the tele-sales department USERIDs, then functional groups may be used.

  Subsystems are a particularly popular basis for job classifications because many applications from third parties set up their own subsystems as part of the package installation process. In addition, many large AS/400 installations find the "command and control" benefits that segment their users into separate subsystems facilitates many of the day-to-day operational tasks. We, therefore, strongly recommend to all medium to large installations that user segmentation by subsystem is put in place.

- **Calibrating the Model**: Once the model has been created, it needs to be calibrated to ensure that it genuinely reflects the real system. Later chapters discuss some of the special issues on calibrating client/server workloads, but it is worth highlighting a point that also applies to non-client/server models here.

In a non-client/server model, there is usually some level of Exceptional Wait (EW) that needs to be added to the interactive response times to ensure that the actual and predicted figures agree. This is done by using F17 as indicated in the following sample displays:

```
                        Measured and Predicted Comparison

                                           Measured     Predicted
Total CPU util . . . . . . . . . . . :        87.4         87.4
Disk IOP util  . . . . . . . . . . . :         5.2          2.2
Disk arm util  . . . . . . . . . . . :        11.1         11.1
Disk IOs per second  . . . . . . . . :      1229.2       1228.1

LAN IOP util . . . . . . . . . . . . :         6.3          2.0
LAN line util  . . . . . . . . . . . :          .0           .0
WAN IOP util . . . . . . . . . . . . :         4.8          2.1
WAN line util  . . . . . . . . . . . :          .0           .0

Interactive:
  CPU util . . . . . . . . . . . . . :        66.0         66.0
  Int rsp time (seconds) . . . . . . :         1.7           .5
  Transactions per hour  . . . . . . :       14108        14095
Non-interactive thruput  . . . . . . :      140122       140174


 Performance estimates -- Press help to see disclaimer.
 F3=Exit   F6=Print   F9=Work with spooled files   F12=Cancel
 F17=Calibrate response time
```

However, the displays show that on this particular model, the exceptional wait added is a rather large 1.2 seconds. It is, of course, correct to add this exceptional wait to ensure that the model correctly reflects the response times, but what about the situation where this workload is measured on a relatively small system and then moved to a much larger machine. This is not an uncommon occurrence and once this 1.2 seconds has been added, it means that it is not possible to get a response time of sub-seconds from this workload. We, therefore, recommend that if capacity planning is going to involve moving a workload to a much larger AS/400 system, the exceptional wait is **not** added in. Instead, we recommend that you mentally note that the exceptional wait is not added, and once the workload has been moved to the much bigger system, add an exceptional wait into the appropriate transactions through the following display.

```
                               Change Transaction

Workload  . . . . :   RQ10       Measured from QPFRLAW650 (L31OCT97R1)
Function  . . . . :   RQ10       Function of RQ10

Type choices, press Enter.
  Transaction Type . . . . . . . . .  2            1=Inter, 2=Non
  Pool ID  . . . . . . . . . . . . .  5            F4 for list
  CPU Priority . . . . . . . . . . .  20
  CPU time . . . . . . . . . . . . .  3.412        Secs (on B10)
  Permanent writes . . . . . . . . .  74.4         0-100 percent
  Chars transferred in . . . . . . .  0
  Chars transferred out  . . . . . .  0
. Exceptional wait  . . . . . . .0                  Msec
  Paging behavior  . . . . . . . . .  *GENERIC     F4 for list

                             ------Database-------     ----Non-database-
                              Reads      Writes        Reads      Writ
  Sync I/Os  . . . . . . .   .0         .0            3.3        .1
  Async I/Os . . . . . . .   .0         .2            .0         .0

F3=Exit   F4=Prompt   F6=Calculate B10 CPU time   F12=Cancel
F13=Change paging behavior
```

From the preceding example, it shows that using this facility gives much more flexibility in that the exceptional wait is added at the transaction level, not as an overall value for all workloads.

- **Exercising the Model**: Once the model has been built, it is important to emphasize that it is only an estimate of the future performance and is based on a large number of assumptions and on approximations of the level of growth. The output from the model (with its neatly printed rows of numbers showing response times measured in tenths of a second and utilizations in tenths of a percent) conveys an aura of accuracy that is simply not there. Hence, we recommend that (in addition to the system generated warnings) statements are made advising the reader of the plan and that it must be treated with care. One of the best ways to emphasize this aspect of capacity planning is to, in fact, develop at least two models. One of the models should be "high growth and based on an absolute maximum sample of performance data" and the other should be "normal growth and based on a representative sample of performance data". By presenting both of these forecasts in your capacity plan, you are forcibly advising your readers that your plan is an approximation. The other point that is worth emphasizing is that the further into the future you model, the greater the margin of error is on your plan.

Chapter 3. The AS/400 System in a Client/Server Environment

Traditionally, the AS/400 system has been considered a source of computing power for processing interactive requests in a commercial environment. More recent developments in application software have resulted in the introduction of complex client/server business applications.

This chapter discusses the role played by the AS/400 system as a server in the client/server computing environment. A brief discussion on client/server concepts is used to introduce the topic. A section is included on the various components that constitute the response times perceived by a user of client/server applications.

3.1 Client/Server Concepts

The components in a client/server computing environment are:

- Databases
- Applications
- Presentation

There have been many analyses on the way these components can be grouped together in designing an effective client/server system. Each architectural design has its own advantages and disadvantages. We present the following categorization based on a report by the Gartner Group as an example of the studies on the location of client/server components.

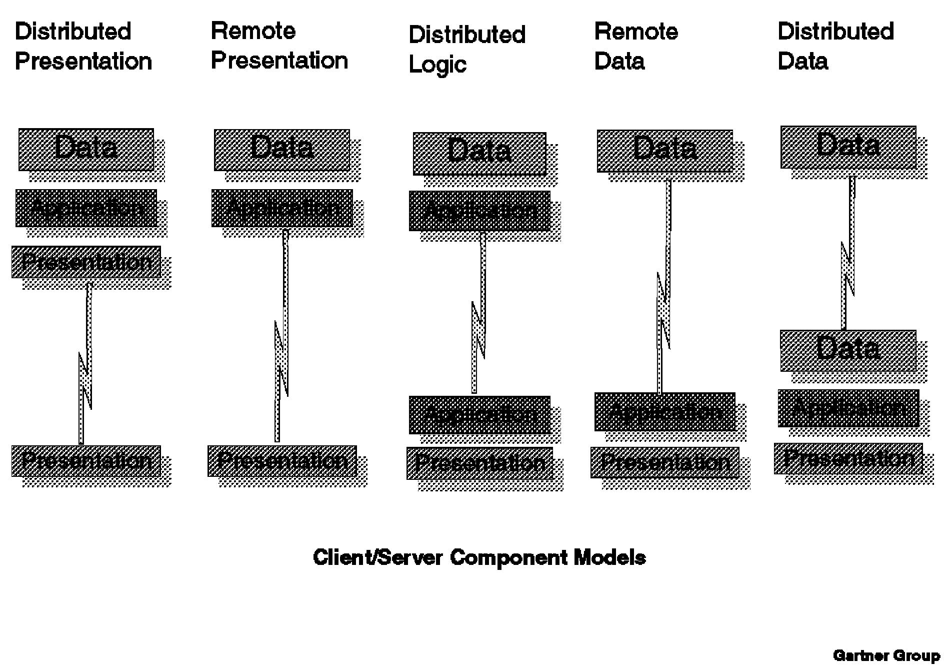


Figure 9. Gartner Group - Client/Server Models

1. **Distributed Presentation**: The data presentation logic is divided between the server and the client.
2. **Remote Presentation**: The presentation layer is confined to the client. Database and application functions are performed by the server.
3. **Distributed Logic**: Database and application logic is associated with the server. Additional application code is located on the client together with the presentation functions.
4. **Remote Data**: The entirety of the information processing logic is on the client together with the presentation. The server provides the necessary support for the database management and services.
5. **Distributed Data**: The database is located at the server. The more "localized data" is retained on the client to minimize data traffic.

In addition, it is possible to design systems having the preceding combinations implemented to varying degrees. While a PC is frequently the client device in most client/server environments, this is not necessarily a requirement.

Some applications use the client exclusively as a presentation device with no processing logic (remote presentation). In this environment, the PC is a good candidate to function as the client. Other applications have a significant amount of information manipulation performed at the client in addition to functioning as a user interface (distributed logic). In some applications, by localizing specific data on the client, the volume of data transmission is minimized (distributed data).

There are many considerations including performance, data security, data integrity, and so on that need to be considered. It is outside the scope of this document to evaluate and review the considerations for client/server application design.

Please refer to *AS/400 Client/Server Performance using the Windows Client*, SG24-4526, and *Client/Server Performance using Application Development Tools*, SG24-4731, for more information on performance and design considerations of client/server applications in an AS/400 server environment.

3.2 Components of Client/Server Response Time

The response time perceived by the user of a client/server application is the sum of many components. At each component, a request is made to the service provider, and some time elapses before the response is received by the requester. The major components of response times are illustrated in Figure 10 on page 25:

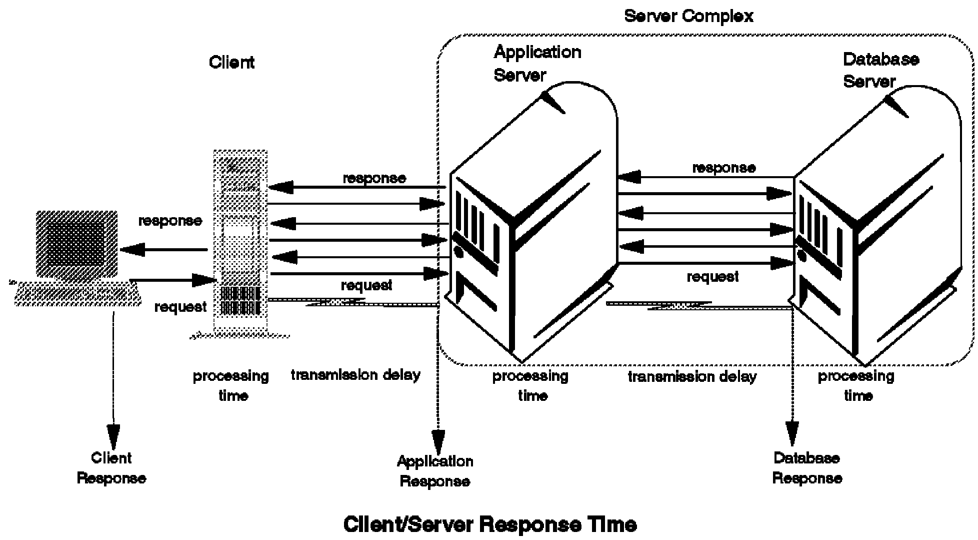


Figure 10. Client/Server Response Time

Depending on the architecture of the specific application, only some of the components may be applicable.

- **Client processing**: A user at the client device completes a request by pressing the Enter button, selecting a Function key, or making a selection with the mouse. The request is initially processed by the client, which may require additional data or information processing functions from an application server. A single request from the user may translate into multiple requests to the application server. Also, after receiving the response to any request of the application server, additional processing may occur prior to completing the response to the user request.
- **Application server processing**: If the client requires additional information processing functions from an application server, the request is processed by this system. The application server may make a request for additional data stored on a database server prior to completion of processing and responding to the request from the client. A single request from the client may translate into multiple requests to the database server.
- **Database server processing**: This component accesses the data requested by the application server and returns the required data to the application server. While the extent of processing on the database server is minimal, there is a need to retrieve large data volumes.
- **Communications transmission delay**: Between each of the previous processing units, the communication links result in some delay. The bandwidth of the communications links and the volume of information to be moved between the participating processors determines the communications delay in addition to the number of line turnarounds between the client and the application server and the application server and the database to complete a request.

3.3 AS/400 Server Job Characteristics

Typically, AS/400 jobs can be differentiated into:

- **Interactive jobs**: These jobs are associated with 5250-type workstations, allowing for users to directly interact with a job running on the AS/400 system. The throughput of interactive jobs is dependent on the availability of requests from users (assuming there are no computer system resource constraints). A particular system configuration can have a throughput capacity limit dependent on the hardware, operating system, application software, and how the application functions are used.

  Often the request reaches the AS/400 system through a communications controller or a workstation controller. Each request results in the job actively using processor cycles over a short period of time compared to relatively long periods during which there is no demand for processor cycles caused by user key-think time. A relatively long wait time between periods of processing is a characteristic of interactive jobs. The AS/400 Performance Tools measure definite request/response boundaries that are used to estimate response times.

- **Batch jobs**: These jobs are initiated by users, but run without any further interaction with the user initiating the job. The throughput of batch jobs is dependent on the availability of system resources such as disk (providing data in response to a request from application) and CPU cycles to process the data made available to the processor. A typical batch job has no external agent to restrain its progress to completion and is usually active over long periods of its elapsed time. Short wait times (compared to that in interactive jobs) between periods of processing are characteristic of batch jobs running in an unconstrained environment. The AS/400 Performance Tools do not recognize or report request/response boundaries.

- **Server jobs**: These AS/400 jobs respond to an external event such as a request from a client device. When the AS/400 system is functioning as a server by responding to requests from a client device, the server job on the AS/400 system is reported by the performance tools as a non-interactive job without any recognition of a request/response boundary. However, these server jobs exhibit "interactive" characteristics in that:

  - Job activity is initiated by an external event (a client request).
  - Progression of the job is interrupted by user key-think time.
  - A unit of activity is represented by a user initiated request (from the client) and a response (from the server).

AS/400 support for data queues and message queues makes the distinction between interactive, batch, and server jobs even less clear. Jobs interacting with data queues, message queues, and client devices are reported as "batch" jobs even though their activity is determined by interactions with the corresponding request queues or client devices.

Activity through the AS/400 Client Access product is reported as an interactive workload with distinct request/response boundaries but the workload may not be typical of an interactive dialog.

It is necessary to distinguish between interactive and non-interactive workload so that we can associate the workload on an AS/400 system with the throughput capacities published by IBM for the various server models of the AS/400 system.

3.4 Server Model Considerations

The AS/400 advanced servers and AS/400e servers are intended for use in environments where the major processing requirements are from non-interactive workloads such as server applications and batch processing. While IBM-5250 type interactive work can be performed on a server model, the performance characteristics of the AS/400 servers work in favor of the non-interactive work at the expense of interactive work. The price/performance of AS/400 servers provide an excellent solution for application workloads that are predominantly non-interactive.

A special range of AS/400 servers called **AS/400e custom servers** has been introduced to provide an economical solution for customers who are migrating from predominantly 5250-type interactive applications to client/server applications. These models are designed to provide balanced capacity and performance in a "mixed" environment of interactive and non-interactive workload. The AS/400e custom servers are offered together with specific Software Vendor (SV) solutions. However, the underlying performance design basis is the same as the AS/400 advanced servers and AS/400e servers.

3.4.1 Server Model Throughput

All AS/400 server models have two published throughput ratings measured in CPW (Commercial Processing Workload) units. The larger CPW value represents the throughput capacity of the model when it is exclusively running non-interactive workload. Alternatively, if the system is running interactive work only, the capacity of the model is represented by the smaller CPW value. The **combined workloads** represented by the maximum interactive and non-interactive CPWs **cannot** be run on a system. This means a model S20-2166 for example, can be operated with a maximum CPW value of 759.0 non-interactive OR with an interactive CPW value of 56.9 and NOT both values added together.

While it is possible to run interactive jobs on a server model, the lower CPW value for interactive workload indicates that there is a penalty in CPU overhead to run interactive work on a server model. The penalty is applicable beyond a particular level of interactive usage. Beyond this threshold value, it becomes increasingly costly from a processor standpoint to run interactive work. This threshold value is expressed as a percentage of the total interactive capacity and is dependent on the model of AS/400 server. These values are:

- Approximately 33% for servers
- Approximately 87% for custom servers

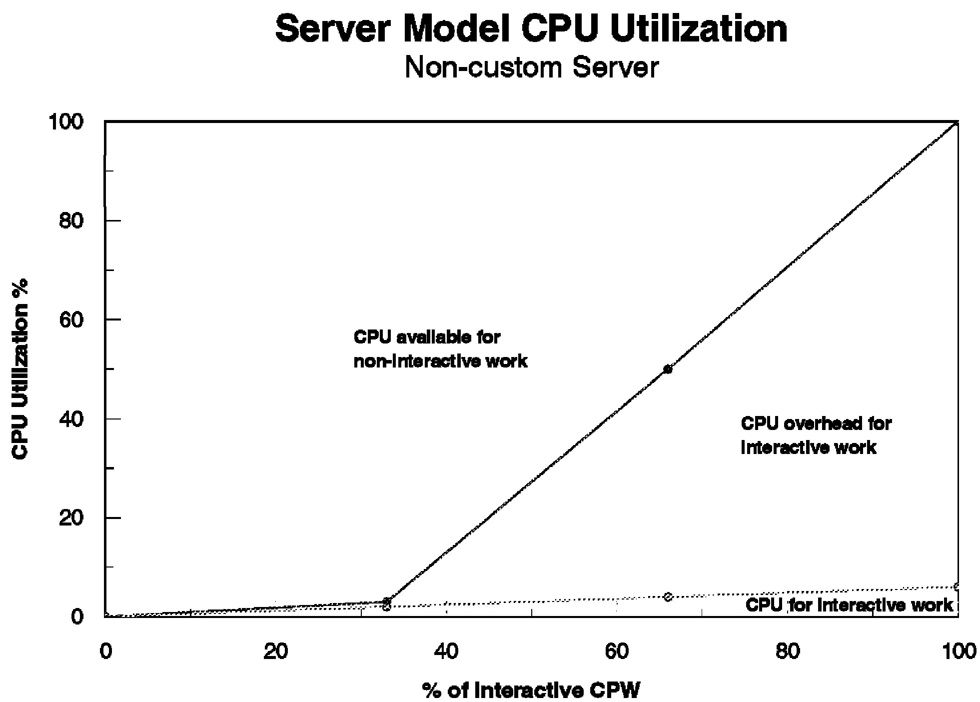


Figure 11. Server Interactive Workload

Figure 11 illustrates the increase in total CPU usage as the amount of interactive usage increases within the relatively small range of interactive CPW capacity. As the interactive workload increases beyond the threshold level, there is a significant reduction in available CPU capacity for non-interactive work.

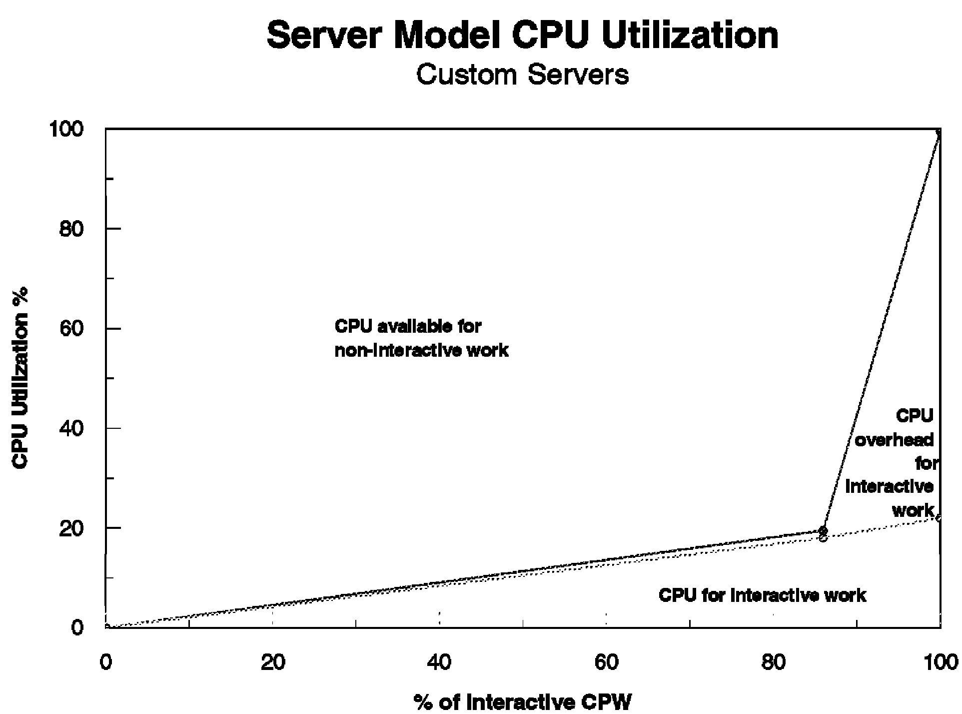


Figure 12. Custom Server Interactive Workload

Figure 12 illustrates the relationship between interactive workload and total CPU utilization in the custom servers. The penalty paid for interactive usage becomes effective at much higher levels of interactive workload. Thus, interactive and non-interactive workload can coexist far more effectively with custom servers.

You must ensure that the selected custom server has adequate capacity to support the planned workload for interactive and non-interactive work as separate entities. The total CPU workload arising from:

- Interactive work
- Non-interactive work
- Any penalty for excessive interactive work

must not exceed the capacity of the processor.

In OS/400 V4R1, the algorithm governing the penalty for running interactive work on a server model has changed. The adverse effect of exceeding the interactive threshold can be better controlled by the customer. Through the use of dynamic prioritization, interactive jobs are placed lower in the priority queue if their processor usage reaches the threshold value. This causes the interactive work to slow down as a result of not getting as many processor cycles and therefore allows for more batch throughput to continue. The non-interactive jobs should have a **priority setting between 20 and 35** and the system value QDYNPTYSCD (Dynamic priority scheduler) set to 1 to benefit from the new algorithm.

3.5 CPW Values in Capacity Planning

The Commercial Processing Workload (CPW) is a synthetic workload used to measure the relative throughput capabilities of the AS/400 family. This workload may not be representative of any particular customer environment.

The overall throughput in any processing environment results from the combined effect of hardware capability and operating system efficiency. Release-dependent enhancements in the operating system improve performance in selected functions with little or no improvement in others.

For example, in OS/400 release 4.1, there are several improvements in the code used by Extended Dynamic SQL functions. However, since the CPW workload does not include these functions, the benefits of the code enhancements have no effect on CPW throughput. On the other hand, improvements in journal, memory, and contention management may effect the CPW throughput more than a particular customer workload that was not impacted by these aspects. Thus, the overall impact of operating system enhancements in a customer environment is heavily dependent on the nature of the application code.

Also, the CPW environment measures throughput at close to the maximum CPU capacity on systems that are configured to ensure no memory or disk constraints, which is often different from most customer situations.

BEST/1 uses performance tables that focus on the hardware power of the systems and less on the specific code paths used by the workload. It also allows for estimating the effect of memory and disk on the overall throughput of the customer environment and handles the effect of job priorities on throughput.

The BEST/1 modelling process is based on average computing resource utilizations and capacities, combined with adjustments for release-dependent operating system enhancements in a general commercial environment. Thus, it is possible that results produced by BEST/1 may be more conservative than those predicted by using basic CPW values. However, considering the levels of approximation associated with the predictive nature of capacity planning, BEST/1 provides a sound basis for estimating hardware requirements to achieve planned workload throughput and average response times.

3.6 Capacity Planning Unit of Measure

In the light of the differences between interactive and non-interactive workload, it is prudent to consider all work performed by the AS/400 processors as **tasks**. Each task represents one or more requests for service and associated responses.

Each request/response bounds an identifiable unit of activity. In this document, we refer to this as a **Capacity Planning Unit of Measure** (CPUM). The boundaries of a **CPUM** can be determined according to the needs and measurability afforded by the specific application being considered. You are free to select the most suitable set of boundaries for the request/response considering measurement facilities available in the business environment, the application suite, and the AS/400 server.

For example, a traditional "batch job" can have a request such as the complete processing of specific input data sets or the achievement of specific completion criteria. In this case, the boundary of this Capacity Planning Unit of Measure (CPUM) is the beginning and end of the batch task with a *CPUM* value of one. The response time is the elapsed time of the task.

In the case of an interaction with a user or client device, the request is initiated by pressing the Enter button, a Function key, or by a selection made with the mouse. The response to this request completes the capacity planning unit of measure (CPUM). This is similar to a "transaction" referred to in interactive AS/400 performance measurements. The computer systems and communications activity and the communications line turnarounds that occur within a capacity planning unit of measure (CPUM) is dependent on the specific client/server or interactive application.

The processing characteristics of a **single request/response** associated with a batch job, an interactive job, or a server job are similar in that it proceeds to completion as quickly as possible provided it does not encounter constraints in processing and disk capacity, and is subject to throughput management attributes such as time slice and execution priority considerations.

3.6.1 Business Transaction

It is important to distinguish these request/response interactions between the user and the client, or between the client and the server complex from a **business transaction**, which is made up of multiple user interactions to complete a single business transaction such as a sales order, purchase order, or journal entry.

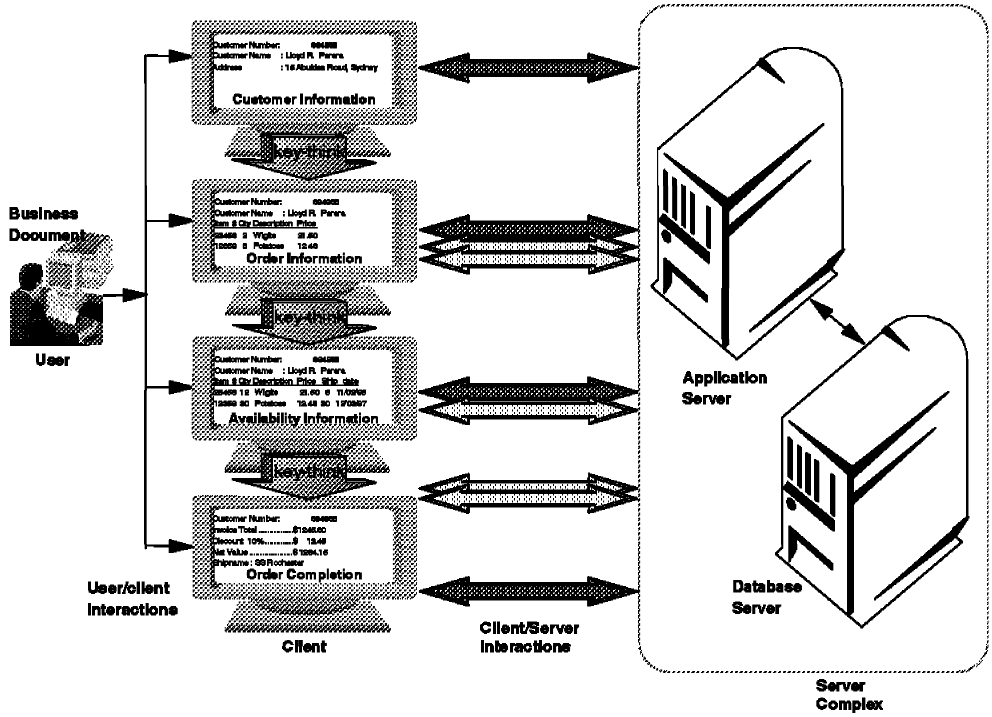


Figure 13. Business Transaction

For example, in completing a single business transaction such as a sales order, the user navigates through multiple interactions with the client device. In Figure 13, which represents a single order entry business transaction, there are four interactions between the user and the client device (PC) with windows showing information on:

- Customer data
- Order requirements
- Stock availability
- Order summary and so on

Each interaction between the user and the client device has varying numbers of requests/responses to the server complex, all of which are transparent to the user. The number of requests to the server may depend on the processing performed by the client for each window displayed.

In this example, you may select any of the following interactions as your CPUM:

- The business transaction
- The user interaction with the client device
- The client interaction with the server complex
- The application server interaction with the database server

3.6.2 Measurement

In certain environments, the application software may provide precise measurements of numbers and response times of each window change at the client. The measurement of response time is taken at the client or at the server. Accordingly, take into consideration the various components that are included in the measure. Some measurements taken at the server application may not have a direct relationship to the response times perceived at the client.

For example, if there is a significant amount of processing performed on the client that results in many interactions with the server application for a single request/response at the client, the measurement at the client indicates a relatively long response time for a single request/response. The same workload measured at the server shows multiple request/response counts with relatively rapid responses from the server. Thus, it may be difficult to correlate the response times and number of interactions perceived by the user with those measured at the server.

Your choice of CPUM depends on the availability of suitable measurements to correlate the number of interactions with response time. Also consider the point at which the measurements are taken and provide for the various components included in the response time.

If the information is measured by the:

- **User**, you may count business transactions and the average time to process a transaction.

  In this case, note that there may be significant variation in the completion time (response time) of a business transaction because of the many occurrences of user reaction time within each business transaction. The user reaction time includes think time and typing-in time in addition to any other human interaction time. These delays reduce the demand on computing resources by reducing user throughput, but must be considered as a component of the overall response time for the business transaction.

- **Client**, you may count the window changes and the average time to refresh a window with the response.

  You need to appreciate that the response time includes not only work performed by the client and server complex, but also communications delays. The single user interaction can result in multiple requests to the server as in the case of a "fat client" performing considerable processing. On the other hand, if the client is only a "presentation" device, then each request to the client results in a single request/response to the server complex. Also depending on other work being performed concurrently on the client, there may be a significant delay on the client also.

- **Application server**, the count reflects the number of requests received by the application server and the time to service the request and includes any time to access the database.

  The response times include application server processing time and database service time and any communication time between the application and database server. The response times perceived by the user at the client include communications and client overhead in addition to the response time measured by the application server.

- **Database server**, the count shows the number of database requests and the time to respond with the data.

  Here the response time measured is the database service times. The response time perceived at the user/client interface includes the client overhead, application service time, and the communications delay.

Chapter 4. BEST/1 for Server Capacity Planning

The objective of any capacity planning effort is to determine the optimum, cost effective configuration required to support a specified workload (typically based on currently measurable levels of activity) such that acceptable response times are delivered to the user.

This section presents an overview of the factors that need to be considered prior to launching on a capacity planning project. These considerations are valid for most capacity planning efforts and is not confined to an AS/400 server environment.

While it is not within the scope of this document to discuss performance tuning and performance optimization, we strongly recommend that every effort be made to make the application, its implementation, and use as efficient as possible in the business environment. If this is not done, as a prerequisite, any problems associated with the inefficiencies are carried through with the increased workload, causing increased resource overheads.

4.1 Introduction

Capacity planning is a predictive process to determine future computing hardware resources to support the estimated changes in computing workload. The increased workload on computing resources can be a result of growth in business volumes or the introduction of new applications and functions to enhance the current suite of applications.

Resulting from its predictive nature, capacity planning can only be an approximation at best. The implementation of the same application in different environments can result in significantly different computing system requirements. There are many factors that can influence the degree of success in achieving the predicted results. These include changes in application design, the way users interact with the application, and the number of users who may use the applications. It is also difficult to determine external factors that affect the distribution of the workload over a given period of time such as phone-in customer orders during a working day.

The objective of this document is to present techniques to assist in the prediction of AS/400 system requirements where the AS/400 system functions as a server in a client/server computing environment. No attempt is made to determine PC client or network requirements.

The functions of the AS/400e capacity planning facility (BEST/1), available as part of the AS/400e Performance Tools Licensed Program (5769-PT1) Manager Feature, are used to predict AS/400 requirements. Detailed information on the use of this function is available in *BEST/1 Capacity Planning Tool*, SC41-3341.

This chapter introduces the key aspects of BEST/1 that need to be considered in capacity planning projects for AS/400 server applications. It is expected to help readers become familiar with the relevant BEST/1 facilities.

The suggested approach assumes that performance data collected with the Performance Monitor is available to build the model. However, the AS/400 Performance Monitor does not record non-5250 transactions, and, hence, there is

© Copyright IBM Corp. 1998

no granularity in the measurement of resource usage in a AS/400 server application. The AS/400 performance measurements do not recognize individual requests to or responses from a server job. Also, all server applications are "batch-like" and no LAN or WAN times are attributed to them. Hence, it is not possible to use the AS/400 measured performance data only for capacity planning if you require some response time indications.

Any references to "transactions" and "response times" from reports produced by Performance Tools for AS/400 (5769-PT1) refer to interactive work produced by 5250 sessions and **do not apply to any server applications** that may be running at the time. The reports produced using the standard performance tools options provide you with information on the total use of resources such as processor, disk, and so on for server jobs.

There are many ways to implement a client/server application with the AS/400 functioning as a server. This chapter provides an overview of the method applicable to server workload modelling in general. Subsequent sections of this document present specific examples of the applicability this process.

Note

This chapter presumes the reader has at least a moderate understanding and experience with BEST/1. Even so, access to the manual *BEST/1 Capacity Planning Tool*, SC41-3341, is recommended when reading this chapter.

The success of the approach covered in this section is dependent on the user's familiarity with the BEST/1 capacity planning interfaces and options as well as "artistic adjustments" by the BEST/1 user to define a non-interactive transaction within a BEST/1 workload.

You may have experience with client/server modelling techniques that are different than those illustrated in this chapter. If you have been successful using these techniques for doing capacity planning, you are encouraged to continue using them. However, the techniques used in this chapter are also believed to be valid.

The key to successful client/server capacity planning is a thorough understanding of the application implementation and use of performance data collected.

4.2 AS/400 Server Modelling

There is no intent to discuss designing applications for performance, performance management techniques, or system tuning options. The projections start from a known, measurable workload and extrapolate the resource utilization and response times using a specified rate of growth. The basic prerequisites are that:

- The application is running optimally at expected levels of performance.
- There are no application dependent constraints that invalidate BEST/1 extrapolations.
- No changes are expected in patterns of application usage or user behavior.

- A particular period of measurable activity on the AS/400 system can be established as being representative of a sustained high peak workload.
- The proposed increase in workload can be related to the activity measured.
- An acceptable average response time estimate is established.
- A suitable capacity planning unit of measure has been determined based on application knowledge and measurability.

The key aspects covered in this section include:

- Model creation using measured data for a non-interactive workload
- Model validation
- Growth analysis

4.2.1 Assumptions

The capacity planning process makes the following assumptions:

- The applications under consideration have been optimized for the environment it is running in with regard to:
  - Functions
  - Business requirements
  - User behavior
- The performance measurements used in the capacity planning project are a good representation of a typical busy workload on the system, including the mix of activity and volume of work.
- There are no application dependent bottlenecks that prevent growth in throughput or improved response times. For example, an application may support only a single communications I/O Processor, or there may be a common code-path within the application that only allows requests to the application server to be single threaded.
- The performance data is collected on a system that has not reached saturation point with regard to any key system resources such as memory, processor, disk, and so on. When system resources are saturated, they introduce overheads such as queuing that are difficult to isolate and eliminate in the modelling process.
- The nature of applications and its complexity do not change significantly as a part of the capacity projection. For example, a version upgrade of an application suite often results in significant changes in resource requirements associated with additional functions, processing architecture, and user behavior.
- There is no change in the overall use of any functions of the application that result in increased utilization of system resources.

If any of these assumptions are not valid, plan to make allowances for these increases in the projected workload increases.

4.2.2 Understand the Application Environment

Consider the suite of applications whose workloads are to be extrapolated to provide for growth together with any satellite applications that it is associated with. Increased workload on one application may cause a cascading effect on related applications and increase their workload as well. For example, a planned increase in the number of users for a sales order entry system can result in increased activity in related systems such as materials management, production planning, accounting, and so on, as well as extend any batch run times such as periodic accounting statements to customers.

You also need a good understanding of the overall architecture of the application. For example, you need to know if the application uses a distributed logic approach with a significant amount of processing being performed at the client. Associated with this type of application, there may be significant workload on the communications bandwidth resulting from high data volumes and line turnarounds. This knowledge can enable you to determine the impact of the capacity of the AS/400 server on the total end-to-end client/server throughput and response, and assist you in setting levels of expectation with regard to results that may be achieved once your recommendations are implemented.

A good understanding of the application can also assist in taking measurements that enable you to create a reasonably accurate workload model for extrapolation.

4.2.3 Select a Capacity Planning Unit of Measure

Identifying a measurable capacity planning unit of measure is critical to the success of a capacity planning project. The selected unit of measure must be clearly understood and have boundaries that can be defined and identified.

4.2.3.1 Measurability

It is important to measure a specific workload with regard to:

- Response time per CPUM
- Number of CPUMs per unit of time
- Components included in the measure
- Overall AS/400 resource utilization such as:
  - CPU
  - Memory
  - Disk and so on

The application provider may have included the capacity to measure this within the application server code. The application server code on the AS/400 system can recognize the start and end for processing a request from a client. The elapsed time between receiving the request and providing the response is the **response time** of the selected capacity planning unit of measure (CPUM). This gives you a value that represents the AS/400 server response time. At the same time the server code can count the number of capacity planning units of measure (CPUMs) and report this at the end of a measured period.

In cases where the application server has to go to a database server for data (such as in a multi-tier implementation), the application server may or may not include the database service time in the response time measured at the application server. If it includes the external database server time, you must consider the impact of any adverse performance on the database server and the communications facilities on the measured response time.

4.2.3.2 Options

In certain applications situations, the measurement facility at the AS/400 server may not be readily available. For example, in an application using standard AS/400 ODBC server code, there is no standard facility to recognize a request/response, measure the response time, and count the number of requests/responses in a given period of time.

User exits available through the server code may be used, or code in the client application can be used to count the defined CPUMs and measure the response times.

If the measurements are made at the client, it should be recognized that, in addition to the AS/400 server response time, the measured response time includes:

- Any delay or overhead at the client
- Communications transmission delays and line turnaround times

You may even use business transactions such as sales orders. invoices, customer inquiries, payment vouchers, and so on as a capacity planning unit of measure. The volume of transactions are easily calculated for a period of time. However, each business transaction is the product multiple interactions with the computer system with some amount of key-think time between each interaction. In this situation, the capacity planning model must make allowances for these external delays.

4.2.4 Determine Growth Objectives

An effective capacity planning project requires an estimation of the proposed increase in workload that the computing system must cope with. Often the business managers predict increases in terms of sales revenue, profit projections, or other business measurements. These business measurement have to be translated into the impact on computing workload.

You must work closely with the customer to establish these increases and gain agreement on the estimated increases. A simple increase in volume at one level of the business can have a cascading effect on the overall computing activity on the system. Ensure that you document the assumptions and the basis on which the increases in computing workload were determined. This helps in reconciling any differences that may be encountered when the changes actually occur.

4.2.5 Set Expectations

A clear understanding of the expected deliverable of the capacity planning project must be agreed upon. It is important to stress the predictive nature of the process, and the fact that there are many factors that can influence the degree of success in achieving the predicted results.

There are margins for error at every stage of the predictive process including:

- Estimation of workload at the business level
- Existence of application dependant bottlenecks
- Translation of the business workload into CPUMs
- Prediction of user behavior
- Determination of periods of peak activity

For these reasons, the conclusions and recommendation of a capacity planning project can only be an approximation.

The customer should also indicate expected average response times. It is normally the practice to specify response times requirements below a particular value (for example, under 1.5 seconds) for a specified percentage of the transactions (for example, 90% of the interactive transactions).

Averages

The BEST/1 modelling facilities use averages, and its predictions of utilization and response times are averages also.

4.3 Measure Workload and System Resource Utilization

This is a critical aspect of the capacity planning project because it forms the basis of extrapolation for the growth estimates provided by the customer.

- The duration over which performance data is collected should provide representative information over a period of high activity. If the customer cannot identify a period of sustained peak activity, you may have to measure performance data over a period of time to establish a suitable time frame from which to develop a capacity planning model.
- Select a Capacity Planning Unit of Measure that can assist you in modelling the growth. The selected CPUM must be measurable in terms of the quantity and the response time over the measured period.
- Identify the various components of response that contribute to the CPUM response time so that the necessary adjustments can be made when determining the AS/400 resource contributions.

Building a workload model for BEST/1 requires the following basic information:

- AS/400 resource utilization information
- Application workload measurements:
  - Number of Capacity Planning Units of Measure (CPUMs)
  - Average response time per CPUM
  - AS/400 resource category used by the applications

The first step is to collect AS/400 performance data and application workload statistics for the selected period of system activity. In a client/server environment, the information may have to be collected from more than one source if all the necessary information is not available through the AS/400 performance data. The two sets of data to be used in building the model should be from the same period in time so that the information can be correlated.

4.3.1 Collect AS/400 Performance Data

Use the standard OS/400 STRPFRMON command for performance data collection to start the performance monitor. This provides information on the AS/400 resource usage during the period. Select the data collection time interval to provide sufficient granularity to identify a representative dataset for modelling.

```
                     Start Performance Monitor (STRPFRMON)
Type choices, press Enter.
Member . . . . . . . . . . . . . . >*GEN           Name, *GEN
Library  . . . . . . . . . . . . >.a.          Name
Text 'description' . . . . . . . >.b.
Time interval (in minutes) . . . >.c.     5, 10, 15, 20, 25, 30, 35...
Stops data collection  . . . . .   *ELAPSED      *ELAPSED, *TIME, *NOMAX
Days from current day  . . . . .   0             0-9
Hour . . . . . . . . . . . . . .   2             0-999
Minutes  . . . . . . . . . . . .   0             0-99
Data type  . . . . . . . . . . .   *ALL          *ALL, *SYS
Trace type . . . . . . . . . . .   *NONE         *NONE, *ALL
Dump the trace . . . . . . . . .   *YES          *YES, *NO
Job trace interval . . . . . . .   .5            .5 - 9.9 seconds
Job types  . . . . . . . . . . .   *DFT          *NONE, *DFT, *ASJ, *BCH...
               + for more values
F3=Exit   F4=Prompt   F5=Refresh   F10=Additional parameters   F12=Cancel
F13=How to use this display        F24=More keys
```

- .a. Enter the name of an existing library to collect performance data.
- .b. Enter an appropriate description for the dataset.
- .c. Enter a time interval for data "snapshots". A value of **5 minutes** is recommended.

Note: The duration over which performance data is collected can be specified at the time STRPFRMON is run or can be stopped by running the ENDPFRMON command.

4.3.2 Collect Application Workload Data

The major components of the application environment that need to be captured are:

- Number of Capacity Planning Units of Measure (CPUMs)
- Average response time per CPUM
- AS/400 resource allocation category used by the applications (such as job numbers, subsystem name, and so on)

Important

Obtaining this information requires an understanding of the client/server application. It is important that the application software has some method to count the selected CPUMs and provide you with information on the response times.

If this information is not available as a function of the client/server application, it is necessary to establish some measure for the preceding components through an alternative method. This may even be the use of a stop-watch to measure response times!

In addition to this, a knowledge of the environment in which the application runs can enable you to make the necessary category selections in the creating the BEST/1 model.

4.4 Creating a Model Using Measured Data

BEST/1 provides the facility to use measured AS/400 performance data in building a model for capacity planning. The objective is to create separate **workloads** within the model to represent the various server applications that are involved in the capacity planning project.

Once the separate workloads are created, they can be reviewed for the effect of independent rates of growth on resource utilization and response time. When the workloads have been created, they can be saved and used in other BEST/1 models combined with other workloads.

The workloads can be defined using the following categories or groups:

- User ID
- Job type
- Job name
- Account code
- Job number
- Subsystem
- Memory pool and so on

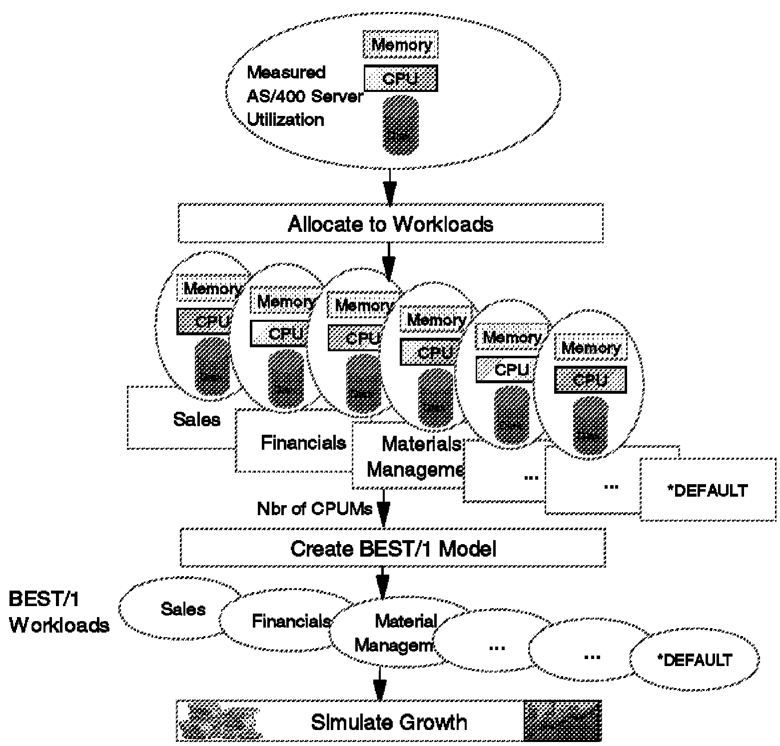


Figure 14. Creating a BEST/1 Model

The major steps in using BEST/1 to create a model with one or more distinct workloads based on your definitions are:

1. Use the STRBEST command to activate the "advanced user level" of BEST/1.
2. Select the performance dataset to be used in modelling.
3. Identify the time period to be considered.
4. Classify the server jobs into suitable workloads.
5. Define the number of Capacity Planning Units of Measure (CPUMs) per hour for each server workload.

6. Create the model.

The remainder of this chapter shows the steps in building a BEST/1 model and validating (calibrating) a model.

4.4.1 Start BEST/1

Sign on the system and start BEST/1 by typing the STRBEST command; press F4, which prompts for input.

```
                         Start BEST/1 (STRBEST)

 Type choices, press Enter.

 BEST/1 data library  . . . . . . .  >.a.     Name, *CURLIB
 Performance data library . . . . .  >.b.     Name
 Log member . . . . . . . . . . . .   *NONE          Name, *NONE
 Log library  . . . . . . . . . . .   *BESTDTAL      Name, *BESTDTAL




 F3=Exit  F4=Prompt  F5=Refresh  F12=Cancel  F13=How to use this display
 F24=More keys
```

- .a. Enter the name of an existing library to contain BEST/1 objects created during modelling.
- .b. Enter the name of the library that has the previously collected performance data.

An initial menu display is shown after the disclaimer display is shown.

4.4.2 BEST/1 for the AS/400 System

On the BEST/1 for the AS/400 menu, select Option 1 (Work with BEST/1 models).

```
                             BEST/1 for the AS/400

 Select one of the following:

      1. Work with BEST/1 models

      5. Create BEST/1 model from performance data

     10. Work with results

     50. About BEST/1

     60. More BEST/1 options


Selection or command
===> .1.

F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
```

4.4.3 Work with BEST/1 Models

Selecting Option 1 (Work with BEST/1 models) creates the following display.

```
                              Work with BEST/1 Models

 Library . . . . .   YOURLIB      Name

 Type options, press Enter.
   1=Create   3=Copy   4=Delete   5=Work with   6=Print   7=Rename

 Opt  Model       Text                                    Date      Time
 .1. .YOURMODEL.
                                                                 Bottom
 Command
 ===>
 F3=Exit   F4=Prompt   F5=Refresh        F9=Retrieve   F12=Cancel
 F15=Sort by model     F16=Sort by text   F19=Sort by date and time
```

Select option 1 to create a model and enter a name for the model you want to build.

4.4.4 Create BEST/1 Model

The following display is shown; select Option 1 (Create from performance data).

```
                              Create BEST/1 Model

Select one of the following:

     1. Create from performance data
     2. Create from predefined and user-defined workloads

Selection
  .1.

F3=Exit   F12=Cancel
```

4.4.5 Create BEST/1 Model from Performance Data

Enter the following information and press Enter.

- .a. Enter the description text for your model.
- .b. Enter the member name containing the AS/400 performance data.
- .c. Enter the name of the library containing the measured AS/400 performance data.

```
                  Create BEST/1 Model from Performance Data

Model . . . . . . . . . . . . . . . :   YOURMODEL

Type choices, press Enter.  Use *SLTHOUR to select an hour-long time period or
  use *SLTITV to select first and last interval of a one to two hour
  time period.  The time period selected should be representative of your peak
  processing activity.

Text . . . . . . . . . . . . . .   .a.

Performance data:
  Member . . . . . . . . . . . .   .b.   Name, F4 for list
  Library  . . . . . . . . . . .   .c.   Name

Start time . . . . . . . . . . .   *SLTHOUR     Time, *FIRST, *SLTHOUR, *SLTITV
Start date . . . . . . . . . . .   *FIRST       Date, *FIRST

Stop time  . . . . . . . . . . .   *LAST        Time, *LAST
Stop date  . . . . . . . . . . .   *LAST        Date, *LAST

F3=Exit   F4=Prompt   F12=Cancel
```

Note: Use the **Start/Stop** specifications to select the performance data you want to include in your model. Use the **Help** key to get assistance on the use of these fields.

4.4.6 Select Time Interval

If you select ***SLTITV** in the previous display, a display that allows you to select the start time and end time of the performance data to be included in the model is shown.

```
                           Select Time Interval

Library  . . . . . . . :   YOURLIB      Performance member . . :   Q972950956

Type option, press Enter.  Select first and last interval.
  1=Select

                              ---Transaction---   --CPU Util---   I/Os per SEC
Opt   Date       Time          Count   Rsp Time   Total   Inter   Sync   Async
      10/22/97   10:01:55         13         .5       4       0      6      10
      10/22/97   10:06:55         16         .1       1       0      5       4
      10/22/97   10:11:54          0         .0       2       0     19      35
      10/22/97   10:16:53          0         .0       2       0      9       7
      10/22/97   10:21:54          0         .0       1       0      5       6
      10/22/97   10:26:54          0         .0       1       0     10       6
      10/22/97   10:31:55          0         .0       1       0     11       9
      10/22/97   10:36:54          0         .0       1       0      5       4
      10/22/97   10:41:55         52         .0       1       0      4       2
      10/22/97   10:46:54         45         .1       1       0     10       5
      10/22/97   10:51:55         13         .0       1       0      4       4
                                                                       More...
F3=Exit   F12=Cancel   F15=Sort by interval        F16=Sort by count
F17=Sort by rsp time   F18=Sort by total CPU util   F19=Sort by total I/Os
```

Select the start and end times **only** by entering a **1**.

4.4.7 Classify Jobs into Workloads

The next display allows you to specify how you want to classify the resource usage into workloads. Select option 2 and press Enter.

```
                              Classify Jobs

Select one of the following:

     1. Use default job classification
     2. Classify jobs into workloads
     3. Use existing job classifications

Selection
   .2.

F3=Exit   F12=Cancel
```

4.4.8 Specify Job Classification Category

The following display allows you to select the job category you want to use in assigning the AS/400 server jobs to workloads. Select the category that can assist you in defining your workloads. For example, if each of the server workloads run in separate subsystems, you might select 6=Subsystem as the classification category.

```
                  Specify Job Classification Category

Type choice, press Enter.

  Category . . . . . . . . . . . . . _            1=User ID
                                                  2=Job type
                                                  3=Job name
                                                  4=Account code
                                                  5=Job number
                                                  6=Subsystem
                                                  7=Pool
                                                  8=Control unit
                                                  9=Comm line
                                                  10=Functional area

F3=Exit   F12=Cancel
```

4.4.9 Edit Job Classifications

The details of this display depend on your selection of the classification category. The following example is based on Job Number:

```
                          Edit Job Classifications

 Enter workload names and category values which are assigned to each workload,
   press Enter.  Jobs with unassigned values become part of workload QDEFAULT.

 Workload    Job Number     Workload    Job Number     Workload    Job Number
 ________    __________     ________    __________     ________    __________
 ________    __________     ________    __________     ________    __________
 ________    __________     ________    __________     ________    __________
 ________    __________     ________    __________     ________    __________
 ________    __________     ________    __________     ________    __________
 ________    __________     ________    __________     ________    __________
                                                                        More....
 F3=Exit   F9=Display values from data   F12=Cancel
 To display values from performance data, press F9
```

Press F9 to analyze your performance data and present the available values within the category selected.

Assign the AS/400 server jobs to as many workloads as you want to create within your model. If your measured data has interactive work in addition to server applications, you can define these as well.

Note:

We do not recommend explicitly assigning *LIC tasks to a specific workload. BEST/1 has internal algorithms to allocate *LIC tasks to the appropriate workload.

```
                          Assign Jobs to Workloads

 Workload . . . . . . . . . . . . . . . . . .

 Type options, press Enter.  Unassigned jobs become part of workload QDEF
   1=Assign to above workload   2=Unassign

                               Number of       CPU       I/O
 Opt  Workload    Job Number  Transactions    Seconds    Count
      DIALOG      008758            0           .000         0
      DIALOG      008759            0           .000         0
      DIALOG      008760            0           .000         0
      DIALOG      008761            0           .000         0
      UPDATE      008762            0         50.175      5351
      ENQUEUE     008763            0           .000         0
      BATCH       008764            0         36.958       470
      BATCH       008765            0           .000         0
      BATCH       008766            0           .000         0
      SPOOL       008767            0          7.549      2646
      UPDATE      008768            0         16.163       822

                                                                        More...
 F3=Exit   F12=Cancel      F15=Sort by workload      F16=Sort by job number
 F17=Sort by transactions   F18=Sort by CPU seconds   F19=Sort by I/O count
```

4.4.10 Specify Paging Behaviors

```
                          Specify Paging Behaviors

 Type choices, press Enter.

               Paging Behavior
 Workload       (F4 for list)

 QDEFAULT         *GENERIC
 SPOOL            *GENERIC
 BATCH            *GENERIC
 ENQUEUE          *GENERIC
 UPDATE           *GENERIC
 DIALOG           *GENERIC



 F3=Exit   F4=Prompt   F12=Cancel
```

Accept the default value of *GENERIC for all workloads.

4.4.11 Define Non-Interactive Transactions

The following display allows you to specify the number of Capacity Planning Units of Measure (CPUMs) for each server workload. Please refer to Section 4.2.3, "Select a Capacity Planning Unit of Measure" on page 36 for a discussion on the selection of CPUMs.

CPUMs per Hour

The value entered is **per hour** regardless of the time period over which the model was built.

Ensure that you change the "Type" field to ***NONE** for all the server workloads.

```
                    Define Non-Interactive Transactions

 Job classification category  . . . . . . :   Subsystem


 Type choices, press Enter.

               ---Activity Counted as Transaction---     Total Transactions
 Workload          Type                 Quantity          when Type = *NONE
 QDEFAULT          *LGLIO                 100.0                          0
 SPOOL            .*NONE.                 100.0                     .  320.
 BATCH            .*NONE.                 100.0                     .  500.
 UPDATE           .*NONE.                 100.0                     . 1220.
 DIALOG           .*NONE.                 100.0                     . 3875.

 Type:  *LGLIO, *CMNIO, *CPUSEC, *PRINT, *NONE

 F3=Exit   F12=Cancel
```

4.4.12 Save Job Classification Member

This allows you to save the job classifications as a member in a library.

```
                         Save Job Classification Member

 Change values if desired, press Enter.

   Member . . . . . . . . . . . . .     .a.    Name
     Library  . . . . . . . . . . .     .b.    Name

   Text . . . . . . . . . . . . . .     .c.

   Replace  . . . . . . . . . . . .   N             Y=Yes, N=No

 F12=Cancel
```

- .a. Enter the member name to contain the classification.
- .b. Check the library name for BEST/1 objects.
- .c. Enter description text for your classification.

4.4.13 Confirm Creation of BEST/1 Model

Check the model and library name and the associated text and press Enter to submit a batch job to create your BEST/1 model. You return to the **Work with BEST/1 Models** display.

```
                        Confirm Creation of BEST/1 Model

 Type choices, press Enter.

   Member . . . . . . . . . . . . .     .a.    Name
     Library  . . . . . . . . . . .     .b.    Name

   Text . . . . . . . . . . . . . .     .c.

   Replace  . . . . . . . . . . . .   N             Y=Yes, N=No

   Job name . . . . . . . . . . . .   CRTBESTMDL    Name, *JOBD
   Job description  . . . . . . . .   QPFRJOBD      Name, *NONE, *USRPRF
     Library  . . . . . . . . . . .     QPFR        Name, *LIBL, *CURLIB


 F12=Cancel
 Member .[redacted]. has been saved
```

- .a. Check the member name to contain the classification.
- .b. Check the library name for BEST/1 objects.
- .c. Check the description text for your classification.

You may issue the Work with Submitted Job (WRKSBMJOB) command to find out when your create model job completes, or issue DSPMSG and look for a *model completion* message, or repetitively use the F5 key (Refresh) until you see your new model name appear on the Work with BEST/1 Models menu. After you see your new model on this display, proceed to the next step.

4.5 Validate the BEST/1 Model

Prior to evaluating the effect of workload growth, review the model created by BEST/1 to ensure that it conforms to the measured workload and response times. This section discusses the process to achieve this.

Model Objectives

A BEST/1 model may be considered to be properly calibrated when the measured data values have response times within .5 seconds of the corresponding predicted data values, and the resource utilizations between the measured and predicted are within 20% of each other.

In server capacity planning, we can ignore any differences in communications resource utilization because we limit the modelling to the server only.

Note: The BEST/1 manual, *BEST/1 Capacity Planning Tool*, SC41-5341, discusses the adjustments required when modelling "server jobs."

We have to rely on our experience with BEST/1 and our understanding of the application and the OS/400 operating environment to effectively adjust or **calibrate** the model.

4.5.1 BEST/1 Workload Components

You may have to verify and modify the workloads if analysis shows that BEST/1's calculations for some important resource usages do not coincide with actual measurements. These include:

- Total CPU utilization
- Disk I/O activity
- Number of non-interactive transactions per hour

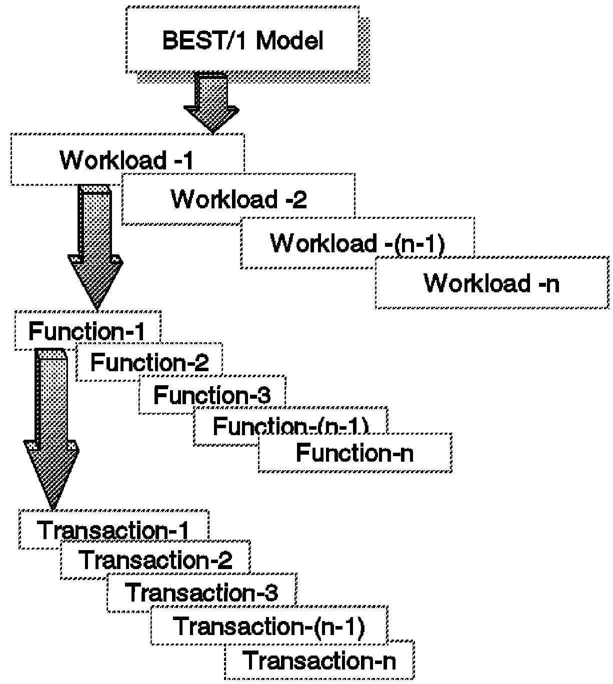


Figure 15. BEST/1 Workload Components

4.5.1.1 Workload

A BEST/1 model is made of one or more user-defined **workloads**, which is the main unit of input to capacity planning.

From the **Work with BEST/1 Model** display, select Option 1 (Work with workloads) to see all the workloads in a BEST/1 model.

```
                            Work with Workloads

Model/Text:   CENTRAL01   Central System

Type options, press Enter.
  1=Create   2=Change   3=Copy   4=Delete   5=Display   6=Print   7=Rename
  8=Save workload to workload member        9=Edit transactions

Opt     Workload       Text

        BATCH          Measured from YOURLIB (Q972950956)
        DIALOG         Measured from YOURLIB (Q972950956)
        ENQUEUE        Measured from YOURLIB (Q972950956)
        QCMN           Comm activity without jobs
        QDEFAULT       Measured from YOURLIB (Q972950956)
        SPOOL          Measured from YOURLIB (Q972950956)
        UPDATE         Measured from YOURLIB (Q972950956)
                                                                      Bottom
F3=Exit   F6=Add saved workload   F9=Add predefined workload   F12=Cancel
F13=Combine workloads
```

4.5.1.2 Function

Each workload is made up of one or more **functions**. Each defined workload created using performance data defaults to one function per user per hour. This function represents all the work done by this workload. Use option 5 to display the functions in a workload.

Note: Non-interactive workloads have a usage mode represented by "N/A".

```
                                Display Workload

 Workload  . . . . . . . :   DIALOG
 Workload text . . . . . :   Measured from YOURLIB (Q972950956)
 Workload type . . . . . :   *NORMAL
 CPU architecture  . . . :   *RISC
 Usage mode  . . . . . . :   4            1=Casual, 2=Interrupted, 3=Steady,
                                            4=N/A
 Type option, press Enter.
   5=Display
                                            Functions  Avg K/T     Inter Tns
 Opt  Function    Function Text              per User   (secs)   per Function
      DIALOG      Function of DIALOG             1.00      N/A            .00

                                                                          Bottom
 F3=Exit   F9=Display chars to comm line resources   F10=Display I/Os to ASPs
```

Two additional values are available at the **Function** level to assist you in calibrating a BEST/1 model. These are:

- Key/think time (not applicable for non-interactive work)
- Additional delays

These parameters can be modified by selecting the **Change Function** option.

```
                                Change Function
Workload  . . . . :   DIALOG      Measured from YOURLIB (Q972950956)
Function  . . . . :   DIALOG

Change fields, press Enter.
  Function text . . . . . . . . . .   Function of DIALOG
  Key/Think time  . . . . . . . . .   N/A            Seconds
  Additional delays . . . . . . . .   .0             Seconds

Transaction    Pool                 Transactions      CPU Time        Total
    Type        ID     Priority     per Function       (Secs)          I/Os
     2           2        19            3875.00         7.666           9.2
     2           1         0               1.00        75.499        7628.7

                                                                          Bottom
Transaction Type: 1=Interactive, 2=Non-interactive

F3=Exit   F6=Work with transactions   F12=Cancel
```

4.5.1.3 Transaction

Each function is made up of one or more **transactions**. Details of a function and the transactions included in the function can be displayed using option 5. The transactions in a server workload are non-interactive. The display shows the transactions per hour.

```
                                Display Function

 Workload  . . . . :   DIALOG      Measured from YOURLIB (Q972950956)
 Function  . . . . :   DIALOG      Function of DIALOG


 Type options, press Enter.
   5=Display

       Transaction   Pool                   Transactions     CPU Time       Total
 Opt       Type       ID    Priority    per Function      (Secs)          I/Os
            2          2       19           3875.00         7.666           9.2
            2          1        0              1.00        75.499        7628.7

                                                                          Bottom
Transaction Type: 1=Interactive, 2=Non-interactive
F3=Exit   F12=Cancel
```

The following display is an example of the information included in a transaction, which is obtained by selecting option 5.

```
                              Display Transaction

 Workload  . . . . :   DIALOG      Measured from YOURLIB (Q972950956)
 Function  . . . . :   DIALOG      Function of DIALOG

   Transaction Type . . . . . . . . :   2            1=Inter, 2=Non
   Pool ID  . . . . . . . . . . . . :   2
   CPU Priority . . . . . . . . . . :   19
   CPU time . . . . . . . . . . . . :   7.666        Secs (on B10)
   Permanent writes . . . . . . . . :   91.0         Percent
   Chars transferred in . . . . . . :   0
   Chars transferred out  . . . . . :   0
   Exceptional wait . . . . . . . . :   .0           Msec
   Paging behavior  . . . . . . . . :   *GENERIC

                                        Reads        Writes
   Sync DB I/Os . . . . . . . . . . :   1.6          .0
   Async DB I/Os  . . . . . . . . . :   .9           .7
   Sync non-DB I/Os . . . . . . . . :   2.9          .4
   Async non-DB I/Os  . . . . . . . :   .0           2.5
Press Enter to continue.

F3=Exit   F12=Cancel   F13=Display paging behavior
```

Adjustments

You can make adjustments to the model at any or all of the following components:

- Workload
- Function
- Transaction

4.5.2 Work with BEST/1 Models

On the Work with BEST/1 Models menu, select Option 5 (Work with) for your model.

```
                              Work with BEST/1 Models

 Library . . . . .   YOURLIB      Name

 Type options, press Enter.
   1=Create   3=Copy   4=Delete   5=Work with   6=Print   7=Rename

 Opt  Model        Text                                   Date      Time
 .5. .YOURMODEL.                            11/20/97  17:26:22
                                                                 Bottom
 Command
 ===>
 F3=Exit   F4=Prompt   F5=Refresh        F9=Retrieve   F12=Cancel
 F15=Sort by model     F16=Sort by text   F19=Sort by date and time
```

4.5.3 Analyze Current Model

On the Work with BEST/1 Model menu, select Option 5 (Analyze current model). This option runs the specified workload against the current configuration in the model.

```
                               Work with BEST/1 Model

 Performance data . . . :   YOURLIB (Q972950956)
 Model/Text . . . . . . :   CENTRAL01   Central System


 Select one of the following:

      1. Work with workloads
      2. Specify objectives and active jobs

      5. Analyze current model
      6. Analyze current model and give recommendations
      7. Specify workload growth and analyze model

     10. Configuration menu
     11. Work with results
                                                                      More...
 Selection or command
 ===> .5.
 F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel   F15=Save current model
 F17=Analyze using ANZBESTMDL        F22=Calibrate model   F24=More keys
```

4.5.4 Work with Results

Use the facilities within this function to evaluate the suitability of the BEST/1 model. If the values predicted by the model are substantially different from the values measured by the AS/400 performance monitor and the application profile measurements, you must make the necessary "artistic adjustments" to the model.

Calibration Expertise

This is where skill and experience play a major role.

```
                              Work with Results

Printed report text . . . . . .   Your Model

Type options, press Enter.
  5=Display   6=Print

Opt     Report Name
__      Measured and Predicted Comparison
__      Analysis Summary
__      Recommendations
__      Workload Report
__      ASP and Disk Arm Report
__      Disk IOP and Disk Arm Report
__      Main Storage Pool Report
__      Communications Resources Report
__      All of the above

F3=Exit   F12=Cancel   F14=Select saved results   F15=Save current result
F18=Graph current results   F19=Append saved results   F24=More keys
```

Some of the possible anomalies are discussed in the following section.

4.5.4.1 Workload Type *BATCHJOB

On the Work with Results menu, you may get a message "`current system CPU cannot handle the measured workload.`" This observation may not coincide with the analysis of AS/400 performance data.

On the Work with Results menu, select Option 5 against **Measured and Predicted Comparison**. This shows a column of measured statistics compared to the values evaluated by the BEST/1 model.

The following example shows the CPU as a highlighted field indicating CPU usage is in excess of 98%. Relatively high predictions for other system resources may also exist.

```
                       Compare Against Measured Values
                                         Measured     Predicted
Total CPU util . . . . . . . . . . . :        16.4         98.4
Disk IOP util  . . . . . . . . . . . :         2.9         16.4
Disk arm util  . . . . . . . . . . . :         2.2         13.2
Disk IOs per second  . . . . . . . . :        26.0        148.6
LAN IOP util . . . . . . . . . . . . :         4.8          1.8
LAN line util  . . . . . . . . . . . :          .8           .3
WAN IOP util . . . . . . . . . . . . :         1.2          5.3
WAN line util  . . . . . . . . . . . :          .0           .0
Interactive:
  CPU util . . . . . . . . . . . . . :         2.1          2.1
  Int rsp time (seconds) . . . . . . :          .1           .6
  Transactions per hour  . . . . . . :         763          763
Non-interactive thruput  . . . . . . :         867         9221

F3=Exit   F6=Print   F9=Work with spooled files   F12=Cancel
```

The BEST/1 user must understand the "server job" implementation included in performance monitor data and manually change the workload attribute from *BATCHJOB to *NORMAL. Making this change and other changes to the model are called *calibrating the model*.

It is important to understand how BEST/1 came up with the predicted CPU utilization of approximately 98%. This can affect the modelling of all "server jobs" because they are not perceived by the AS/400 performance measurements as interacting with any external agent and, hence, treated as a batch job.

Modelling "Batch Jobs"

BEST/1 assigns a workload type of *BATCHJOB when all jobs in the BEST/1 workload have met **all** of the following criteria:

- All jobs are non-interactive jobs.
- All jobs were active for at least 95% of the time the performance monitor collected data.
- All storage pools used by the jobs service only non-interactive jobs.

The *BATCHJOB workload type was designed to assist in modeling traditional batch job run-time modeling, which gives all additional CPU capabilities to a *BATCHJOB workload.

However, this algorithm has problems when the actual non-interactive environment contains "server type" jobs that periodically wait for work to do, similar to an interactive workstation job. Server jobs include an invoice print server that waits for notification of an order completion, an order entry client/server job, and so on that waits for information to be received from the client.

If the workload is identified by BEST/1 as *BATCHJOB, BEST/1 assigns all **available CPU** not used by other workloads to the server workloads, causing a high predicted CPU utilization.

4.5.4.2 Transactions per Function

Ensure that the number of transactions per function match the value input to the model.

Adjusting Transactions

BEST/1 always takes some of the unassigned OS/400 system work (pool 2) and Licensed Internal Code (LIC) work (pool 1) and "adds" a portion to all BEST/1 workloads because some system and LIC work was required to do things such as manage the job and perform disk and communications I/O.

For non-interactive workloads that are classified as *BATCHJOB, the number of active jobs is set to the number of jobs that contribute to the workload. This may not reflect the true profile of the workload.

If the workload is made up of non-interactive jobs but are **not** classified as *BATCHJOB, the number of active jobs is set to 1.

For example, you may have had six active jobs but the workload was really derived from only four.

You may have to examine the number of **Active Jobs** shown in the following display by selecting option 2 from the "Work with BEST/1 Model" menu:

```
                     Specify Objectives and Active Jobs

Model/Text:   Your Model
Type changes, press Enter.

                     Workload    Active  ----Interactive-----   Non-inter
Workload    Connect  Type         Jobs   Rsp Time     Thruput    Thruput

ODBCWL      *LOCAL   *BATCHJOB   ___6.0   ____.0        ____0     ______0
PFRMON      *LOCAL   *NORMAL     ___1.0   ____.0        ____0     ______0
QDEFAULT    *LOCAL   *NORMAL     ____.1   ____.0        ____0     _____0
QDEFAULT    *LAN     *NORMAL     ___1.0   ____.0        ____0     ______0


                                                                   Bottom
F3=Exit   F11=Show all quantities   F12=Cancel   F15=Sort by connect type
F19=Work with workloads
```

This difference can result in a erroneous reporting of the number of non-interactive transactions for the function.

4.5.4.3 Manual Calibration of Resource Usage

BEST/1 allows you to manually calibrate resource utilization such as CPU and disk by adjusting workload functions until the predicted utilizations are a close approximation to the measured values. Manual calibration requires a good understanding of the parameters you are changing.

Model Calibration

Use **F22** to enter manual calibration mode. While you are in manual calibration mode, any changes to the number of active jobs, or functions per user **do not affect total workload throughput ("transactions" per hour)**. This is because values for transactions per function are modified to offset changes to either active jobs or functions per user. In addition, all paging coefficients are recalculated during analysis.

After manual calibration, return to the **Work with BEST/1 Model** menu and press F22 to exit the manual calibration mode.

4.5.4.4 Calibrate Response Times

Server applications are reported as non-interactive jobs, and any external interfaces are not recognized. Thus, BEST/1 does not recognize the impact of any communications delays or overhead.

However, if you need to represent an "interactive response time" to a server application (because the user perceives performance from this viewpoint), you can adjust BEST/1's representation of response time by adjusting the **Additional delays** value in the workload specification.

The response time indication visible in the **Display Workload Report** display should be considered as representing **internal response time**.

For example, on the "Work with Results" menu, select option 5 to display the **Workload Report**. You can then display "response times details" using **F11**. Add an amount equal to the difference between the perceived user response time and the internal response time shown in the Workload Report.

```
                           Display Workload Report
Period:   Analysis
                             Total   --------Rsp Time Secs spent in--------
Workload    Type  Connect  Rsp Time    CPU     I/O    Pool    Comm   Other
QDEFAULT      1   *LOCAL        .2      .1      .0      .0      .0      .0
QDEFAULT      1   *LAN          .6      .1      .0      .0      .4      .0
DIALOG        2   *LOCAL        .9      .5      .4      .0      .0      .0
PFRMON        2   *LOCAL     211.6    27.4   183.6      .0      .0      .0
QDEFAULT      2   *LOCAL      14.0     2.9    11.1      .0      .0      .0
QDEFAULT      2   *LAN        14.0     2.9    11.1      .0      .0      .0

Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F11=Workload summary   F12=Cancel
```

Go to the Work with BEST/1 Model menu and select Option 1 (Work with workloads) followed by Option 2 (Change) for the workload. Then press F6 (Work with functions), select Option 2 (Change) for the workload, and increase the additional delay.

```
                               Change Function
Workload  . . . . :   DIALOG   Function of DIALOG
Function  . . . . :   DIALOG
Change fields, press Enter.
  Function text . . . . . . . . . .   Function of DIALOG
  Key/Think time  . . . . . . . . .   N/A             Seconds
  Additional delays . . . . . . . .    0.4            Seconds
Transaction    Pool                 Transactions      CPU Time         Total
    Type        ID     Priority     per Function       (Secs)           I/Os
     2           5        20              200.00         2.891          53.3
     2           2         0                 .25       127.521            .0
     2           1         0                 .25        58.105        8351.9

                                                                       Bottom
Transaction Type: 1=Interactive, 2=Non-interactive
F3=Exit   F6=Work with transactions   F12=Cancel
```

Exceptional Wait time can also be added at the **transaction** level of the workload.

```
                                  Display Transaction

 Workload  . . . . :   DIALOG      Measured from YOURLIB (Q972950956)
 Function  . . . . :   DIALOG      Function of DIALOG

   Transaction Type . . . . . . . . :   2           1=Inter, 2=Non
   Pool ID  . . . . . . . . . . . . :   2
   CPU Priority . . . . . . . . . . :   19
   CPU time . . . . . . . . . . . . :   7.666       Secs (on B10)
   Permanent writes . . . . . . . . :   91.0        Percent
   Chars transferred in . . . . . . :   0
   Chars transferred out  . . . . . :   0
   .Exceptional wait. . . . . . . . :   .0     Msec
   Paging behavior  . . . . . . . . :   *GENERIC

                                        Reads       Writes
   Sync DB I/Os . . . . . . . . . . :   1.6         .0
   Async DB I/Os  . . . . . . . . . :   .9          .7
   Sync non-DB I/Os . . . . . . . . :   2.9         .4
   Async non-DB I/Os  . . . . . . . :   .0          2.5
Press Enter to continue.

F3=Exit   F12=Cancel   F13=Display paging behavior
```

4.6 Saving the BEST/1 Model

When calibration of the BEST/1 model is complete, save the model and workloads so that we can use it later for modelling growth of this workload and for modelling it with other measured data workloads. This is especially important if considerable manual adjustments are made to the model.

1. Go back to the Work with BEST/1 Model menu and select Option 1 (Work with Workloads).
2. Select Option 8 (Save workload to workload member) for each of the workloads and save the workload.

```
                      Save Workload to Workload Member
Change values if desired, press Enter.
  Member . . . . . . . . . . . .  .a      Name
    Library  . . . . . . . . . .  .b      Name
  Text . . . . . . . . . . . . .  .c
  Replace  . . . . . . . . . . .  N             Y=Yes, N=No
  CPU architecture . . . . . . .  *RISC         *CISC, *RISC

F12=Cancel
```

- .a. Enter a member name to contain the workload.
- .b. Enter the library name for BEST/1 objects.
- .c. Enter a description text for your workload.

The workload member is saved to a file called QACYWKLS in your library. You can use this workload in any other system, or you can add this workload to any other model from any other system.

On the Work with BEST/1 Model menu, select F15 (Save the current model), enter a name (Member) of the model, and press Enter to save the model.

```
                              Save Current Model

Change values if desired, press Enter.

Save to Model member:
  Member . . . . . . . . . . . .  .a      Name
    Library  . . . . . . . . . .  .b      Name
  Text . . . . . . . . . . . . .  .c
  Replace  . . . . . . . . . . .  N              Y=Yes, N=No

Externally described member information:
  Save . . . . . . . . . . . . .  N              Y=Yes, N=No
  Member . . . . . . . . . . . .  *MEMBER        Name, *MEMBER
    Library  . . . . . . . . . .    *LIB         Name, *LIB
  Text . . . . . . . . . . . . .
  Replace  . . . . . . . . . . .  *REP           Y=Yes, N=No, *REP
```

- .a. Enter a member name to contain the model.
- .b. Enter the library name for BEST/1 object.
- .c. Enter a description text for your model.

Generating an "Externally described member" is for experienced BEST/1 users who want to save the model in a format that can be downloaded to a personal computer for later processing by user-written programs. Discussing these capabilities is beyond the scope of this redbook. So we select *N* for the "Save" option.

4.7 Plan for Growth

You can now make simulations of the effect of various hardware configurations on a selection of growth rates by typing your options into the model. Note that each of the individual workloads that you defined in the model can have separate growth rates.

Note: The growth projections presented by the BEST/1 model evaluate only the capacity of AS/400 configuration. Some of the benefits of operating system enhancements are also included on average. However, the model does not take into consideration any specific operating system code-path advantages or disadvantages that the application may encounter.

BEST/1 also cannot predict if the extrapolation of the measured workload results in growth beyond any functional or operational limitations of the application that may exist. These have to be managed separately.

4.7.1 Work with BEST/1 Models

From the main BEST/1 menu, select option 1 to display the "Work with BEST/1 models" menu. Select option 1 on the following menu to display the models in the selected library.

```
                              Work with BEST/1 Models

 Library . . . . .   YOURLIB      Name

 Type options, press Enter.
   1=Create   3=Copy   4=Delete   5=Work with   6=Print   7=Rename

 Opt  Model       Text                                      Date      Time

 .5.  CENTRAL01   Central System                            11/20/97  18:13:15
      CENTRAL02   Central System                            11/20/97  20:15:32
      APPLICN01   Application Server                        11/23/97  08:10:11
      DATABAS01   Database Server                           11/23/97  16:43:08

                                                                         Bottom


Command
===>
F3=Exit   F4=Prompt   F5=Refresh        F9=Retrieve   F12=Cancel
F15=Sort by model     F16=Sort by text   F19=Sort by date and time
```

4.7.2 Work with Selected BEST/1 Model

Select option 7 to specify the workload growth amounts and the number of periods you want to model growth.

```
                               Work with BEST/1 Model

 Performance data . . . :   YOURLIB (Q972950956)
 Model/Text . . . . . . :   CENTRAL01   Central System

 Select one of the following:

      1. Work with workloads
      2. Specify objectives and active jobs

      5. Analyze current model
      6. Analyze current model and give recommendations
      7. Specify workload growth and analyze model

     10. Configuration menu
     11. Work with results

 Selection or command
 ===> .7.

 F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel   F15=Save current model
 F17=Analyze using ANZBESTMDL        F22=Calibrate model   F24=More keys
 Model CENTRAL01 has been read
```

4.7.3 Specify Workload Growth and Analyze Model

This display allows you to grow all the workloads by a single percentage for up to 10 periods.

```
                      Specify Growth of Workload Activity

 Type information, press Enter to analyze model.
   Determine new configuration . . . . . . . . . . .  Y    Y=Yes, N=No
   Periods to analyze  . . . . . . . . . . . . . . .  1    1 - 10

   Period 1  . . . . . . . . . . . . .  Period 1       Name
   Period 2  . . . . . . . . . . . . .  Period 2       Name
   Period 3  . . . . . . . . . . . . .  Period 3       Name
   Period 4  . . . . . . . . . . . . .  Period 4       Name
   Period 5  . . . . . . . . . . . . .  Period 5       Name

             ------Percent Change in Workload Activity-------
 Workload    Period 1  Period 2  Period 3  Period 4  Period 5
 *ALL              .0      20.0      20.0      20.0      20.0

                                                                    Bottom
 F3=Exit   F11=Specify growth by workload   F12=Cancel
 F13=Display periods 6 to 10   F17=Analyze using ANZBESTMDL
```

Press **F11** to display the workloads in the model that allows you to specify growth rates for each workload.

```
                      Specify Growth of Workload Activity

 Type information, press Enter to analyze model.
   Determine new configuration . . . . . . . . . . .  Y.    Y=Yes, N=No
   Periods to analyze  . . . . . . . . . . . . . . .  1.  1 - 10

   Period 1  . . . . . . . . . . . . .  .Period 1.       Name
   Period 2  . . . . . . . . . . . . .  .Period 2.       Name
   Period 3  . . . . . . . . . . . . .  .Period 3.       Name
   Period 4  . . . . . . . . . . . . .  .Period 4.       Name
   Period 5  . . . . . . . . . . . . .  .Period 5.       Name

             ------Percent Change in Workload Activity-------
 Workload    Period 1  Period 2  Period 3  Period 4  Period 5
 BATCH             .0      20.0      20.0      20.0      20.0
 DIALOG            .0      15.0      20.0      25.0      30.0
 ENQUEUE           .0      15.0      20.0      25.0      30.0
 QCMN              .0      20.0      20.0      20.0      20.0
 QDEFAULT          .0      20.0      20.0      20.0      20.0
 SPOOL             .0      10.0      10.0      10.0      10.0
                                                                     More...
 F3=Exit   F11=Specify total growth   F12=Cancel   F13=Display periods 6 to 10
 F17=Analyze using ANZBESTMDL
```

You can either have BEST/1 determine the required upgrade system, or you can modify the configuration and review the effect on system resource usage as the workload increases.

Status messages are issued as growth analysis is performed. When analysis is complete, the following display is shown.

```
                              Work with Results

Printed report text . . . . . .   Central System


Type options, press Enter.
  5=Display   6=Print

Opt     Report Name
 _      Measured and Predicted Comparison
 _      Analysis Summary
 _      Recommendations
 _      Workload Report
 _      ASP and Disk Arm Report
 _      Disk IOP and Disk Arm Report
 _      Main Storage Pool Report
 _      Communications Resources Report
 _      All of the above

                                                                    Bottom
F3=Exit   F12=Cancel   F14=Select saved results   F15=Save current results
F18=Graph current results   F19=Append saved results   F24=More keys

Model has been analyzed
```

4.7.4 Automatic Upgrade to Advanced Servers

If your data was collected on an AS/400 **Advanced Series** and you require BEST/1 to select suitable systems from the **Advanced Servers**, perform the following steps:

1. On the BEST/1 for AS/400 display, take Option 60 (More BEST/1 options).
2. Select Option 10 (Hardware characteristics menu).
3. Select Option 1 (Work with CPU models).
4. Find the CPU model used in your measured data.
5. For this CPU and Model, select 2 (Change).

   Change the *Upgrade to family* parameter to ***POWERSRV** as shown in the following display.

```
                              Change CPU Model

CPU model  . . . . . . . . . . . . . . . . :   2143
Min/Max storage size (MB)  . . . . . . . . :   256    1024

Type information, press Enter.
  System unit . . . . . . . . . . . . .   9406         9402, 9404, 9406
  Architecture  . . . . . . . . . . . .   *RISC        *CISC, *RISC
  Relative performance (B10 = 1.0):
    Normal  . . . . . . . . . . . . . .   19.70
    Server  . . . . . . . . . . . . . .                (Blank if not Server)
  Number of processors  . . . . . . . .   1
  Currently available . . . . . . . . .   Y            Y=Yes, N=No
  Family  . . . . . . . . . . . . . . .   *POWERAS     Name
  Upgrade to family . . . . . . . . . .   *POWERSRV      *NONE, name


                                                 Minimum      Maximum
  Disk IOPs  . . . . . . . . . . . . . . . . .   0            16
  Multifunction IOPs . . . . . . . . . . . . .   1            1
                                                                    More...
F3=Exit   F6=Specify storage sizes      F9=Specify connections to disk IOPs
F11=Specify connections to disk drives   F12=Cancel   F24=More keys
```

6. Press Enter.
7. Return to the BEST/1 for the AS/400 display.

4.7.4.1 High Priority Threshold

BEST/1 defaults to making its CPU utilization recommendations and conclusions based on CPU utilization of priority 20 or higher jobs. Many server jobs such as QZDAINIT run at priority 20 by default.

If your server jobs run at lower priorities (value greater than 20) and you want these to be included in values used by BEST/1, you can modify the guidelines and thresholds as follows:

1. Select Option 60 for **More BEST/1 Options**.
2. Select Option 11 for **Analysis parameter menu**.
3. Select Option 1 for **Edit utilization guidelines**.
4. The following display is shown:

```
                         Edit Utilization Guidelines

 Type changes, press Enter.  Values are used for Analysis recommendations

 Max Hardware Util  . . . . . . . . . . .   *GUIDE    *GUIDE, *THRESH
 Priority for CPU util  . . . . . . . . .   20


                                            Guideline        Threshold
 CPU util . . . . . . . . . . . . . . . .   70               80
 Disk IOP util  . . . . . . . . . . . . .   70               80
 Disk arm util  . . . . . . . . . . . . .   40               50
 Communications IOP util  . . . . . . . .   45               50
 Communications line util . . . . . . . .   35               40
 Local WS controller util . . . . . . . .   35               40
 LAN controller util  . . . . . . . . . .   40               50
 WAN WS controller util . . . . . . . . .   40               50


                                                                      Bottom
F3=Exit   F12=Cancel
```

You can change the required guideline and threshold values on this display.

The following table shows the guidelines for CPU utilization for various n-way processors of the AS/400 system:

	CPU Utilization Guideline
1-way	70%
2-way	76%
4-way	81%
8-way	89%
12-way	92%

Chapter 5. Example for Application Using Sockets Interface

This chapter uses performance data collected when running the Lawson Software client/server accounting and payroll applications in the Customer Benchmark Center in IBM Rochester in November, 1997. The Lawson applications were volume tested on two models of the AS/400 system:

- Model 40S (CPW rating of 27.0/9.4) up to 150 client/server users
- Model 650 (CPW rating of 1,794.0) up to 2500 client/server users

The chapter covers the following capacity planning aspects for both of the AS/400 models used in the testing:

- Generating and collecting performance data
- Model creation using user-defined job classification
- Growth analysis
- Comparison of predicted capacity requirement with actuals

Varying levels of client/server workload were generated using the Mercury Interactive volume testing tool, LoadRunner. Five Mercury scripts were created to represent five typical business transactions:

- RQ10 - Requisition Entry
- AP20 - Accounts Payable Invoice Entry
- PR51 - Pay Check Stub Inquiry
- PR13 - Employee Taxes
- GL95 - General Ledger Account Analysis

5.1 Lawson Software Client/Server Architecture and Volume Test Setup

Lawson's client/server applications run over a TCP/IP network by using the sockets interface. If the client PC can successfully "ping" an AS/400 system, all of the prerequisite hardware and software for running a Lawson client/server application are present.

5.1.1 Lawson Software Client/Server Architecture

The Lawson client software runs on all versions of Microsoft Windows. Since the client software only performs presentation functions for the Lawson GUI (not the application business rules), the Lawson applications have a "thin client" architecture. This architecture minimizes the load on the PC and the load on the network. Stored procedures on the AS/400 system handle the application business rules and the database access. The application business rules are contained in RPG programs that run on the AS/400 system. These programs access data in the DB/400 database directly using native RPG operation codes for both batch and online processes. Obviously batch application processes such as the payroll calculation phase run entirely in the AS/400 system without any involvement of the network or PCs.

In the Lawson client/server applications, the TCP/IP sockets connection to the AS/400 system is done through a user-defined port number. The Lawson client/server setup program establishes this port number and corresponding subsystem based on user input parameters. There is a Lawson written TCP/IP service program "listening" for connections on this port. The TCP/IP service program creates a Lawson application server job for each connected user that

© Copyright IBM Corp. 1998

runs as a "Batch Immediate" job. This architecture runs online jobs efficiently on both server and system models of the AS/400 system.

5.1.2 Lawson Sockets Volume Test Setup

For the purposes of the volume testing, it was decided to use the Lawson client/server setup program to establish five port numbers attached to five subsystems. This arrangement allowed each of the however many interactive jobs for each of the five client/server business transactions to run in one of five specially created AS/400 subsystems. On the AS/400 model 650, work management entries were, in turn, used to arrange for all of the jobs from these five subsystems to run in the shared main storage pool *SHRPOOL1 (this pool had no other jobs running in it). On the model 40S, the five subsystems directed their work into *BASE because main storage was much more constrained.

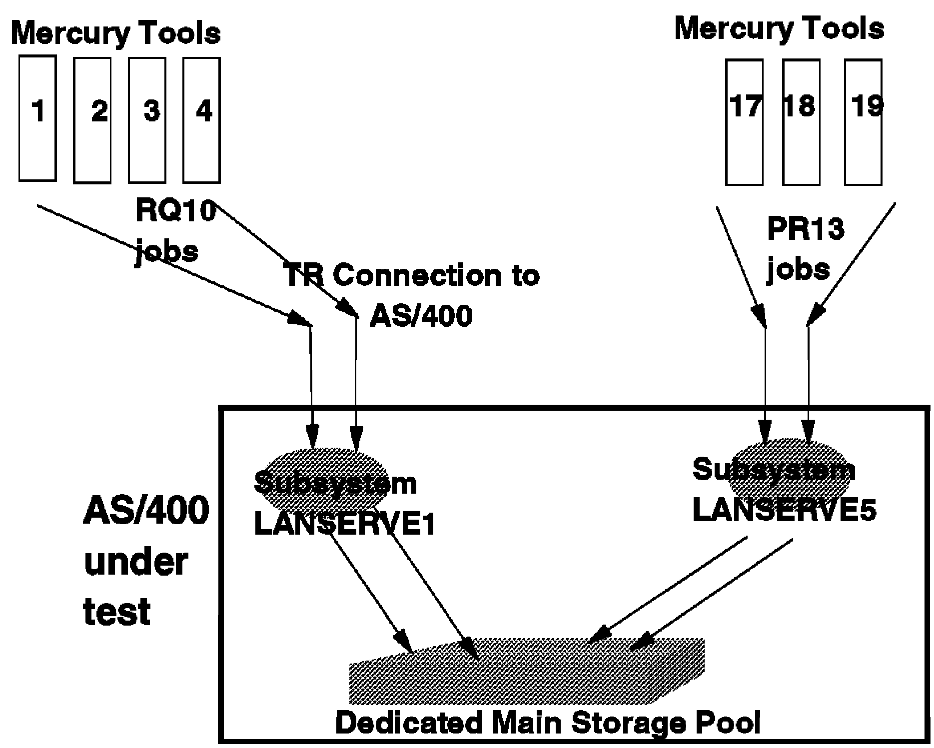


Figure 16. Setup for Volume Tests

5.2 Volume Testing on the AS/400 Model 40S

A series of volume tests (VTs) were run with the Mercury tools connected to the three different AS/400 models.

5.2.1 Generating and Collecting Performance Data on Model 40S

Up to 19 Mercury tools were connected through token-ring to the AS/400 systems with each tool capable of hosting up to 167 independent client/server jobs. The number of client/server users per each business transaction was kept as equal as possible. So when there were a total of 25 client/server users active, there were five users active for each of the five scripts. However, the iteration rates of the scripts running on the tools was different to reflect the relative usage that real users make to these business transactions. A base iteration rate was determined for each of the scripts based on the experience of Lawson technical support consultants. These base iteration rates are as shown in the following table:

Table 1. Model 40S Base Iteration Rates by Script					
Base Iteration Rate of Each Job	**RQ10 Script**	**AP20 Script**	**PR51 Script**	**PR13 Script**	**GL95 Script**
Range (secs)	180-450	90-270	360-1080	60-180	360-1080
Average (secs)	315	180	720	120	720
Base BTs / hour / user	11.43	20	5	30	5

5.2.1.1 Volume Testing of Model 40S

The scripts were run at the base iteration rate with equal numbers of users on each script. The number of active client/server users was gradually increased with periods of time when the number of active users was held steady to allow Performance Tools/400 to collect at least two 5-minute intervals of data at that number of active users. For the model 40S, there were four such time intervals when the number of active client/server users was held steady. Performance Tools/400 System Reports were printed for each of these periods and the following table was created:

Table 2. Model 40S Workload Build Up		
Time Interval	**Total Number of Client/Server Users**	**Total CPU Utilization % from System Report**
14:44-14:54	25	12.8%
15:04-15:14	50	29.5%
15:44-15:54	100	56.1%
16:19-16:29	150	88.6%

As was previously stated, the client/server workload was generated by Mercury Interactive LoadRunner scripts running on PCs attached to the AS/400 system by token-ring. Within LoadRunner, there is a facility to control the rate at which the business transaction scripts iterate. This iteration rate can be set as one iteration every so many seconds (that is, one client/server BT every so many seconds). The number of seconds for a BT is a random number selected by the tool from between a specified range. The idea behind this facility is to introduce a level of variability into the workload, even when the number of active users is kept constant. This is, of course, what we find in workloads generated by real users. The base iteration rates used on the model 40S VT and the number of client/server business transactions (BTs) resulting from this setting for the various total numbers of active users is shown in the following table:

Table 3 (Page 1 of 2). Base Rate of Business Transactions (BTs) by Script						
Total Number of Users	**RQ10 Script**	**AP20 Script**	**PR51 Script**	**PR13 Script**	**GL95 Script**	**Total All Scripts**
25 users	57.14	100	25	150	25	375.14
50 users	114.28	200	50	300	50	714.28
100 users	228.57	400	100	600	100	1428.57

Table 3 (Page 2 of 2). Base Rate of Business Transactions (BTs) by Script						
Total Number of Users	**RQ10 Script**	**AP20 Script**	**PR51 Script**	**PR13 Script**	**GL95 Script**	**Total All Scripts**
150 users	342.86	600	150	900	150	2142.86

5.2.2 Best/1 Model Creation Using Data from Model 40S Volume Test

All of the Best/1 modelling was done at the *ADVANCED level. We were, therefore, careful to always start the modelling by using the STRBEST command and prompting on F4 to ensure that the tool did **not** invoke the *BASIC option. *We also strongly advise that once Best/1 has been started, option 1 from the following display should be used*. Option 5 should be avoided, as at some PTF levels, Best/1 drops the user back to *BASIC if option 5 is used. The problem with the *BASIC mode is that it is not possible to specify job classifications and use many of the other facilities described later.

```
                                BEST/1 for the AS/400

Select one of the following:

     1. Work with BEST/1 models

     5. Create BEST/1 model from performance data

    10. Work with results

    50. About BEST/1

    60. More BEST/1 options




Selection or command
===>    1

F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
```

*Figure 17. Ensuring *ADVANCED Mode Retained*

A Best/1 model was created based on the performance data from the time interval when the number of active client/server users was kept at a total of 25 (that is, five on each script). In the preceding table, we calculated the number of non-interactive transactions per hour for each of the scripts. Each script's jobs were running in a particular subsystem and the following table shows the mapping between scripts, subsystems, line IDs, socket addresses, and Mercury tools.

Table 4. Mapping of Scripts to Subsystems for 40S and S20 VTs				
Scripts	**Subsystems**	**Line IDs**	**Socket Addresses**	**Mercury Hosts**
RQ10	LANSERVE1	PIDLAN	7011	1 to 4
AP20	LANSERVE2	PIDLAN1	7012	5 to 8
PR51	LANSERVE3	PIDLAN2	7013	9 to 12
GL95	LANSERVE4	PIDLAN3	7014	13 to 16
PR13	LANSERVE5	PIDLAN4	7015	17 to 19

We were able to assign the Lawson subsystems to Best/1 workloads, which, as you see in the following display, were assigned the same name as the scripts that generated the client/server transactions:

```
                              Assign Jobs to Workloads

 Workload . . . . . . . . . . . . . . . . .

 Type options, press Enter.  Unassigned jobs become part of workload QDEF
AULT.
   1=Assign to above workload   2=Unassign

                                Number of       CPU        I/O
 Opt  Workload     Subsystem   Transactions    Seconds     Count
                                     0          3.636      1317
                   JOBSCHED          0           .000         0
                   LASERVER          0           .000         0
      RQ10         LASERVE1          0          6.657       296
      AP20         LASERVE2          0           .543        43
      PR51         LASERVE3          0         12.904      1388
      GL95         LASERVE4          0         11.698      1062
      PR13         LASERS5          0         33.641       853
                   QBATCH            0           .000         0
                   QCMN              0           .008         9
      QPFRMON      QCTL              4          2.854       364
                                                                   More...
F3=Exit   F12=Cancel       F15=Sort by workload     F16=Sort by subsystem
F17=Sort by transactions   F18=Sort by CPU seconds   F19=Sort by I/O count
```

Figure 18. Assigning Subsystems to Best/1 Workloads

We were also able to specify the number of non-interactive transactions per hour in the subsequent "Define Non-Interactive Transactions" display shown in Figure 19 on page 70.

```
                       Define Non-Interactive Transactions

 Job classification category  . . . . . . :   Subsystem


 Type choices, press Enter.

               ---Activity Counted as Transaction---     Total Transaction
 Workload          Type                  Quantity         when Type = *NONE
 QDEFAULT          *LGLIO                  100.0                    0
 GL95              *NONE                   100.0             25
 PR51              *NONE                   100.0             25
 AP20              *NONE                   100.0             100
 RQ10              *NONE                   100.0             57
 QPFRMON           *LGLIO                  100.0                    0
 PR13              *NONE                   100.0             150




 Type:  *LGLIO, *CMNIO, *CPUSEC, *PRINT, *NONE


 F3=Exit   F12=Cancel
```

Figure 19. Define Non-Interactive Transactions

The options were now taken to submit the model for creation. This resulted in the following display, which shows that we produced a fairly "good" model (that is, a model that appears to predict the system utilizations reasonably well).

```
                        Measured and Predicted Comparison

                                         Measured     Predicted
 Total CPU util . . . . . . . . . . . :        12.9         12.9
 Disk IOP util  . . . . . . . . . . . :          .9          1.2
 Disk arm util  . . . . . . . . . . . :         1.6          1.6
 Disk IOs per second  . . . . . . . . :         9.3          9.3

 LAN IOP util . . . . . . . . . . . . :         4.3           .0
 LAN line util  . . . . . . . . . . . :          .0           .0
 WAN IOP util . . . . . . . . . . . . :         1.8          1.7
 WAN line util  . . . . . . . . . . . :          .0           .0

 Interactive:
   CPU util . . . . . . . . . . . . . :          .1           .1
   Int rsp time (seconds) . . . . . . :          .1           .2
   Transactions per hour  . . . . . . :          24           24
 Non-interactive thruput  . . . . . . :         545          603


 Performance estimates -- Press help to see disclaimer.
```

Figure 20. Measured and Predicted Comparison

It is interesting to note that "no calibration" was required in the creation of this model as Best/1 had correctly assigned our "Batch Immediate" jobs supporting the client/server work as type *NORMAL. The model had also estimated the response times for the client/server jobs as shown in Figure 21 on page 71.

```
                              Display Workload Report

 Period:   Analysis
                     CPU     Thruput  -------Response Times (Secs)------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QPFRMON       1       .1          24        .2        .2       .0       .0
AP20          2       .1         100        .1        .1       .0       .0
GL95          2      2.0          25       6.6       6.6       .0       .0
PR13          2      5.7         150       2.1       2.1       .0       .0
PR51          2      2.2          25       6.9       6.9       .0       .0
QDEFAULT      2      1.2         120        .4        .4       .0       .0
QPFRMON       2       .5           0        .6        .6       .0       .0
RQ10          2      1.1          57       1.2       1.2       .0       .0







 Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
 Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-analyze   F11=Response time detail      F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

Figure 21. Display Workload Report

5.2.3 Growth Analysis for AS/400 Model 40S

Having created what appears to be a fairly good model, it is now time to exercise the model. This is done by asking the model to forecast the system resource required to support the larger numbers of users that we already had system reports for. We chose the option to apply growth to the model and entered the growth parameters shown in the following figure:

```
                     Specify Growth of Workload Activity

Type information, press Enter to analyze model.
  Determine new configuration . . . . . . . . . .   N    Y=Yes, N=No
  Periods to analyze  . . . . . . . . . . . . . .   4    1 - 10

  Period 1  . . . . . . . . . . . . .  25 users      Name
  Period 2  . . . . . . . . . . . . .  50 users      Name
  Period 3  . . . . . . . . . . . . .  100 user      Name
  Period 4  . . . . . . . . . . . . .  150 user      Name
  Period 5  . . . . . . . . . . . . .  Period 5      Name

            ------Percent Change in Workload Activity-------
Workload    Period 1  Period 2  Period 3  Period 4  Period 5
*ALL             .0    100       100       50         20.0




F3=Exit   F11=Specify growth by workload   F12=Cancel
F13=Display periods 6 to 10   F17=Analyze using ANZBESTMDL
```

Figure 22. Specify Growth of Workload Activity

The growth was applied to all the workloads, even though you can use F11 to apply different levels of growth to different workloads.

After a short pause, Best/1 produced new output, forecasting the levels of utilization across the system at the specified increased levels of workload. The "Analysis Summary" is shown in the following display:

```
                           Display Analysis Summary

                        Stor     CPU  -Disk IOPs--  -Disk Ctls--  -Disk Arms--
Period    CPU Model    (MB)    Util   Nbr    Util   Nbr    Util   Nbr     Util
25 users  40S 2109       96    12.9     2     1.2     5      .0     5      1.6
50 users  40S 2109       96    25.7     2     2.3     5      .0     5      3.2
100 user  40S 2109       96    52.1     2     5.9     5      .1     5      8.2
150 user  40S 2109       96    82.3     2    17.4     5      .3     5     24.1



          ----Inter Rsp Time----  -------Inter--------  -----Non-Inter------
Period     Local    LAN     WAN   CPU Util    Trans/Hr  CPU Util    Trans/Hr
25 users      .2      .0      .0        .1          24      12.8       603
50 users      .2      .0      .0        .2          48      25.6      1206
100 user      .2      .0      .0        .4          96      51.7      2412
150 user      .3      .0      .0        .6         144      81.7      3618





F3=Exit   F10=Re-analyze   F11=Alternative view         F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

Figure 23. Display Analysis Summary

The summary shows that the workload generated by 150 client/server users drives the CPU in the model 40S to well above the guideline figure for "good performance" on a system with one main processor, which is, of course, 70%. The following "Workload Report" shows that Best/1 forecasts the response time of the client/server jobs will degrade significantly, which we expect when the CPU utilization is so high.

```
                              Display Workload Report

Period:   25 users
                    CPU     Thruput  -------Response Times (Secs)------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QPFRMON       1       .1        24        .2        .2       .0       .0
AP20          2       .1       100        .1        .1       .0       .0
GL95          2      2.0        25       6.6       6.6       .0       .0
PR13          2      5.7       150       2.1       2.1       .0       .0
PR51          2      2.2        25       6.9       6.9       .0       .0
QDEFAULT      2      1.2       120        .4        .4       .0       .0
QPFRMON       2       .5         0        .6        .6
RQ10          2      1.1        57       1.2       1.2       .0       .0

-------------------------------------------------------------------------
Period:   50 users
                    CPU     Thruput  -------Response Times (Secs)------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QPFRMON       1       .2        48        .2        .2       .0       .0
AP20          2       .2       200        .1        .1       .0       .0
GL95          2      4.0        50       7.3       7.3       .0       .0
PR13          2     11.4       300       2.4       2.4       .0       .0
PR51          2      4.4        50       7.1       7.1       .0       .0
QDEFAULT      2      2.3       240        .5        .5       .0       .0
QPFRMON       2      1.0         0        .6        .6       .0       .0
RQ10          2      2.3       114       1.4       1.4       .0       .0

-------------------------------------------------------------------------
Period:   100 user
                   CPU     Thruput  -------Response Times (Secs)------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QPFRMON       1       .4        96        .2        .2       .0       .0
AP20          2       .4       400        .1        .1       .0       .0
GL95          2      8.2       100      11.4      11.4       .0       .0
PR13          2     22.9       600       3.9       3.9       .0       .0
PR51          2      9.0       100       8.8       8.8       .0       .0
QDEFAULT      2      4.7       480        .7        .7       .0       .0
QPFRMON       2      2.0         0        .7        .7       .0       .0
RQ10          2      4.6       228       2.3       2.3       .0       .0

-------------------------------------------------------------------------
Period:   150 user
                   CPU     Thruput  -------Response Times (Secs)-------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QPFRMON       1       .6       144        .3        .3       .0       .0
AP20          2       .6       600        .4        .4       .0       .0
GL95          2     13.6       150      33.2      33.2       .0       .0
PR13          2     35.3       900      11.2      11.2       .0       .0
PR51          2     14.6       150      16.7      16.7       .0       .0
QDEFAULT      2      7.5       720       2.4       2.4       .0       .0
QPFRMON       2      3.0         0        .8        .8       .0       .0
RQ10          2      7.2       342       6.5       6.5       .0       .0

 Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
 Performance estimates -- Press help to see disclaimer.
```

Figure 24. Display Workload Report

5.2.4 Comparison of Predicted Capacity Requirement with Actuals for 40S

From the preceding output, the following simple spread sheet and graph were created. The spread sheet shows that Best/1 consistently under estimated the total CPU required to support the increased levels of the client/server workload.

Period	Time Interval	Total Number of client/server users	Total CPU % from System Report	Total CPU % from Best/1
P1	14:44-14:54	25	12.8	12.9
P2	15:04-15:14	50	29.5	25.7
P3	15:39-15:49	100	56.1	52.1
P4	16:19-16:29	150	88.6	82.3

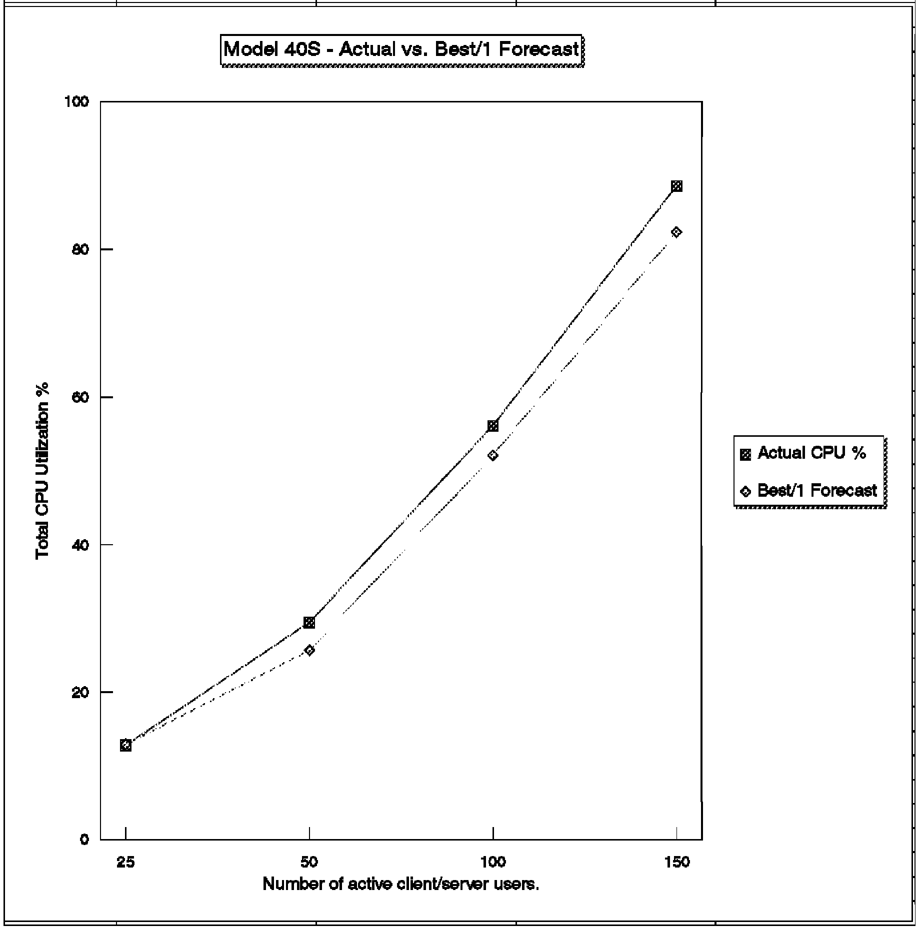


Figure 25. Spread Sheet and Graphs of Actual CPU versus Best/1 Forecast for AS/400 Model 40S

One of the reasons that Best/1 under estimates the utilization levels is that there is no mechanism within the Best/1 tool to apply an additional workload for higher levels of resource contention. Clearly, when we increase the workload, we cause a corresponding increase in file locking and so on, and this leads to the higher levels of utilization that Best/1 does not allow for.

5.3 Volume Testing of Model 650

The Lawson Software and data libraries were loaded onto an AS/400 Model 650 and the same Mercury Interactive LoadRunner scripts were again used to generate a client/server workload. As shown in the following figures, the initial client/server workload made little impact on such a large system. This was deliberate because part of the testing not covered in this document was to test the client/server workload with heavy concurrent batch work. A "snapshot" of the system with this heavy concurrent batch work is shown in the following display:

```
                              Display by Subsystem

Member . . . . . . . :   L1028R2        Elapsed time . . . . :   00:09:
  Library  . . . . . :     QPFRLAW1

Type options, press Enter.  Press F6 to display all jobs.
  5=Display jobs

                           CPU         Job         Tns      Average      Disk
Option    Subsystem       Util       Count       Count     Response       I/O
          *MACHINE         .83          46           0          .00      4669
          LASERVER         .00           0           0          .00         0
          LASERVE1        1.35         388           0          .00      4448
          LASERVE2         .21         383           0          .00         2
          LASERVE3        2.46         223           0          .00     53802
          LASERVE4        1.56         317           0          .00      7337
          LASERVE5        8.76         385           0          .00     24335
          LAWSON           .00           0           0          .00         0
  5       QBATCH         83.46           8           0          .00    114071
          QCMN             .00           0           0          .00         0
                                                                      More...
F3=Exit   F6=Display all jobs   F12=Cancel   F14=Display by job type
F15=Display by interval
```

Figure 26. Display by Subsystem

This batch workload had been especially adapted to make use of the eight CPUs in the model 650. This can be seen on the following display, which shows all eight jobs running in QBATCH and taking almost equal amounts of CPU:

```
                                    Display Jobs

Subsystem . . . . :   QBATCH           Member  . . . . . :   L1028R2
Elapsed time  . . :   00:09:59           Library . . . . :     QPFRLAW1

Type options, press Enter.
  5=Display job detail

                                            Job      CPU      Tns     Avg   Disk
Option  Job         User        Number  Type    Util     Count    Rsp    I/O
        PR140       LAIDEMO     024186  BCH    10.54         0     .0  14244
        PR140       LAIDEMO     024191  BCH    10.48         0     .0  14536
        PR140       LAIDEMO     024190  BCH    10.47         0     .0  14735
        PR140       LAIDEMO     024174  BCH    10.45         0     .0  15565
        PR140       LAIDEMO     024205  BCH    10.42         0     .0  13576
        PR140       LAIDEMO     024196  BCH    10.39         0     .0  13955
        PR140       LAIDEMO     024182  BCH    10.34         0     .0  14313
        PR140       LAIDEMO     024211  BCH    10.33         0     .0  13147

F3=Exit          F12=Cancel   F15=Sort by job   F16=Sort by job type   Bottom
F19=Sort by CPU   F24=More keys
```

Figure 27. Display Jobs

5.3.1 Generating and Collecting Performance Data for Model 650

The same script iteration rates were used but the number of users was significantly increased. The number of users was increased but held level at three points, and these points and the resulting total CPU utilization are shown in the following table:

Table 5. Model 650 Workload Build Up

Time Interval	Total Number of Client/Server Users	Total CPU Utilization % from System Report
10:00-10:10	1,000	8.2%
10:45-10:55	1,500	12.1%
12:10-12:20	2,000	15.0%

Using the script iteration rates shown in the section that described the Model 40S volume test resulted in the transaction rates per script and in total shown in the following table:

Table 6. Base Rate of Business Transactions by Script for Model 650

Total Number of Users *	RQ10 Script	AP20 Script	PR51 Script	PR13 Script	GL95 Script	Total BTs/Hour for All Scripts
1 user	11.43	20	5	5	30	
1000 users	2286	4000	1000	1000	6000	14286
1500 users	3429	6000	1500	1500	9000	21429
2000 users	4527	8000	2000	2000	12000	28572

* There were always the same number of users per script, so when the total number of users was 1000, the number of users per script was 200 and so on.

5.3.2 Model Creation Using Data from Model 650 Volume Test

A Best/1 model was created from the performance data time intervals at the time that the number of users was held at 1000 (see Figure 28).

```
                              Select Time Interval

 Library  . . . . . . . . :   QPFRLAW1      Performance member . . :   L1028

 Type option, press Enter.  Select first and last interval.
   1=Select

                             ---Transaction---   --CPU Util---   I/Os per
 Opt   Date        Time        Count    Rsp Time   Total    Inter   Sync   Async
       10/28/97   09:40:03       33        1.6      17        0     484      86
       10/28/97   09:45:02       17        1.4      10        0     283      36
       10/28/97   09:50:02        3        1.0       7        0     113      16
       10/28/97   09:55:03        6        1.0       9        0     148      21
       10/28/97   10:00:02       48         .7       9        0      98      13
  .1.  10/28/97   10:05:02        6         .8       8        0      98      16
  .1.  10/28/97   10:10:02       19         .6       8        0      76      11
       10/28/97   10:15:02       10         .6      14        0     270      45
       10/28/97   10:20:02       17        1.6       9        0      77       8
       10/28/97   10:25:03       38        1.6      10        0      88       9
       10/28/97   10:30:03        0         .0      10        0      94      10
                                                                          More...
```

Figure 28. Select Time Interval

Again, the mapping of client/server jobs to specific subsystems was used to build the model by using a Best/1 job classification based on the subsystem (see Figure 29 on page 79).

```
                      Assign Jobs to Workloads

Workload . . . . . . . . . . . . . . . . .

Type options, press Enter.  Unassigned jobs become part of workload QDEFAULT
  1=Assign to above workload   2=Unassign

                               Number of       CPU       I/O
Opt  Workload    Subsystem   Transactions    Seconds     Count
                                       0      10.712      4181
                 LASERVER              0        .000         0
     RQ10        LASERVE1              0      34.561      2430
     AP20        LASERVE2              0       5.555         0
     PR51        LASERVE3              0      85.443     41654
     GL95        LASERVE4              0      34.526      3406
     PR13        LASERVE5              0     213.916     13059
                 LAWSON                0        .000         0
                 QBATCH                0        .000         0
                 QCMN                  0        .000         0
     QPFRMON     QCTL                  0       1.798       338
     OPERATIONS  QINTER               25       1.801       489
                 QSERVER               0        .000         0
                 QSPL                  0        .000         0
                 QSYSWRK               0        .024         0

F3=Exit   F12=Cancel      F15=Sort by workload      F16=Sort by subsystem
F17=Sort by transactions   F18=Sort by CPU seconds   F19=Sort by I/O count
```

Figure 29. Assign Jobs to Workloads

On the model 650, we carefully assigned the client/server subsystems to an exclusive *SHRPOOL to allow us to study their behavior more closely. Using the calculated figures for each script of client/server transactions per hour, we were able to define non-interactive transactions for each of the five client/server workloads shown in Figure 30.

```
                Define Non-Interactive Transactions

Job classification category  . . . . . . :   Subsystem


Type choices, press Enter.

               ---Activity Counted as Transaction---     Total Transactions
Workload           Type                  Quantity         when Type = *NONE
QDEFAULT           *LGLIO                  100.0
OPERATIONS         *LGLIO                  100.0
QPFRMON            *LGLIO                  100.0
PR13               *NONE                   100.0              6000
GL95               *NONE                   100.0              1000
PR51               *NONE                   100.0              1000
AP20               *NONE                   100.0              4000
RQ10               *NONE                   100.0              2286


                                                                     Bottom
Type:  *LGLIO, *CMNIO, *CPUSEC, *PRINT, *NONE

F3=Exit   F12=Cancel
```

Figure 30. Define Non-Interactive Transactions

The model was then created based on the performance time intervals for the 1000 active client/server users. The following display shows that there was a reasonable agreement between the model and the measured values:

```
                  Measured and Predicted Comparison

                                       Measured     Predicted
Total CPU util . . . . . . . . . . . :        8.3          8.3
Disk IOP util  . . . . . . . . . . . :        3.9          1.5
Disk arm util  . . . . . . . . . . . :         .8          1.0
Disk IOs per second  . . . . . . . . :      152.1        162.0

LAN IOP util . . . . . . . . . . . . :        3.8           .0
LAN line util  . . . . . . . . . . . :         .0           .0
WAN IOP util . . . . . . . . . . . . :        4.0          1.4
WAN line util  . . . . . . . . . . . :         .0           .0

Interactive:
  CPU util . . . . . . . . . . . . . :         .0           .0
  Int rsp time (seconds) . . . . . . :         .7           .2
  Transactions per hour  . . . . . . :        150          150
Non-interactive thruput  . . . . . . :      14419        14474


Performance estimates -- Press help to see disclaimer.
```

Figure 31. Measured and Predicted Comparison

Again, no "calibration" was required and the following Workload Report gave Best/1's estimates of the client/server job's AS/400 server response times:

```
                       Display Workload Report

Period:   Analysis
                    CPU     Thruput  -------Response Times (Secs)-------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
OPERATIONS    1       .0        150        .2        .2       .0       .0
AP20          2       .1       4000        .0        .0       .0       .0
GL95          2       .7       1000        .3        .3       .0       .0
OPERATIONS    2       .0          0       6.5       6.5       .0       .0
PR13          2      4.5       6000        .2        .2       .0       .0
PR51          2      1.8       1000       2.0       2.0       .0       .0
QDEFAULT      2       .3        120       1.0       1.0       .0       .0
QPFRMON       2       .0          0       9.2       9.2       .0       .0
RQ10          2       .7       2286        .1        .1       .0       .0




                                                                      Bottom
Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-analyze   F11=Response time detail     F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

Figure 32. Display Workload Report

5.3.3 Growth Analysis for Model 650

Growth was applied to the model using the Best/1 facility shown in the following display. The percentage increases applied were 150% and 133%. These increases were applied to all the workloads except the OPERATIONS workload. The OPERATIONS workload was due to the "green screen" transactions that were entered by the personnel who were monitoring the AS/400 performance (that is, they were mostly WRKSYSSTS and WRKACTJOB commands).

These growth figures should have increased the client/server workload to be equivalent to 1500 users and then to 2000 users. This allows us to consider the accuracy of the Best/1 model as we also have the actual performance data for the system when it was supporting these numbers of users.

```
                       Specify Growth of Workload Activity

 Type information, press Enter to analyze model.
   Determine new configuration . . . . . . . . . . .   N    Y=Yes, N=No
   Periods to analyze  . . . . . . . . . . . . . . .   3    1 - 10

   Period 1  . . . . . . . . . . . . . .  1000user       Name
   Period 2  . . . . . . . . . . . . . .  1500user       Name
   Period 3  . . . . . . . . . . . . . .  2000user       Name
   Period 4  . . . . . . . . . . . . . .  Period 4       Name
   Period 5  . . . . . . . . . . . . . .  Period 5       Name

             ------Percent Change in Workload Activity-------
 Workload    Period 1  Period 2  Period 3  Period 4  Period 5
 AP20             .0      50.0      33.3      20.0      20.0
 GL95             .0      50.0      33.3      20.0      20.0
 OPERATIONS       .0       0.0       0.0      20.0      20.0
 PR13             .0      50.0      33.3      20.0      20.0
 PR51             .0      50.0      33.3      20.0      20.0
 QDEFAULT         .0      50.0      33.3      20.0      20.0
                                                                  More...
F3=Exit   F11=Specify total growth   F12=Cancel   F13=Display periods 6 to 10
F17=Analyze using ANZBESTMDL
```

Figure 33. Specify Growth of Workload Activity

The model was re-analyzed with these growth figures, and the Analysis Summary that was produced is shown in Figure 34 on page 82.

```
                         Display Analysis Summary

                     Stor     CPU  -Disk IOPs--  -Disk Ctls--  -Disk Ar
 Period    CPU Model  (MB)    Util  Nbr    Util   Nbr    Util   Nbr
 1000user  650 2240  17408     8.3    7     1.5    14      .5   116
 1500user  650 2240  17408    12.2    7     2.3    14      .7   116
 2000user  650 2240  17408    16.1    7     3.0    14      .9   116




          ----Inter Rsp Time----  -------Inter--------  -----Non-Inter--
 Period     Local    LAN     WAN  CPU Util    Trans/Hr  CPU Util    Tran
 1000user      .2      .0      .0       .0         150       8.2       1
 1500user      .2      .0      .0       .0         150      12.1       2
 2000user      .2      .0      .0       .0         150      16.0       2


                                                                    Bottom
F3=Exit   F10=Re-analyze   F11=Alternative view         F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

Figure 34. Display Analysis Summary

The Workload Report for this model is shown in Figure 35 on page 83 as this is where we receive the Best/1 forecast of the client/server response times for each of our scripts.

Display Workload Report

Period: 1000user

Workload	Type	CPU Util	Thruput per Hour	-------Response Times (Secs)------- Internal	Local	LAN	WAN
OPERATIONS	1	.0	150	.2	.2	.0	.0
AP20	2	.1	4000	.0	.0	.0	.0
GL95	2	.7	1000	.3	.3	.0	.0
OPERATIONS	2	.0	0	6.5	6.5	.0	.0
PR13	2	4.5	6000	.2	.2	.0	.0
PR51	2	1.8	1000	2.0	2.0	.0	.0
QDEFAULT	2	.3	120	1.0	1.0	.0	.0
QPFRMON	2	.0	0	9.2	9.2	.0	.0
RQ10	2	.7	2286	.1	.1	.0	.0

Period: 1500user

Workload	Type	CPU Util	Thruput per Hour	-------Response Times (Secs)------- Internal	Local	LAN	WAN
OPERATIONS	1	.0	150	.2	.2	.0	.0
AP20	2	.2	6000	.0	.0	.0	.0
GL95	2	1.1	1500	.3	.3	.0	.0
OPERATIONS	2	.0	0	6.5	6.5	.0	.0
PR13	2	6.7	9000	.2	.2	.0	.0
PR51	2	2.7	1500	2.0	2.0	.0	.0
QDEFAULT	2	.5	180	1.0	1.0	.0	.0
QPFRMON	2	.0	0	9.3	9.3	.0	.0
RQ10	2	.9	2743	.1	.1	.0	.0

Period: 2000user

Workload	Type	CPU Util	Thruput per Hour	-------Response Times (Secs)------- Internal	Local	LAN	WAN
OPERATIONS	1	.0	150	.2	.2	.0	.0
AP20	2	.2	7998	.0	.0	.0	.0
GL95	2	1.5	1999	.3	.3	.0	.0
OPERATIONS	2	.0	0	6.5	6.5	.0	.0
PR13	2	9.0	11997	.2	.2	.0	.0
PR51	2	3.6	1999	2.0	2.0	.0	.0
QDEFAULT	2	.6	240	1.0	1.0	.0	.0
QPFRMON	2	.1	0	9.3	9.3	.0	.0
RQ10	2	1.0	3292	.1	.1	.0	.0

Bottom

Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-analyze   F11=Response time detail   F12=Cancel

Figure 35. Display Workload Report

5.3.4 Comparison of Forecast with Actual for AS/400 Model 650

The data from the model was then tabulated with data from Performance Tools/400 System Reports for the appropriate time intervals and the results are shown in Figure 36.

Period	Time Interval	Total Number of client/server users	Total CPU % from System Report	Total CPU % from Best/1
P1	10:00-10:10	1,000	8.2	8.3
P2	10:45-10:55	1,500	12.1	12.2
P3	12:10-12:20	2,000	15.0	16.1

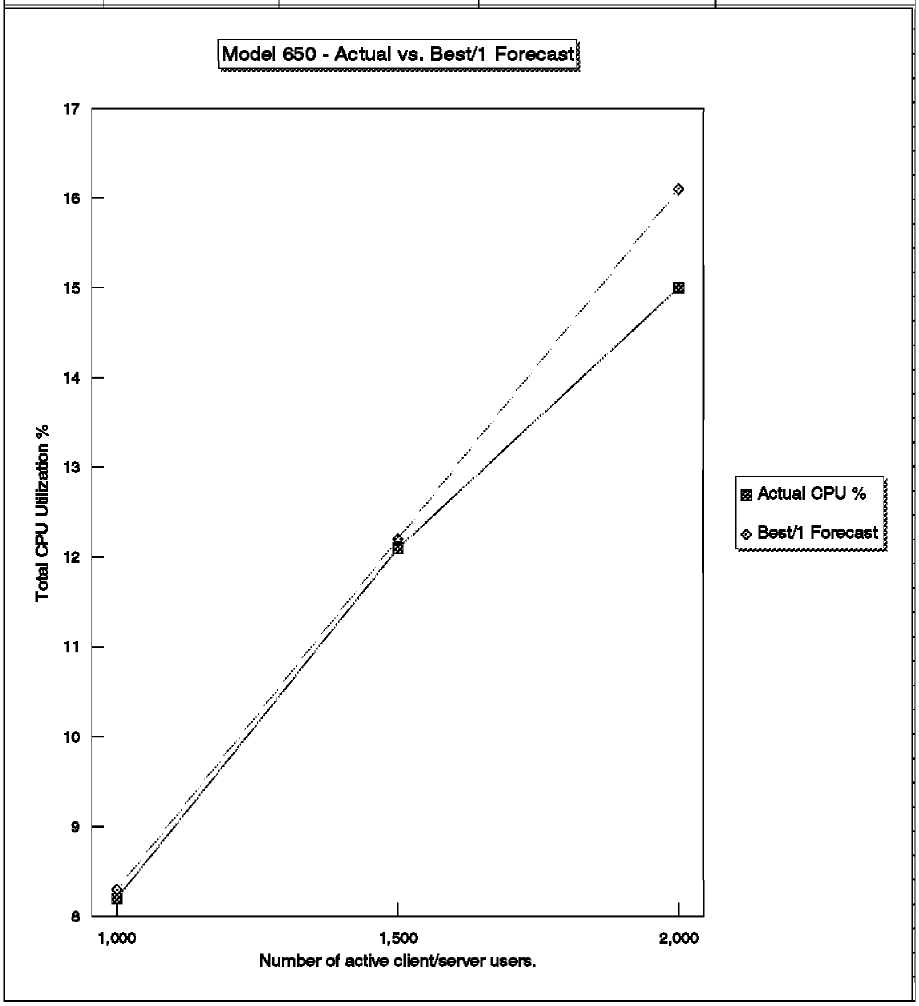


Figure 36. Spread Sheet and Graphs of Actual CPU versus Best/1 Forecast

5.4 Volume Test on 650 with External Response Time Measurements

In the testing that has been described up to this point, we did not have any external and independent measurements of end-user response times. The testing that has been previously described on the model 650 did not put a heavy client/server load on the system due to the need to also test the batch processing. In the following test, the script iteration rates was quadrupled and the number of active client/server users was increased to 2502 to put a much heavier client/server load on the model 650. There were also modifications made to the scripts that unfortunately meant that we could not migrate Best/1 workloads from the earlier testing. The adjustments to the scripts made the AP20 script add a heavier load onto the AS/400 server, and the changes made to script PR51 reduced the load that it placed on the server.

However, the following test does provide some reasonably useful verification of the Best/1 estimates of the client/server response times.

5.4.1 Revised Iteration Rates by Script

The rate of the five scripts was adjusted so they effectively cycled four times faster. The resultant iteration rates and the aggregate business transaction rates per script are shown in the following tables:

Table 7. Model 650 - Revised Iteration Rates by Script

Base Iteration Rate of Each Job	RQ10 Script	AP20 Script	PR51 Script	PR13 Script	GL95 Script
Range (secs)	72-108	36-54	144-216	24-36	144-216
Average (secs)	90	45	180	30	180
BTs / hour / user	40	80	20	120	20

Table 8. Revised Rate of Business Transactions (BTs) by Script for Model 650

Total Number of Users	RQ10 Script	AP20 Script	PR51 Script	PR13 Script	GL95 Script	Totals
1 user	40	80	20	120	20	
Users / script	500	500	501	500	501	2502
2502 users	20 000	40 000	10 020	60 000	10 020	140 040

5.4.2 Revised Model Creation for Model 650

The preceding client/server transaction rates were used to develop a model based on the subsystems as shown in Figure 37 on page 86. It is interesting to note the much higher levels of system activity and the impact of the script changes (that is, the AP20 workload is now much heavier and the PR51 workload is much lighter). The first step was, of course, to select the three time intervals shown on the following display:

```
                             Select Time Interval

Library  . . . . . . . :   QPFRLAW650    Performance member . . :   L03N

Type option, press Enter.  Select first and last interval.
  1=Select

                              ---Transaction---   --CPU Util---   I/Os per Sec
Opt   Date       Time          Count   Rsp Time   Total   Inter   Sync   Async
      11/03/97   11:15:10          4         .5      73       0    429      18
      11/03/97   11:20:10         17         .6      71       0    363       9
 1    11/03/97   11:25:10         10        1.1      69       0    329       4
      11/03/97   11:30:09         14         .6      69       0    330       4
 1    11/03/97   11:35:10         13        1.0      70       0    340       4
      11/03/97   11:40:09          7        7.6      61       0    680       8
      11/03/97   11:45:09          0         .0      41       0    182       4
      11/03/97   11:50:09          0         .0      26       0    145       5
      11/03/97   11:55:10          4         .2      27       0    207      13
      11/03/97   12:00:10          0         .0      23       0    219      16
      11/03/97   12:05:10          0         .0      19       0    205      17
                                                                          More...
```

Figure 37. Select Time Interval

Within the creation of the Job Classification, workloads were then designated, which again mapped to the subsystems for each of the five scripts (see Figure 38).

```
                            Assign Jobs to Workloads

Workload . . . . . . . . . . . . . . . . .

Type options, press Enter.  Unassigned jobs become part of workload QDEFAULT
  1=Assign to above workload   2=Unassign

                                Number of        CPU        I/O
Opt  Workload    Subsystem   Transactions     Seconds      Count
                                     0         39.498      22842
     RQ10        LASERVE1            0        476.658      33054
     AP20        LASERVE2            0        306.493      16424
     PR51        LASERVE3            0        232.409      15652
     PR13        LASERVE4            0       3234.492     185915
     GL95        LASERVE5            0        692.911      51852
                 LAWSON              0           .000          0
                 QBATCH              0           .000          0
                 QCMN                0           .000          0
                 QCTL                0          5.332        448
     OPERATIONS  QINTER             37          1.827        228
                 QSERVER             0           .000          0
                 QSPL                0           .020         36
                 QSYSWRK             0           .065         12

                                                                         Bottom
F3=Exit   F12=Cancel      F15=Sort by workload      F16=Sort by subsystem
F17=Sort by transactions   F18=Sort by CPU seconds   F19=Sort by I/O count
```

Figure 38. Assign Jobs to Workloads

The next step was, of course, to again define the non-interactive client/server transactions using the figures from the preceding business transaction iteration rate table.

```
                        Define Non-Interactive Transactions

 Job classification category  . . . . . . :   Subsystem


 Type choices, press Enter.

                ---Activity Counted as Transaction---     Total Transacti
 Workload           Type                  Quantity         when Type = *N
 QDEFAULT           *LGLIO                  100.0              0
 OPERATIONS         *LGLIO                  100.0              0
 GL95               *NONE                   100.0              10020
 PR13               *NONE                   100.0              60000
 PR51               *NONE                   100.0              10020
 AP20               *NONE                   100.0              40000
 RQ10               *NONE                   100.0              20000




 Type:  *LGLIO, *CMNIO, *CPUSEC, *PRINT, *NONE

 F3=Exit   F12=Cancel
```

Figure 39. Define Non-Interactive Transactions

The model was then submitted for creation, and the following display shows that there was quite a good match between the "Predicted" and the "Actual".

```
                         Measured and Predicted Comparison

                                             Measured      Predicted
 Total CPU util . . . . . . . . . . . :          70.0          70.0
 Disk IOP util  . . . . . . . . . . . :           6.7           3.5
 Disk arm util  . . . . . . . . . . . :           1.9           1.9
 Disk IOs per second  . . . . . . . . :         437.0         439.5

 LAN IOP util . . . . . . . . . . . . :           8.6            .0
 LAN line util  . . . . . . . . . . . :            .0            .0
 WAN IOP util . . . . . . . . . . . . :           5.8           3.6
 WAN line util  . . . . . . . . . . . :            .0            .0

 Interactive:
   CPU util . . . . . . . . . . . . . :            .0            .0
   Int rsp time (seconds) . . . . . . :            .9            .1
   Transactions per hour  . . . . . . :           148           148
 Non-interactive thruput  . . . . . . :        140243        140300


 Performance estimates -- Press help to see disclaimer.
 F3=Exit   F6=Print   F9=Work with spooled files   F12=Cancel
 F17=Calibrate response time
```

Figure 40. Measured and Predicted Comparison

Having satisfied ourselves that we have a reasonable model, again without the need for any "calibration", we can review the "Workload Report" shown in Figure 41 on page 88.

```
                           Display Workload Report

Period:   Analysis
                   CPU     Thruput  -------Response Times (Secs)-------
Workload    Type   Util   per Hour  Internal   Local     LAN      WAN
OPERATIONS    1      .0        148        .1      .1      .0       .0
AP20          2     4.3      40000        .0      .0      .0       .0
GL95          2     9.7      10020        .2      .2      .0       .0
OPERATIONS    2      .3          0       1.6     1.6      .0       .0
PR13          2    45.1      60000        .2      .2      .0       .0
PR51          2     3.2      10020        .1      .1      .0       .0
QDEFAULT      2      .8        120       2.7     2.7      .0       .0
RQ10          2     6.6      20000        .1      .1      .0       .0

                   CPU     Thruput  -------Response Times (Secs)-------
Workload    Type   Util   per Hour  Internal   Local     LAN      WAN
OPERATIONS    1      .0        148        .1      .1      .0       .0
AP20          2     4.3      40000        .0      .0      .0       .0
GL95          2     9.7      10020        .2      .2      .0       .0
OPERATIONS    2      .3          0       1.6     1.6      .0       .0
PR13          2    45.1      60000        .2      .2      .0       .0
PR51          2     3.2      10020        .1      .1      .0       .0
QDEFAULT      2      .8        120       2.7     2.7      .0       .0
RQ10          2     6.6      20000        .1      .1      .0       .0


Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-analyze   F11=Response time detail      F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

Figure 41. Display Workload Report

5.4.3 Comparison of Best/1 Response Time Forecasts with Actuals

The Performance Tools/400 DSPPFRDTA command was used to display the following details of the performance of the model 650 for the three 5-minute time intervals when the number of active client/server jobs was steady at 2502.

```
                         Select Time Intervals to Display

 Member . . . . . . . :   L03NOV97R1      Library  . . . . . . :   QPFRLA

 Type options, press Enter.
   1=Select

                   Transaction  --CPU Util---  --High---  -Pool Fault-
 Opt  Date   Time   Count   Rsp  Tot  Int  Bch  Dsk  Unit  Mch  Usr  ID
      11/03  11:15      4    .5   73    0   72    6  0116    2  381  05
      11/03  11:20     17    .6   71    0   69    4  0002    0  342  05
   1  11/03  11:25     10   1.1   69    0   68    4  0004    1  321  05
   1  11/03  11:30     14    .5   69    0   68    4  0003    0  322  05
   1  11/03  11:35     13   1.0   70    0   69    5  0013    1  329  05
      11/03  11:40      7   7.5   61    0   57   34  0006    4  637  05
      11/03  11:45      0    .0   41    0   40    2  0033    0  175  05
      11/03  11:50      0    .0   26    0   25    3  0082    1  139  05
      11/03  11:55      4    .2   27    0   29    3  0043    0  178  05
      11/03  12:00      0    .0   23    0   22    4  0001    2  189  05
      11/03  12:05      0    .0   19    0   18    3  0004    0  179  05
                                                                        More...
F3=Exit   F5=Refresh   F11=Display histogram   F12=Cancel   F13=Select all
F14=Deselect all
```

Figure 42. Select Time Intervals to Display

The preceding selection led to the following system summary display:

```
                             Display Performance Data

 Member . . . . . . . .   L03NOV97R1           F4 for list
   Library  . . . . . .     QPFRLAW650

 Elapsed time . . . . :   00:14:59        Version  . . . . . . :   4
 System . . . . . . . :   CBC658A         Release  . . . . . . :   1.0
 Start date . . . . . :   11/03/97        Model  . . . . . . . :   650-22
 Start time . . . . . :   09:20:09        Serial number  . . . :   10-28T
 QPFRADJ  . . . . . . :   0               QDYNPTYSCD . . . . . :   1

 CPU utilization (priority) . . . . . . . . . . . :           .02
 CPU utilization (other)  . . . . . . . . . . . . :         69.86
 Job count  . . . . . . . . . . . . . . . . . . . :          2514
 Transaction count  . . . . . . . . . . . . . . . :            37
 Transactions per hour  . . . . . . . . . . . . . :           148
 Average response (seconds) . . . . . . . . . . . :           .91
 Disk utilization (percent) . . . . . . . . . . . :          1.62
 Disk I/O per second  . . . . . . . . . . . . . . :         362.9


F3=Exit      F4=Prompt   F5=Refresh   F6=Display all jobs   F10=Command entry
F12=Cancel   F24=More keys
```

Figure 43. Display Performance Data

From the System Summary, we can easily view the utilization by subsystem and the activity by main storage pool.

```
                              Display by Subsystem

 Member . . . . . . . :   L03NOV97R1      Elapsed time . . . . :   00:14:
   Library  . . . . . :     QPFRLAW650

 Type options, press Enter.  Press F6 to display all jobs.
   5=Display jobs

                            CPU        Job         Tns      Average      Disk
 Option    Subsystem        Util       Count       Count    Response      I/O
           *MACHINE         1.05         42           0         .00     22831
           LASERVE1         6.62        500           0         .00     33054
           LASERVE2         4.26        500           0         .00     16424
           LASERVE3         3.23        501           0         .00     15652
           LASERVE4        44.97        500           0         .00    185915
           LASERVE5         9.63        501           0         .00     51852
           LAWSON            .00          0           0         .00         0
           QBATCH            .00          0           0         .00         0
           QCMN              .00          0           0         .00         0
           QCTL              .07          1           0         .00       448
                                                                        More...
                              Display Pool Detail

 Member . . . . . . . :   L03NOV97R1      Elapsed time . . . . :   00:14:
   Library  . . . . . :     QPFRLAW650

 Type options, press Enter.
   5=Display pool intervals

              Size     Act     Tns     Avg  Active->    Wait->    Active-
 Opt  Pool     (K)     Lvl    Count    Rsp    Wait    Ineligible  Ineligib
       01   ++++++      0        0      .0       .7          .0         .0
       02   ++++++    346        0      .0      4.9          .0         .0
       03   ++++++      2        0      .0      1.0          .0         .0
       04   965280    786       37      .8      2.4          .0         .0
       05   ++++++    500        0      .0   2329.1          .0         .0
       06    10000      6        0      .0       .0          .0         .0




                                                                       Bottom
F3=Exit   F11=Display expert cache   F12=Cancel   F15=Sort by pool
F24=More keys
```

Figure 44. Display by Subsystem

There are two points that emerge from the preceding displays:

1. From the subsystem details, we can see that the script changes have been effective in increasing the utilization of AP20 and reducing the resource required for GL95. It is also interesting to note that the number of active jobs corresponds exactly to the expected numbers. This is another indication that by the time we did this volume test, earlier script problems had been completely resolved.
2. From the Pool Details, we can see that the "Active to Wait Transitions" for system pool 5 (the pool that the client/server jobs were running in) was 2329.1 transitions per minute. This figure is remarkably close to the calculated figure for the client/server transactions per minute, which is 140 040/60 (or 2334).

The A-W transitions in a pool have always been used as a good guide to the number of traditional "green screen" transactions per minute so it is useful to have this strong evidence that it is still a good indicator of activity in the client/server world.

On this test, we were able to extract from LoadRunner (the volume test software that we were using), figures for the response times measured at the LoadRunner tools. This is close to the "end-user" response time as it includes two of the three components of client/server response time. That is, it includes not only the "server response time" but also any "line time". In this case, the "line time" is extremely small because, although we had quite a high workload, we had five 16 megabit token-ring cards connecting the LoadRunner tools to the AS/400 system, one for each script.

The following table summarizes the preceding information by showing the Best/1 response time estimates and the LoadRunner measurements of response time at the time when the workload was held steady at 2502 client/server users.

Table 9. Comparison of LoadRunner Response Times with Best/1 Estimates

Scripts	**RQ10**	**AP20**	**PR51**	**PR13**	**GL95**
LoadRunner Resp Time	0.20 sec	0.07 sec	0.20 sec	0.30 sec	0.05 sec
Best/1 Estimate Resp Time	0.1 sec	0.0 sec	0.1 sec	0.2 sec	0.2 sec
BT per hour	20 000	40 000	10 020	60 000	10 020
CPU %	6.62	4.62	3.23	44.97	9.63

Although all of these response times are short, which makes the comparison rather difficult, there is a reasonable level of agreement here with Best/1 getting four out of the five response times almost precisely correct. It is hoped to do more testing on this aspect of Best/1 in the near future, and hopefully publish the results, probably on the Web. However, in the mean time, these results are not definitive and care should, therefore, be exercised when quoting response times derived from Best/1 using this approach.

5.5 Summary

The method worked well with the performance data from the sockets volume tests. In particular, the ease with which the various models were created was pleasing; there was no need to carry out "model calibration" on any of the models that were developed.

The main points that emerged are as follows:

- Best/1 appeared to be able to forecast CPU with reasonable accuracy for the client/server workloads using the method described in the earlier chapters of this document.
- The position on Best/1's ability to forecast client/server response times is less clear cut and needs further work. However, there is some evidence that the tool does a reasonable job of this task, but clearly particular caution needs to be used in this area.
- The A-W transitions are a good indicator of the number of client/server transactions per minute being performed in that pool. Testing did show that

the A-W transitions did begin to exceed the client/server transactions per minute when the CPU was pushed above its guideline figure for good performance. This is, of course, what also happens with traditional applications. When jobs are having to queue excessively for CPU, some of them will not complete their work on a single pass through the CPU; this results in multiple A-W transitions per transaction.

Chapter 6. Working Example of an APPC Based Application

In this chapter, we want to use the recommended approach to analyze a working example of an application using APPC interfaces.

This chapter covers the following capacity planning aspects:

- Model creation using user-defined job classification
- Growth analysis

6.1.1 Objective

The objective of this chapter is to give you the capability to identify appropriate performance data for creating an accurate APPC application model and use that model for modeling workload growth.

6.2 Introduction to the APPC Based Application

This example is developed from a banking application. The bank's AS/400 server (9406-310#2044 CISC system with an RPR rating of 20.2 or CPW rating of 56.5) connects to the remote sites through X.25, DDN, PSTN, and so on. The AS/400 functions as an application server and database server. At the remote site, an SCO UNIX server is the front end to ASCII banking terminals. There are approximately 150 banking terminals running the application. The APPN network is illustrated by the following diagram:

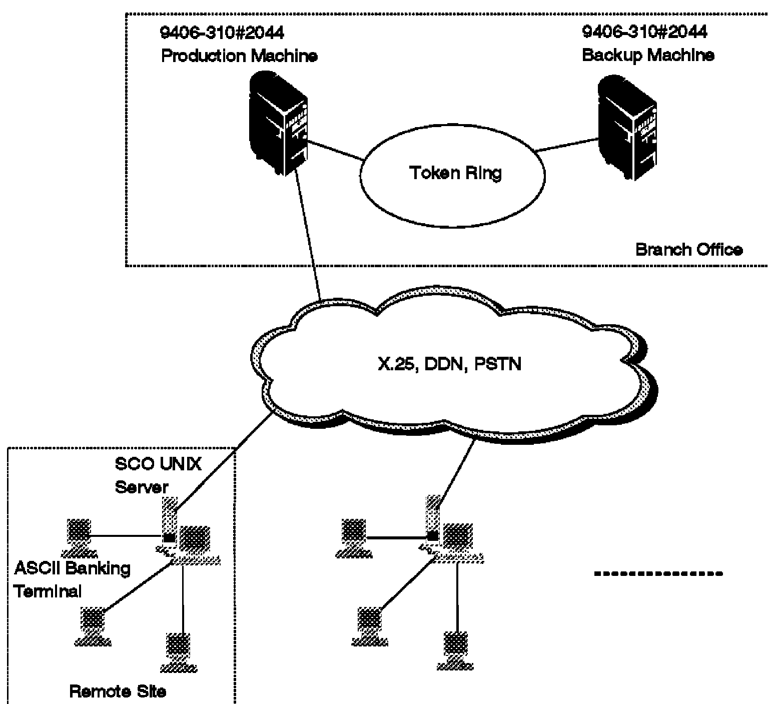


Figure 45. Typical APPN Network

On the SCO UNIX server, a C language program using UNIX APPC APIs communicates a C program on an AS/400 server data queue.

The user at an ASCII banking terminal initiates a request. The SCO UNIX server transfers the data to an AS/400 prestart job through a data queue. The AS/400 system provides the application functions and data. The following diagram illustrates the application architecture:

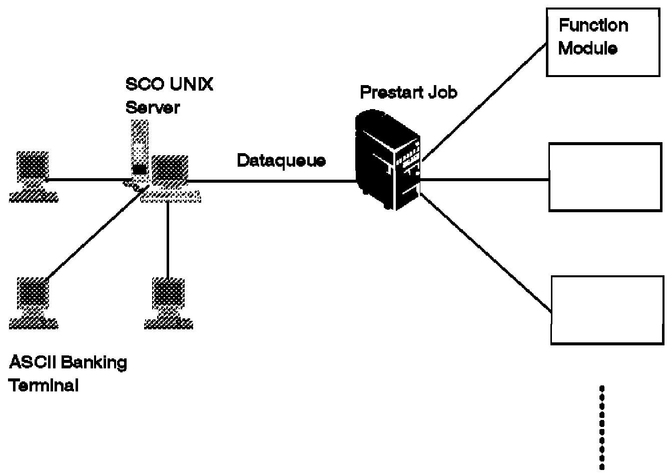


Figure 46. APPC Working Model

6.3 Performance Data Collection

In our example, we selected the busiest hour in the busiest day. We use this one hour's data to create the model. The description of the data is summarized in the following table:

Table 10. APPC Modeling Performance Data

Application	Banking Comprehensive Application
Number of clients	150
Transactions per hour	1800

To collect the performance data, we issued the following command:

```
STRPFRMON
```

```
                        Start Performance Monitor (STRPFRMON)

 Type choices, press Enter.

 Member . . . . . . . . . . . . . .   *GEN          Name, *GEN
 Library  . . . . . . . . . . . . > QPFRBANK      Name
 Text 'description' . . . . . . .   *SAME

 Time interval (in minutes) . . .   15            5, 10, 15, 20, 25, 30,
 Stops data collection  . . . . .   *ELAPSED      *ELAPSED, *TIME, *NOMAX
 Days from current day  . . . . .   0             0-9
 Hour . . . . . . . . . . . . . .   2             0-999
 Minutes  . . . . . . . . . . . .   0             0-99
 Data type  . . . . . . . . . . .   *ALL          *ALL, *SYS
 Trace type . . . . . . . . . . .   *NONE         *NONE, *ALL
 Dump the trace . . . . . . . . .   *YES          *YES, *NO
 Job trace interval . . . . . . .   .5            .5 - 9.9 seconds
 Job types  . . . . . . . . . . .   *DFT          *NONE, *DFT, *ASJ, *BCH
               + for more values
 Start database monitor . . . . .   *NO           *YES, *NO
                                                                      More...
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

6.4 Creating a Model with User-Defined Job Classification

We want to create a BEST/1 model to use as the base for projecting the impact of growth in workload on the system. We also want to model the APPC work separate from any other work on the system to ensure we get an accurate model. Therefore, we assigned all the APPC work to a specific workload and isolated the APPC workload from other system activity.

The remainder of this chapter shows the steps in building the BEST/1 model and selecting different capacity planning options based on the percentage of growth to determine the AS/400 system requirements.

Note

Not every BEST/1 display is shown in every step. Only the more significant displays are shown.

6.4.1 Creating the Model from Performance Data

1. Sign on the system.
2. Start BEST/1 by typing the STRBEST command and pressing F4, which prompts for input.

```
                         Start BEST/1 (STRBEST)

Type choices, press Enter.

BEST/1 data library  . . . . . .   A970121C      Name, *CURLIB
Performance data library . . . .   QPFRBANK      Name
Log member . . . . . . . . . . .   *NONE         Name, *NONE
Log library  . . . . . . . . . .   *BESTDTAL     Name, *BESTDTAL
User level . . . . . . . . . . .   *ADVANCED     *ADVANCED, *BASIC


F3=Exit  F4=Prompt  F5=Refresh  F12=Cancel  F13=How to use this display
F24=More keys
```

3. On the BEST/1 for the AS/400 menu, select Option 1 (Work with BEST/1 models).
4. On the Work with BEST/1 Models menu, select Option 1 (Create):

   ```
   Opt   Model
    1.   Model_name
   ```

   where `Model_name` is the name of the model to be created.

   We used **CSAPPCMDL** as the model name.
5. On the Create BEST/1 Model menu, select Option 1 (Create from performance data).
6. On the Create Model from Performance Data menu, we specified the performance member containing the data and indicated the time period to be used in model creation:

```
                   Create BEST/1 Model from Performance Data

Model . . . . . . . . . . . . . . :   CSAPPCMDL

Type choices, press Enter.  Use *SLTHOUR to select an hour-long time period or
  use *SLTITV to select first and last interval of a one to two hour
  time period.  The time period selected should be representative of your peak
  processing activity.

Text . . . . . . . . . . . . .   Model for C/S APPC application example

Performance data:
  Member . . . . . . . . . . .   Q972680829   Name, F4 for list
  Library  . . . . . . . . . .   QPFRBANK     Name

Start time . . . . . . . . . .   09:15:00     Time, *FIRST, *SLTHOUR, *SLTITV
Start date . . . . . . . . . .   *FIRST       Date, *FIRST

Stop time  . . . . . . . . . .   10:13:59     Time, *LAST
Stop date  . . . . . . . . . .   *LAST        Date, *LAST

F3=Exit   F4=Prompt   F12=Cancel
```

   You may select the parameter of *SLTHOUR for Start time. When you press Enter, you see a display that allows you to choose a start time; the system automatically gets an hour's data for you.
7. On the Classify Jobs menu, select Option 2 (Classify jobs into workloads):

   We want to specifically define our own workload so we can model APPC work separately from other work on the measured system.
8. In the Specify Job Classification Category, select **category 6 (Subsystem)**.

9. Press F9 (Display values from data). This runs a query against the collected data (Q972680829), and produces a list of subsystems, number of transactions, CPU times, and so on. The Account, Communicate, Credit, and Deposit subsystem contains the APPC workload.

10. Assign the workloads as follows:

    - APPCWL - APPC Workload
    - ATMSWL - ATM Workload
    - MIMIXWL - Mimix High Availability Workload
    - WTRWL - Printing Workload

```
                             Assign Jobs to Workloads

 Workload . . . . . . . . . . . . . . . . .   _________

 Type options, press Enter.  Unassigned jobs become part of workload QDEFAULT
   1=Assign to above workload   2=Unassign

                                Number of       CPU       I/O
 Opt  Workload    Subsystem   Transactions    Seconds     Count
      APPCWL      ACCOUNT            0       2694.109    397744
      ATMSWL      ATMMS              0        775.757     69256
      APPCWL      COMUNICATE         0        614.874     52907
      ATMSWL      TERMINAL         155        336.652      4630
                                     0        323.236     29601
      APPCWL      CREDIT             0        295.932     41068
      MIMIXWL     MIMIXSBS           0        275.539      6937
                  STARTUP          126         63.754     12367
                  BACKTOP            0         15.871       634
      APPCWL      DEPOSIT            0          5.100        32
      WTRWL       PRINTER            0          4.922      1050
                                                                      More...
F3=Exit   F12=Cancel       F15=Sort by workload      F16=Sort by subsystem
F17=Sort by transactions   F18=Sort by CPU seconds   F19=Sort by I/O count
```

    The jobs that you do not assign to any workload are included in the QDEFAULT workload.

11. On the Specify Paging Behaviors menu, accept the default *GENERIC for all workloads (APPCWL, ATMSWL, MIMIXWL, WTRWL, and QDEFAULT).

12. Define a non-interactive transaction:

    The following display shows the Define Non-Interactive Transactions menu. It is important that you define the APPC (client/server) transaction in terms acceptable to the customer.

    On this display, you define a unit of work that corresponds to a meaningful business unit of work. BEST/1 defaults to 100 *LGLIO (Logical I/Os) for a non-interactive transaction.

    By typing *NONE in the Type column for workload APPC, you can specify a number of business or user transactions in the Total Transactions column. This value is in transactions per hour. Knowledge of the activity during the measured period tells us that the total transactions in one hour is 1800.

```
                     Define Non-Interactive Transactions

Job classification category  . . . . . . :   Subsystem


Type choices, press Enter.

               ---Activity Counted as Transaction---     Total Transactions
Workload          Type                 Quantity          when Type = *NONE
QDEFAULT          *LGLIO                 100.0                        0
WTRWL             *LGLIO                 100.0                        0
MIMIXWL           *LGLIO                 100.0                        0
ATMSWL            *LGLIO                 100.0                        0
APPCWL           .*NONE            100.0                .1800


Type:  *LGLIO, *CMNIO, *CPUSEC, *PRINT, *NONE

F3=Exit   F12=Cancel
```

Press Enter and the Save Job Classification Member display is shown.

```
                        Save Job Classification Member
Change values if desired, press Enter.
  Member . . . . . . . . . . . .   CSAPPCJC _     Name
    Library  . . . . . . . . . .     A970121C     Name
  Text . . . . . . . . . . . . .   APPC Job Classification_____
  Replace  . . . . . . . . . . .   N              Y=Yes, N=No


F12=Cancel
```

Press Enter and the Confirm Creation of BEST/1 Model display is shown.

13. Confirm the creation of the BEST/1 model:

    On this display, you are prompted with your previously specified model name (in this publication, CSAPPCMDL) and descriptive text. You may change these values if you want to.

```
                       Confirm Creation of BEST/1 Model
Type choices, press Enter.
  Model  . . . . . . . . . . . .   CSAPPCMDL      Name
    Library  . . . . . . . . . .     A970121C     Name
  Text  . . . . . . . . . . . . . .   Model for C/S APPC application
  Replace  . . . . . . . . . . .   N              Y=Yes, N=No
  Job name . . . . . . . . . . .   CRTBESTMDL     Name, *JOBD
  Job description  . . . . . . .   QPFRJOBD       Name, *NONE, *USRPRF
    Library  . . . . . . . . . .     QPFR         Name, *LIBL, *CURLIB


F12=Cancel
```

BEST/1 submits the model creation to job queue QBATCH.

14. Now you are back in the Work with BEST/1 Models menu. The CRTBESTMDL job is running.

    You may issue the Work with Submitted Job (WRKSBMJOB) command to find out when your create model job completes, or issue DSPMSG and look for a `model completion` message. You can also repetitively use the F5 key (Refresh) until you see your new model name appear on the Work with BEST/1 Models menu. After you see your new model on this display, proceed to the next step.

15. On the Work with BEST/1 Models menu, select Option 5 (Work with) for your model (CSAPPCMDL).

16. On the Work with BEST/1 Model menu, select Option 5 (Analyze current model).

```
                                Work with Results

 Printed report text . . . . . .   Model for C/S APPC application example



 Type options, press Enter.
   5=Display   6=Print

 Opt     Report Name
 __      Measured and Predicted Comparison
 __      Analysis Summary
 __      Recommendations
 __      Workload Report
 __      ASP and Disk Arm Report
 __      Disk IOP and Disk Arm Report
 __      Main Storage Pool Report
 __      Communications Resources Report
 __      All of the above

 F3=Exit   F12=Cancel   F14=Select saved results   F15=Save current result
 F18=Graph current results   F19=Append saved results   F24=More keys
Model has been Analyzed
```

17. On the Work with Results menu, select Option 5 (Display) next to **Measured and Predicted Comparison**:

    This shows a column of model performance statistics that should be compared to measured performance data.

```
                      Measured and Predicted Comparison

                                         Measured     Predicted
 Total CPU util . . . . . . . . . . . :      79.3          79.3
 Disk IOP util  . . . . . . . . . . . :       8.2           8.8
 Disk arm util  . . . . . . . . . . . :       4.4           4.4
 Disk IOs per second  . . . . . . . . :     178.0         177.8

 LAN IOP util . . . . . . . . . . . . :       4.6           4.3
 LAN line util  . . . . . . . . . . . :        .4            .3
 WAN IOP util . . . . . . . . . . . . :       6.9           6.4
 WAN line util  . . . . . . . . . . . :       3.0           1.5

 Interactive:
   CPU util . . . . . . . . . . . . . :       5.5           5.5
   Int rsp time (seconds) . . . . . . :        .8           7.8
   Transactions per hour  . . . . . . :       282           282
 Non-interactive thruput  . . . . . . :      7976          7976

 Performance estimates -- Press help to see disclaimer.
 F3=Exit   F6=Print   F9=Work with spooled files   F12=Cancel
```

You can see the measured CPU utilization is the same as predicted CPU utilization. The model we created corresponds to the measured environment.

6.4.2 Calibrating the Model

We can see that we have created a good model. The predicted CPU utilization matches the measured CPU utilization, but the response time of the APPC transaction does not reflect the real situation. In following section, we show you how to calibrate the response time to match the measured value, if there was a difference.

Note

If the predicted CPU utilization does not match the measured CPU utilization, we should calibrate the model further.

On the Work with Results menu, select Option 5 (Display the Workload Report). Then use F11 (Response time) detail to determine the BEST/1 estimate of the non-interactive response time components associated with the APPCWL workload.

```
                           Display Workload Report

Period:   Analysis
                           Total   --------Rsp Time Secs spent in--------
Workload    Type  Connect  Rsp Time     CPU      I/O     Pool     Comm    Other
ATMSWL        1   *LOCAL      13.9     13.6       .1       .0       .2       .0
QDEFAULT      1   *LOCAL        .7       .3       .3       .0       .0       .0
APPCWL        2   *LOCAL       1.3      1.0       .3       .0       .0       .0
ATMSWL        2   *LOCAL      11.5      9.9      1.6       .2       .0       .0
MIMIXWL       2   *LOCAL      46.1     46.0       .1       .0       .0       .0
QDEFAULT      2   *LOCAL       2.9      2.8       .2       .0       .0       .0
WTRWL         2   *LOCAL        .2       .1       .1       .0       .0       .0


                                                                           Bottom
Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-Analyze   F11=Workload summary   F12=Cancel
```

The highlighted APPCWL 1.3 second response time should be considered as an internal response time because it does not take into account the communication line time nor the client time. Recall that the performance monitor cannot assign WAN or LAN communications bytes sent and received to client/server applications, though it does do this for 5250/3270-type workloads and transactions.

If you want to be more accurate, try to find out the "User Perceived Response Time." The user response time is:

```
User Perceived Response Time = Server Internal Time + Communications Time + Client Time
```

6.4.3 Saving the User-Defined Workload

At this point, we can consider that the APPCWL workload accurately reflects the APPC server application workload. We want to save this workload so that we can use it later for modeling the growth of this workload and for modeling it with other measured data workloads if necessary. This is especially important since we had to make some manual adjustments to the model and do not want to have to do this every time you want to model this workload, either separately or with other applications that you have measured.

1. Go back to the Work with BEST/1 Model menu and select Option 1 (Work with Workloads).
2. Select Option 8 (Save workload to workload member) for APPCWL workload and save the workload.

```
                         Save Workload to Workload Member
 Change values if desired, press Enter.
   Member . . . . . . . . . . . .   APPCWL         Name
     Library  . . . . . . . . . .     A970121C     Name
   Text . . . . . . . . . . . . .   Measured from QPFRBANK (Q972680829)
   Replace  . . . . . . . . . . .   N              Y=Yes, N=No
   CPU architecture . . . . . . .   *CISC          *CISC, *RISC



 F12=Cancel
```

Note that the RISC versions of BEST/1 have the additional parameter "CPU architecture ... *CISC or *RISC." This enables BEST/1 to more accurately perform growth analysis by knowing if the workload was based on a CISC system or a RISC system.

3. On the Work with BEST/1 Model menu, select F15 (Save the current model), changing the name (Member) of the model to CSAPPCMDLC, and changing the text field as shown to indicate the model has been **calibrated.**

```
                               Save Current Model

 Change values if desired, press Enter.

 Save to Model member:
   Member . . . . . . . . . . .   CSAPPCMDLC         Name
     Library  . . . . . . . . .     A970121C         Name
   Text . . . . . . . . . . . .   Model for APPC Appl. Calibrated
   Replace  . . . . . . . . . .   N              Y=Yes, N=No

 Externally described member information:
   Save . . . . . . . . . . . .   N              Y=Yes, N=No
   Member . . . . . . . . . . .   *MEMBER        Name, *MEMBER
     Library  . . . . . . . . .     *LIB         Name, *LIB
   Text . . . . . . . . . . . .   Model for APPC Appl. Calibrated
   Replace  . . . . . . . . . .   *REP           Y=Yes, N=No, *REP
```

Generating an "Externally described member" is for experienced BEST/1 users who want to save the model in a format that can be downloaded to a personal computer for later processing by user-written programs. Discussing these capabilities is beyond the scope of this redbook. So we select the *Save No* option.

6.5 Growth Analysis

Now that we have a good model, we can do a series of workload growth analysis using different system types (CISC to CISC growth, CISC to RISC growth, and traditional model to server model).

Consider the situation where the customer is planning to grow 100% of its business in the first year, 50% in the second year, and 50% in the third year. In the following steps, we explore the effect of this growth on the current system.

1. Go to the Work with BEST/1 Models display and select the CSAPPCMDL model. Select Option 5 (Work with) for CSAPPCMDL.

```
                        Work with BEST/1 Models

Library . . . . .   A970121C     Name

Type options, press Enter.
  1=Create   3=Copy   4=Delete   5=Work with   6=Print   7=Rename

Opt  Model       Text                                  Date      Tim

_    __________
_    CSAPPCMDLC  Model for APPC appl. Calibrated       11/19/97  11:05:01
5    CSAPPCMDL   Model for C/S APPC application example    11/18/97  14:34:31



Command
===>
F3=Exit   F4=Prompt   F5=Refresh        F9=Retrieve   F12=Cancel
F15=Sort by model     F16=Sort by text   F19=Sort by date and time
```

2. From the Work with BEST/1 Model menu, select Option 7 (Specify workload growth and analyze model). You see the Specify Growth of Workload Activity display.

   The following display shows all the parameter changes specified in this analysis.

```
                  Specify Growth of Workload Activity

Type information, press Enter to analyze model.
  Determine new configuration . . . . . . . . . .   N    Y=Yes, N=No
  Periods to analyze  . . . . . . . . . . . . . .   4    1 - 10

  Period 1  . . . . . . . . . . . . .   19971119      Name
  Period 2  . . . . . . . . . . . . .   19981119      Name
  Period 3  . . . . . . . . . . . . .   19991119      Name
  Period 4  . . . . . . . . . . . . .   20001119      Name
  Period 5  . . . . . . . . . . . . .   Period 5      Name

            ------Percent Change in Workload Activity-------
Workload    Period 1  Period 2  Period 3  Period 4  Period 5
APPCWL           .0     100.0      50.0      50.0      20.0
ATMSWL           .0      20.0      20.0      20.0      20.0
MIMIXWL          .0      20.0      20.0      20.0      20.0
QCMN             .0      20.0      20.0      20.0      20.0
QDEFAULT         .0      20.0      20.0      20.0      20.0
WTRWL            .0      20.0      20.0      20.0      20.0
                                                                    Bottom
F3=Exit   F11=Specify total growth   F12=Cancel   F13=Display periods 6 to 10
F17=Analyze using ANZBESTMDL
```

   Note: You can decide to grow all of the workloads or only the APPCWL workload according to your business requirements. In our case, we do not take ATMs, printers, and MIMIX/400 into account. Press F11 (Specify growth by workload).

3. Press Enter to perform the analysis. On completion of the analysis, the following display is shown.

```
                         Work with Results

 Printed report text . . . . . .   Model for C/S APPC application example



 Type options, press Enter.
   5=Display   6=Print

 Opt     Report Name
  _      Measured and Predicted Comparison
  5      Analysis Summary
  _      Recommendations
  _      Workload Report
  _      ASP and Disk Arm Report
  _      Disk IOP and Disk Arm Report
  _      Main Storage Pool Report
  _      Communications Resources Report
  _      All of the above

                                                                   Bottom
F3=Exit   F12=Cancel   F14=Select saved results   F15=Save current results
F18=Graph current results   F19=Append saved results   F24=More keys

Model has been analyzed
```

4. Review all of the reports to find out the non-interactive response time for the APPCWL workload and the total expected CPU utilization.

   Always include reviewing the Recommendations report when examining the results of growth analysis.

   We show only the "Display Analysis Summary" and the extended "Display Workload Report".

```
                         Display Analysis Summary

                        Stor     CPU  -Disk IOPs--  -Disk Ctls--  -Disk Arms--
 Period    CPU Model    (MB)    Util   Nbr    Util   Nbr    Util   Nbr    Util
 19971119  310 2044      320    79.3     3     8.8     6     3.8    19     4.4
 19981119  310 2044      320    99.0     3    12.0     6     5.5    19     6.5
 19991119  310 2044      320    99.0     3    12.2     6     5.2    19     6.1
 20001119  310 2044      320    99.0     3    12.3     6     4.7    19     5.5


                                                                        Bottom
           ----Inter Rsp Time----  -------Inter--------  -----Non-Inter------
 Period     Local     LAN     WAN   CPU Util     Trans/Hr  CPU Util    Trans/Hr
 19971119     7.9      .0      .0      5.5          282     73.8        7976
 19981119      .6      .0      .0       .7          152     98.3        7114
 19991119      .6      .0      .0       .8          182     98.2        8429
 20001119      .6      .0      .0      1.0          218     98.0       10405


                                                                        Bottom
 F3=Exit   F10=Re-analyze   F11=Alternative view        F12=Cancel
 F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

Let's look at the Workload Report.

```
                              Display Workload Report

Period:   19971119
                     CPU     Thruput  -------Response Times (Secs)-------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
ATMSWL        1      4.9         155     13.6     13.9       .0       .0
QDEFAULT      1       .6         126       .6       .7       .0       .0
APPCWL        2     54.4        7200      1.3      1.3       .0       .0
ATMSWL        2     11.9         164     11.5     11.5       .0       .0
MIMIXWL       2      4.0          60     46.1     46.1       .0       .0
QDEFAULT      2      3.4         478      2.9      2.9       .0       .0
WTRWL         2       .1          60       .2       .2       .0       .0
-------------------------------------------------------------------------
Period:   19981119
                     CPU     Thruput  -------Response Times (Secs)-------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QDEFAULT      1       .7         152       .6       .6       .0       .0
APPCWL        2     96.2        7024      8.0      8.0       .0       .0
ATMSWL        2       .6           0      4.4      4.4       .0       .0
QDEFAULT      2      1.5           0    396.6    396.6       .0       .0
WTRWL         2       .1          72       .2       .2       .0       .0
-------------------------------------------------------------------------
Period:   19991119
                     CPU     Thruput  -------Response Times (Secs)-------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QDEFAULT      1       .8         182       .6       .6       .0       .0
APPCWL        2     95.6        8320      6.9      6.9       .0       .0
ATMSWL        2       .7           0      4.4      4.4       .0       .0
QDEFAULT      2      1.8           0    395.8    395.8       .0       .0
WTRWL         2       .1          86       .2       .2       .0       .0
-------------------------------------------------------------------------
Period:   20001119
                     CPU     Thruput  -------Response Times (Secs)-------
Workload    Type    Util    per Hour  Internal   Local     LAN      WAN
QDEFAULT      1      1.0         218       .6       .6       .0       .0
APPCWL        2     94.9       10271      5.8      5.8       .0       .0
ATMSWL        2       .9           0      4.5      4.5       .0       .0
QDEFAULT      2      2.1           0    395.1    395.1       .0       .0
WTRWL         2       .1         104       .2       .2       .0       .0
```

You can see the APPCWL workload takes up most of the CPU. The response time is 8.0 seconds in the year 1998 compared to 1.3 seconds currently.

6.5.1 Upgrading to a RISC Server

Now that you have seen the effect of an upgrade to the CISC system, you probably want to know the effect of an upgrade to a server model.

BEST/1 growth analysis normally models a traditional system-to-traditional system or a server system-to-server system. It does not automatically upgrade between a traditional system and a server system.

In this analysis, we specify the server configuration.

1. From the BEST/1 for AS/400 menu, select Option 1 (Work with BEST/1 models).
2. Select Option 5 (Work with) to work with the CSAPPCMDL model.
3. Select menu Option 10 (Configuration menu).
4. Select menu Option 1 (Change CPU and other resource values).
5. Change CPU Model to **2157**.
6. Change Main storage (MB) to **1024**.
7. Change Release level to **V3R7M0**.
8. Press Enter.

You get a message that says you have to calibrate the main storage size. Press F17 (Re-scale pool size).

9. Press Enter.

   You get the message:

   `RISC CPU SELECTED.  CHECK THAT EACH WORKLOAD HAS APPROPRIATE WORKLOAD TYPE.`

10. Read the help text for the message:

    This message calls your attention to the CISC to RISC consideration for migrating CPU intensive or disk I/O intensive applications to RISC. For the considerations, please refer to Appendix D of *AS/400 Client/Server Performance for Windows Client*, SG24-4526-01.

11. Press Enter.

    Now you get another set of messages; the first one says:

    `Disk IOP feature 6501 is not allowed on this CPU.`

    If you scroll through the remaining messages, you see a lot of disk and communication hardware that does not "migrate to" a RISC server model. You can complete a lot of manual delete and add hardware configuration steps with BEST/1, or...

12. On the Work with BEST/ Model display, select 10 (Configuration menu).

    The following display is shown.

```
                                Configuration

CPU Model . . . . . . . . . :   2157     Main stor (MB)  . . . . . . :   1024
                                         Main stor pools . . . . . . :   10
Disk IOPs . . . . . . . . . :   3
Disk ctls . . . . . . . . . :   6        Comm IOPs . . . . . . . . . :   6
Disk arms . . . . . . . . . :   19       Comm lines  . . . . . . . . :   9
ASPs  . . . . . . . . . . . :   1        Local WS ctls . . . . . . . :   2
                                         LAN ctls  . . . . . . . . . :   1
                                         WAN WS ctls . . . . . . . . :   8

Select one of the following:

     1. Change CPU and other resource values
     2. Work with disk resources
     3. Edit ASPs
     4. Edit main storage pools
     5. Work with communications resources

Selection or command
===> _________________________________________________________
F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel   F13=Check configuration
F17=Correct configuration           F24=More keys
```

13. On the Configuration menu, select F17 (Correct configuration).

    Let BEST/1 do your work.

    You get the following summary of changes!

```
                         Configuration Changes

The following changes have been made to your configuration:

1 9152 IOP(s) deleted
1 9162 IOP(s) created
2 6501 IOP(s) deleted
2 6502 IOP(s) created
15 6602-050 arm(s) created
2 9337-220 disk ctl(s) deleted
2 6602-050 disk ctl(s) created
```

It takes a long time for you to make all these changes manually.

Anytime BEST/1 does a major set of changes such as this, make sure you review the BEST/1 Recommendations report. You may find you want to make a few manual changes.

14. Return to the Work with BEST/1 Model display.
15. Select menu Option 7 (Analyze current model).
16. Introduce the parameters as shown in the following display:

```
                    Specify Growth of Workload Activity

Type information, press Enter to analyze model.
  Determine new configuration . . . . . . . . . .   Y    Y=Yes, N=No
  Periods to analyze  . . . . . . . . . . . . . .   4    1 - 10

  Period 1  . . . . . . . . . . . . .   19971119      Name
  Period 2  . . . . . . . . . . . . .   19981119      Name
  Period 3  . . . . . . . . . . . . .   19991119      Name
  Period 4  . . . . . . . . . . . . .   20001119      Name
  Period 5  . . . . . . . . . . . . .   Period 5      Name

            ------Percent Change in Workload Activity-------
Workload    Period 1  Period 2  Period 3  Period 4  Period 5
APPCWL            .0     100.0      50.0      50.0      20.0
ATMSWL            .0      20.0      20.0      20.0      20.0
MIMIXWL           .0      20.0      20.0      20.0      20.0
QCMN              .0      20.0      20.0      20.0      20.0
QDEFAULT          .0      20.0      20.0      20.0      20.0
WTRWL             .0      20.0      20.0      20.0      20.0

                                                                  Bottom
F3=Exit   F11=Specify total growth   F12=Cancel   F13=Display periods 6 to 10
F17=Analyze using ANZBESTMDL
```

These are the same "growth parameters" we used previously.

17. Press Enter to analyze.

Look at the Analysis Summary and Workload Report displays.

```
                              Display Analysis Summary

                       Stor    CPU  -Disk IOPs--  -Disk Ctls--  -Disk Arms--
 Period    CPU Model   (MB)    Util  Nbr    Util   Nbr    Util   Nbr   Util
 19971119  53S 2157    1024     9.7    3     8.6     6     1.7    19    8.1
 19981119  53S 2157    1024    16.9    3    14.1     6     3.1    19   15.0
 19991119  53S 2157    1024    24.3    3    19.7     6     4.5    19   21.9
 20001119  53S 2157    1024    35.2    3    27.7     6     6.6    19   32.3



           ----Inter Rsp Time----  -------Inter--------  -----Non-Inter------
 Period     Local    LAN     WAN   CPU Util    Trans/Hr  CPU Util    Trans/Hr
 19971119      .7     .0      .0       .7         282      9.0         7976
 19981119      .7     .0      .0       .8         338     16.1        15333
 19991119      .8     .0      .0      1.0         406     23.3        22721
 20001119      .8     .0      .0      1.2         487     34.0        33746

                                                                       Bottom
F3=Exit   F10=Re-analyze   F11=Alternative view         F12=Cancel
F15=Configuration menu     F17=Analyze multiple points  F24=More keys
```

We can see that the CPU utilization is 35.2% in the year 2000. This means 53S can support the business growth until the year 2000. However, there are many twinaxial workstations in use currently. The server model limitation of a maximum seven twinaxial connections prevents you from continuing with this upgrade approach.

6.5.2 Automatic Upgrade to a RISC System

1. Issue the STRBEST command using F4:

```
                          Start BEST/1 (STRBEST)

Type choices, press Enter.

BEST/1 data library  . . . . . .   A970121C      Name, *CURLIB
Performance data library . . . .   QPFRBANK      Name
Log member . . . . . . . . . . .   *NONE         Name, *NONE
Log library  . . . . . . . . . .   *BESTDTAL     Name, *BESTDTAL
User level . . . . . . . . . . .   *ADVANCED     *ADVANCED, *BASIC

F3=Exit  F4=Prompt  F5=Refresh  F12=Cancel  F13=How to use this display
F24=More keys
```

   In this last analysis, we want to upgrade 2044 to a RISC system model.

   Note: If you are modeling a traditional system and you want to have BEST/1 automatically upgrade to a RISC *server* system, you must manually change your CISC CPU model "Upgrade to family" parameter to a **Power Server** value.

   Once you make this change, BEST/1 can automatically upgrade from a traditional system to a server system when doing growth modeling.

2. Return to the BEST/1 for the AS/400 display.
3. Select menu Option 1 (Work with BEST/1 models).
4. Select Option 5 (Work with):

```
Opt    Model
5      CSAPPCMDL
```

This should be your base CISC model.

5. On the Work with BEST/1 Model display, select Option 5 (Analyze current model).

6. On the Work with Results display, select Option 5 for Analysis Summary and then use F11 to compare measured to predicted.

   Confirm that the model is still the base model.

7. Return to the Work with BEST/1 Model display and select Option 7 (Specify workload growth) and analyze model.

   Fill in the parameter values as shown in the following display:

```
                     Specify Growth of Workload Activity

 Type information, press Enter to analyze model.
   Determine new configuration . . . . . . . . . .  Y    Y=Yes, N=No
   Periods to analyze  . . . . . . . . . . . . . .  4    1 - 10

   Period 1  . . . . . . . . . . . . .  19971119      Name
   Period 2  . . . . . . . . . . . . .  19981119      Name
   Period 3  . . . . . . . . . . . . .  19991119      Name
   Period 4  . . . . . . . . . . . . .  20001119      Name
   Period 5  . . . . . . . . . . . . .  Period 5      Name

             ------Percent Change in Workload Activity-------
 Workload    Period 1  Period 2  Period 3  Period 4  Period 5
 APPCWL            .0     100.0      50.0      50.0      20.0
 ATMSWL            .0      20.0      20.0      20.0      20.0
 MIMIXWL           .0      20.0      20.0      20.0      20.0
 QCMN              .0      20.0      20.0      20.0      20.0
 QDEFAULT          .0      20.0      20.0      20.0      20.0
 WTRWL             .0      20.0      20.0      20.0      20.0

F3=Exit   F11=Specify total growth   F12=Cancel   F13=Display periods 6 to 10
F17=Analyze using ANZBESTMDL
```

8. Look at several of the result reports.

```
                            Display Analysis Summary

                            Stor     CPU  -Disk IOPs--  -Disk Ctls--  -Disk Arms--
 Period    CPU Model        (MB)    Util   Nbr    Util   Nbr    Util   Nbr    Util
 19971119  600 2136          320    74.6     1    56.4    15      .1    15    15.6
 19981119  510 2144          320    78.5     2    48.4     4      .8     4   105.9
 19991119  530 2151          512    69.0     8    12.7    20      .3    20    38.3
 20001119  530 2152          512    65.1     8    17.7    29      .3    29    39.8



           ----Inter Rsp Time----  -------Inter--------  -----Non-Inter------
 Period      Local     LAN    WAN  CPU Util     Trans/Hr  CPU Util    Trans/Hr
 19971119      6.0      .0     .0       5.2          282      69.4        7976
 19981119     64.1      .0     .0       3.7          338      74.8       15333
 19991119      2.6      .0     .0       2.7          406      66.3       22721
 20001119      1.9      .0     .0       2.1          487      62.9       33746

F3=Exit   F10=Re-analyze   F11=Alternative view        F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

You can see BEST/1 automatically selected 530#2152 to support the third year's business. The CPU utilization in the third year is 65.1. Notice that the number of Disk IOPs and Disk Arms has been increased.

Considering that a customer may want to upgrade to a recently announced model, we can change the CPU model to 2238 and re-analyze to get the following results:

```
                           Display Analysis Summary

                          Stor     CPU  -Disk IOPs--  -Disk Ctls--  -Disk Arms--
 Period    CPU Model      (MB)    Util   Nbr    Util   Nbr    Util   Nbr    Util
 19971119  640 2238       2560     9.8     3     8.8     6     3.8    19     4.4
 19981119  640 2238       2560    17.2     3    14.3     6     7.0    19     8.3
 19991119  640 2238       2560    24.7     3    20.0     6    10.2    19    12.5
 20001119  640 2238       2560    35.7     3    28.2     6    15.0    19    19.3



           ----Inter Rsp Time----  -------Inter--------  -----Non-Inter------
 Period      Local     LAN      WAN   CPU Util    Trans/Hr  CPU Util    Trans/Hr
 19971119       .5      .0       .0         .7         282       9.2        7976
 19981119       .5      .0       .0         .8         338      16.4       15333
 19991119       .5      .0       .0        1.0         406      23.7       22721
 20001119       .6      .0       .0        1.2         487      34.5       33746

F3=Exit   F10=Re-analyze   F11=Alternative view         F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

```
                          Display Workload Report

Period:   19971119
                      CPU     Thruput  -------Response Times (Secs)-------
Workload    Type     Util    per Hour  Internal   Local     LAN      WAN
ATMSWL        1        .6         155        .4        .6       .0       .0
QDEFAULT      1        .1         126        .3        .4       .0       .0
APPCWL        2       6.8        7200        .4        .4       .0       .0
ATMSWL        2       1.5         164       2.2       2.2       .0       .0
MIMIXWL       2        .5          60        .7        .7       .0       .0
QDEFAULT      2        .4         478        .2        .2       .0       .0
WTRWL         2        .0          60        .1        .1       .0       .0
-------------------------------------------------------------------------
Period:   19981119
                      CPU     Thruput  -------Response Times (Secs)-------
Workload    Type     Util    per Hour  Internal   Local     LAN      WAN
ATMSWL        1        .7         186        .4        .6       .0       .0
QDEFAULT      1        .1         152        .3        .4       .0       .0
APPCWL        2      13.5       14400        .4        .4       .0       .0
ATMSWL        2       1.8         197       2.4       2.4       .0       .0
MIMIXWL       2        .6          72        .8        .8       .0       .0
QDEFAULT      2        .5         573        .3        .3       .0       .0
WTRWL         2        .0          72        .1        .1       .0       .0
```

Period: 19991119

Workload	Type	CPU Util	Thruput per Hour	Response Times (Secs) Internal	Local	LAN	WAN
ATMSWL	1	.9	224	.4	.7	.0	.0
QDEFAULT	1	.1	182	.4	.4	.0	.0
APPCWL	2	20.3	21600	.4	.4	.0	.0
ATMSWL	2	2.1	237	2.6	2.6	.0	.0
MIMIXWL	2	.7	86	.9	.9	.0	.0
QDEFAULT	2	.6	688	.3	.3	.0	.0
WTRWL	2	.0	86	.1	.1	.0	.0

Period: 20001119

Workload	Type	CPU Util	Thruput per Hour	Response Times (Secs) Internal	Local	LAN	WAN
ATMSWL	1	1.1	268	.5	.8	.0	.0
QDEFAULT	1	.1	218	.4	.5	.0	.0
APPCWL	2	30.4	32400	.5	.5	.0	.0
ATMSWL	2	2.5	284	3.1	3.1	.0	.0
MIMIXWL	2	.9	104	1.0	1.0	.0	.0
QDEFAULT	2	.7	826	.3	.3	.0	.0
WTRWL	2	.0	104	.1	.1	.0	.0

The following graphs summarize the CPU utilization and response time.

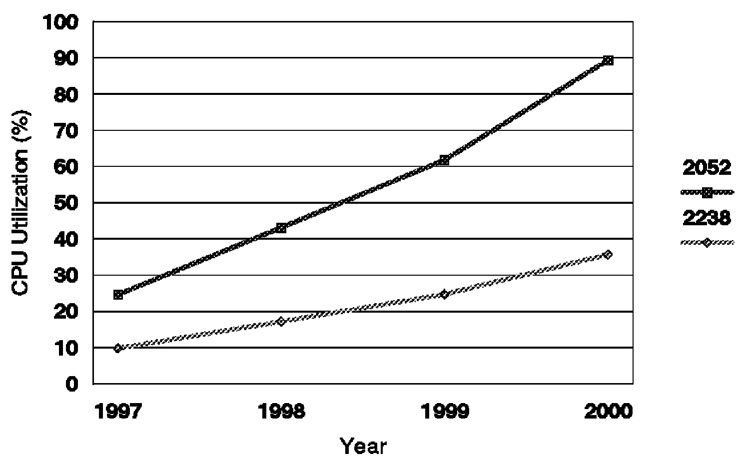


Figure 47. BEST/1 Graph of CPU Utilization

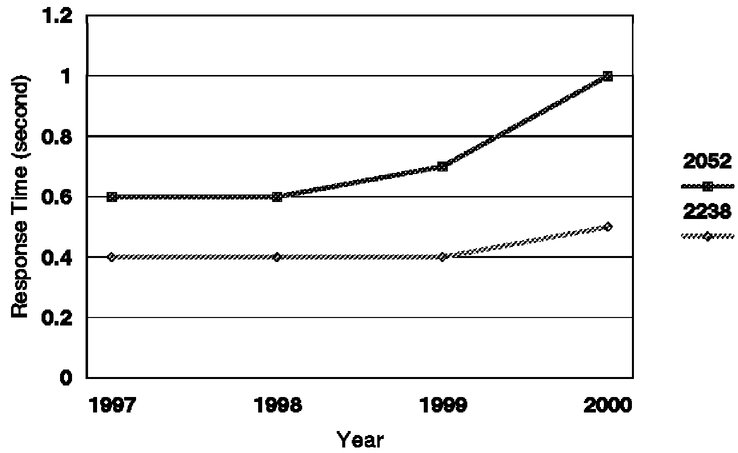


Figure 48. BEST/1 Graph of Response Time

6.6 Client/Server Capacity Planning Summary

You have validated and calibrated the model for "server jobs" that ran in their own storage pool and had to adjust for the fact that the Performance Monitor does not fully support "non-interactive transactions".

You have also seen how to specify growth without upgrading the current hardware configuration. Also you have upgraded the processor from a traditional system to a RISC server model system.

This chapter has calibrated the original BEST/1 model based on:

- Measured CISC V3R1 performance monitor data based an APPC application.
- A deep understanding of the application and the actual response times.

After calibrating the base model to correspond "close enough" to the actual customer environment, we performed various BEST/1 growth modeling analysis, keeping within the CISC family of systems and then analysis modeling growth to the RISC family of systems.

Chapter 7. Internet HTTP Capacity Planning

The AS/400 Internet Connection/400 software allows you to be an HTTP server on a TCP/IP network (for example, the Internet or an intranet). The HTTP server can be configured with numerous options that can help manage the server and optimize the performance. However, the AS/400 Performance Tools cannot really help much in finding out how much of a burden the serving activity is putting on the system and even much less, helping to do capacity planning.

The following sections show you how to collect the right information when doing Web serving, feeding this information into the analytical model (Best/1), and extrapolating the data.

7.1 Objectives

The objectives of this chapter are:

- Explain the usage of Web management software to assist in analyzing the usage and access trends of an HTTP server running on an AS/400 system.
- Show how to use the AS/400 Performance Monitor to gather performance data to assist in general HTTP performance management.
- Show how to use the BEST/1 Capacity Planning Tool to build a model based on HTTP performance data and predict the size of an AS/400 system required if the HTTP workload increases.

7.2 Overview

The first thing we must do is collect and analyze HTTP data. This shows us trends and also when our peek serving periods occur. It also shows, at a detail level, which URL objects are being accessed most frequently.

The second thing we must do is analyze the AS/400 performance using AS/400 Performance Tools reports.

Using the collected performance monitor data, chose a period of heavy demand as a basis for building a BEST/1 capacity model. This allows us to see the impact on AS/400 resources if the workload is increased and do some "what-if" analysis if we change the AS/400 configuration.

7.3 HTTP Access Log Analysis

1. Sign on to the AS/400 system.
2. Check if your AS/400 system is configured to collect Access Log information and the format that this data is written to disk. On the AS/400 command line, type:

   ```
   WRKHTTPCFG
   ```

   Press the Enter key.
3. Scroll down to the line where you see the word Accesslog. This tells us if access logging is turned on or not. To get useful information about who is accessing your HTTP server and which objects are being accessed, you must

© Copyright IBM Corp. 1998

have access logging turned on. Next look at the logging format. Usually this is best set to a common format. This means that it can be used on multiple platforms as its easy to convert (for example, EBCDIC to ASCII, which is shown later).

The following display is shown:

```
                      Work with HTTP Configuration
                                                         System:   RC
  Configuration name . . . . . . . :   CONFIG

 Type options, press Enter.
   1=Add   2=Change   3=Copy   4=Remove   5=Display   13=Insert

       Sequence
 Opt    Number    Entry

        01710     #  1 - AccessLog ACCESSLOG
        01720     # Access Log file, ACCESSLOG, created in
        01730     #    Integrated File System directory, httplog.
        01740     #  2 - AccessLog /httplog/accesslog
        01750     #
        01760     AccessLog acc01 5000
        01770     #
        01780     # Syntax:
        01790     #   AccessLog              <Access_Log_FileName>
        01800     #   ErrorLog               <Error_Log_FileName>
                                                                    M
 F3=Exit   F5=Refresh   F6=Print List   F12=Cancel   F17=Top   F18=Bottom
 F19=Edit Sequence
```

The access log file is created in library QUSRSYS and has an initial size of 5000 Kilobytes.

4. Use DSPPFM to display some of the data in this file to your workstation. It is in a document type format, which makes it easy for the next step.

5. Create a folder in QDLS with the name of Webstats (or similar). Use the WRKFLR command and option 1, to create. After creating the folder, change its security to PUBLIC *USE or PUBLIC *ALL.

6. Now we are going to convert the AS/400 Access Log (which is in EBCDIC) to a PC document (which is in ASCII format) and can be used by our HTTP management tool. Use the Copy to PC Document (CPYTOPCD) command to do this. The following display is shown:

```
                     Copy To PC Document (CPYTOPCD)

 Type choices, press Enter.

 From file  . . . . . . . . . . .   acc01        Name
   Library  . . . . . . . . . . .     qusrsys    Name, *LIBL, *CURLIB
 To folder  . . . . . . . . . . .   Webstats

 From member  . . . . . . . . . .   *FIRST       Name, *FIRST
 To document  . . . . . . . . . .   accesslog01  Name, *FROMMBR
 Replace document . . . . . . . .   *NO          *NO, *YES
 Translate table  . . . . . . . .   *DFT         Name, *DFT, *NONE
   Library  . . . . . . . . . . .                Name, *LIBL, *CURLIB
 Format of PC data  . . . . . . .   *TEXT        *TEXT, *NOTEXT








 F3=Exit   F4=Prompt   F5=Refresh   F10=Additional parameters   F12=Cance
```

7. The next step is to copy this file from the AS/400 shared folders directory to your PC. If you are using Win95, this is best done through a network drive to \QDLS\Webstats\. Create a folder on your PC in the root directory called alog and copy the file alog.log into this directory.
8. Although, you can analyze the log file directly and try to find the number of hits or visits per hour, it is much more convenient to obtain one of the many Web Statistic Programs available on the market place. Please search on the Web for more information. Also, demo copies can be obtained through the Internet (search on Web usage with yahoo), but if you use it for more than 30 days, you must purchase it.

Run the Web statistics program against your copied log file and determine how many hits or visits per hour you had on your server. Note that number for later use in the capacity planning section of this chapter.

The following images are taking from the Web to illustrate what kind of information can be expected from the output of some Web statistic programs:

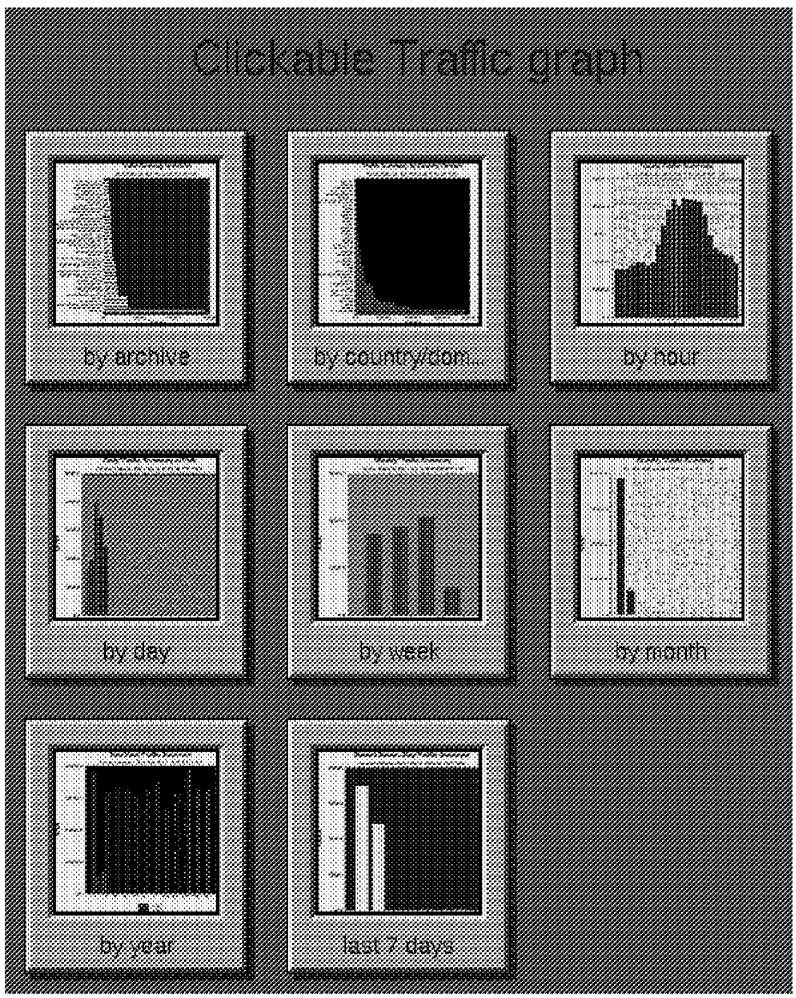


Figure 49. Example of HTTP Statistics from a Web Statistics Program

As you can see, there are many different views on the log file and using such a tool makes it easy to select the right time frame of heavy use of our server.

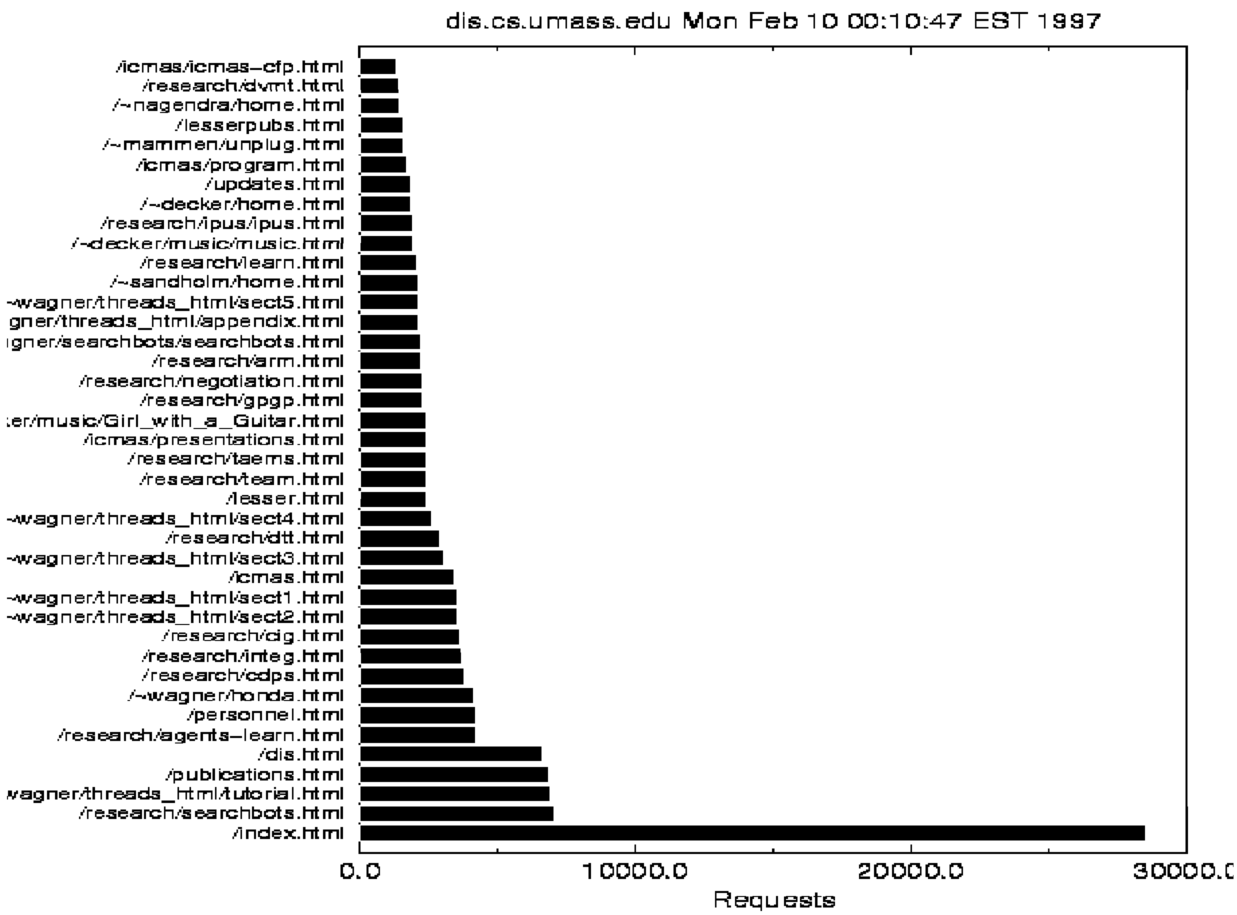


Figure 50. Example HTTP Statistics Based on HTML Pages

The previous information has nothing to do with performance, but gives a good overview of what pages have been requested and what pages are used most of the time. This information might be useful to optimize the most used pages for performance, minimize the number of icons, number of bytes, number of frames, and so on.

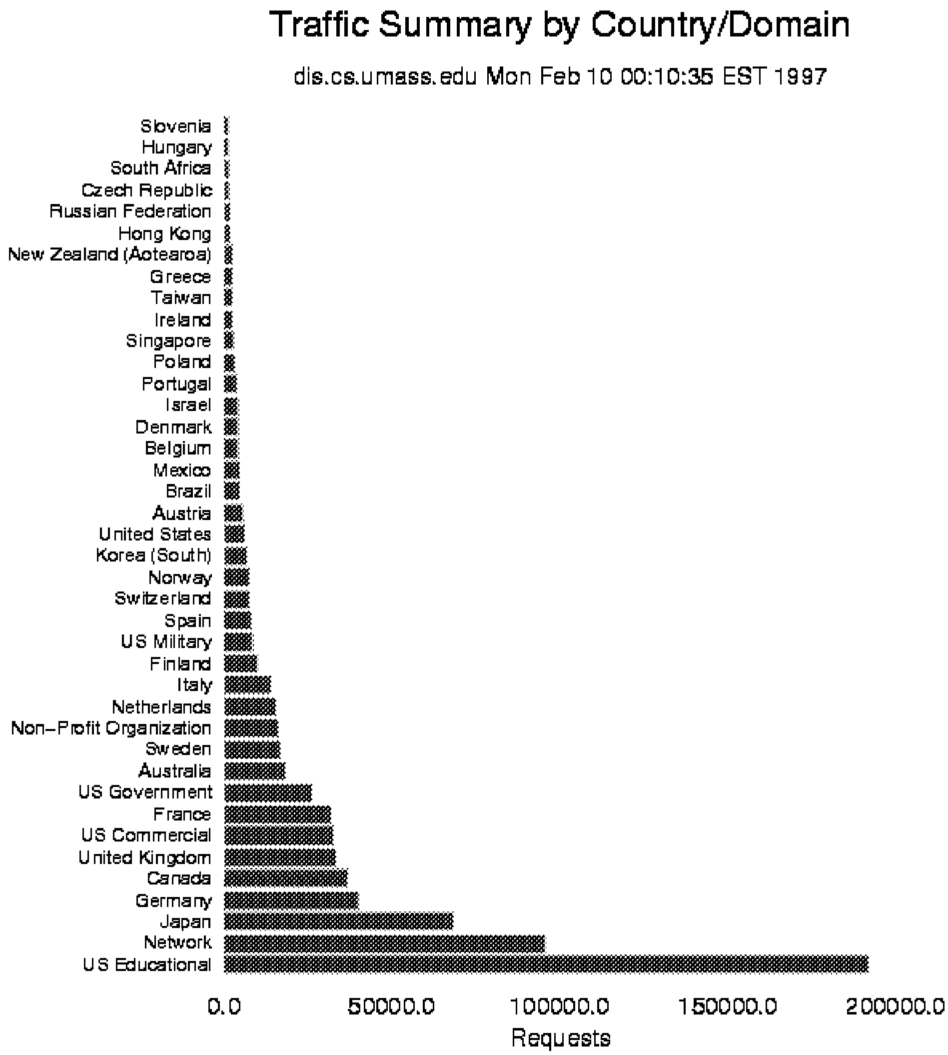


Figure 51. Example HTTP Statistics Based on Country Usage

The statistic based on the country usage might be helpful to determine if a second server (shadow) in the country where most of the requests are coming from is beneficial.

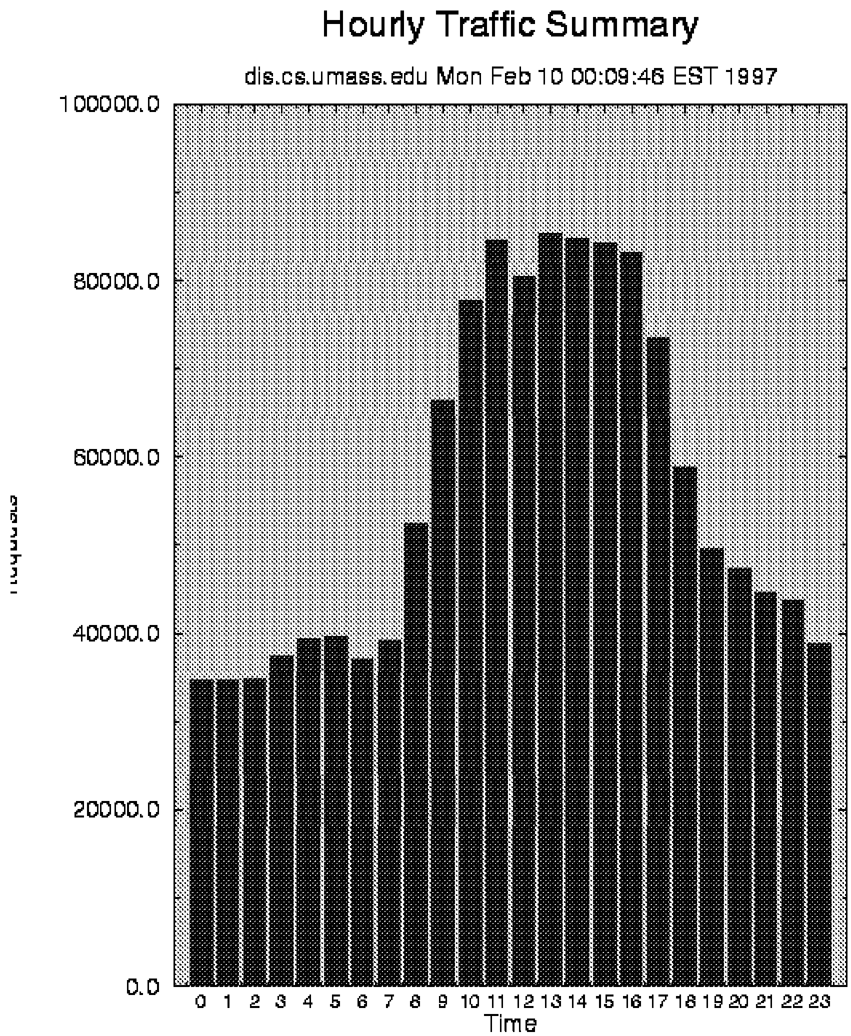


Figure 52. Example HTTP Statistics Based on Hour

The hourly statistic can be used to determine the right time of the day to find the peak server workload for capacity planning.

It is important to note the number of hits or visits per hour so we can use it as a base for modeling in the capacity planning section of this chapter.

7.4 Using AS/400 Performance Tools Data

In the first part, we looked at the performance of the AS/400 HTTP server at the micro level based on data obtained from the Access Log. This was acquired at the object level, but showed little information as to how it affected the AS/400 system. Now look at performance data collected by the performance monitor while the AS/400 system was serving HTTP accesses. The data we use was obtained during a period of time when only HTTP serving was being performed. There were no other user applications running, and no interactive work was being performed.

1. On the command line, enter the following command:

   ```
   GO PERFORM
   ```

   The Performance Tools Main Menu is shown.

2. Select Option 3, Print Performance Report, which gives you a selection of Performance Data Library members.
3. First, run a System Summary Report on the data collected.
4. On the next display, use F6 to select all Sections.
5. Select the times where the system had the most hits.
6. Press Enter twice; this submits the job to batch. When the job is finished (use WRKSBMJOB to ascertain its status), review the spooled file on your display.
7. While your job is running, perform the same basic steps from 3-7, but this time use Option 2, Component Report. Select the same time period as in the System Report.
8. Now look at the two performance reports and, in particular, a few places where it is worth checking to see if the system is correctly tuned, and how much resources are being used.
9. Bring up the AS/400 System Report that you just ran (use WRKSPLF to find all of the spooled files created by your team). Notice first that no interactive transactions were processed. Why?
10. Most of the work is being done in Batch at priority 25. This is the HTTP servers.
11. Look at the total CPU utilization.
12. Now let's look at the Component Report. Select one of the 15-minute samples that had the greatest amount of CPU used.

    You can identify the HTTP server jobs by looking for the following values:

    ```
    QHTTPnnnnn
    ```

    ```
    The user-id being used is
    ```

    ```
    QTMHHTTP
    ```

13. In the IOP utilization of the Component, it shows the utilization for all IOPs, including the Comms IOPs. Look at the number of bytes transmitted by the IOP.

7.5 AS/400 HTTP Capacity Planning

This section focuses on building a capacity planning model using BEST/1 that takes performance data collected from an AS/400 production HTTP server. We will build the model, taking only those jobs that have an impact on the servers capacity. This information was collected from the Component Report in Part 2. We assume that we estimated our server usage will increase by 10 times that of today, and we want to know if our server can handle this amount, or if we need to upgrade.

1. From the AS/400 command line, use the AS/400 command:

    ```
    STRBEST to get to the BEST/1 capacity planning tool:

    STRBEST BESTDTALIB(libary)
    PFRDTALIB(library)
              LOGMBR(*NONE) LOGLIB(*BESTDTAL)
    USRLVL(*ADVANCED)
    ```

2. The following display is shown:

```
                              BEST/1 for the AS/400

 Select one of the following:

      1. Work with BEST/1 models

      5. Create BEST/1 model from performance data

     10. Work with results

     50. About BEST/1

     60. More BEST/1 options




 Selection or command
 ===>

 F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel
```

3. Select Option 1, Work with BEST/1 models.
4. The following display is shown:

```
                             Work with BEST/1 Models

 Library . . . . .   QGPL         Name

 Type options, press Enter.
   1=Create   3=Copy   4=Delete   5=Work with   6=Print   7=Rename

 Opt  Model      Text                                   Date      Tim
  1   HTTPSERV

   (No models in library)






 Command
 ===>
 F3=Exit   F4=Prompt   F5=Refresh        F9=Retrieve   F12=Cancel
 F15=Sort by model     F16=Sort by text   F19=Sort by date and time
```

Create a new model called HTTPSERV. Make sure you create it in your library.

5. The following display is shown:

```
                             Create BEST/1 Model

Select one of the following:

     1. Create from performance data
     2. Create from predefined and user-defined workloads

Selection
   1

F3=Exit   F12=Cancel
```

6. Select option 1 to create from performance data. The following display is shown:

```
                  Create BEST/1 Model from Performance Data

Model . . . . . . . . . . . . . . . :   HTTPSERV

Type choices, press Enter.  Use *SLTHOUR to select an hour-long time per
  use *SLTITV to select select first and last interval of a one to two h
  time period.  The time period selected should be representative of you
  processing activity.

Text . . . . . . . . . . . . .   Measured HTTP serving

Performance data:
  Member . . . . . . . . . . .   HTTPSERV     Name, F4 for list
  Library  . . . . . . . . . .   QPFRDATA     Name

Start time . . . . . . . . . .   *SLTITV      Time, *FIRST, *SLTHOUR, *S
Start date . . . . . . . . . .   *FIRST       Date, *FIRST

Stop time  . . . . . . . . . .   *LAST        Time, *LAST
Stop date  . . . . . . . . . .   *LAST        Date, *LAST


F3=Exit   F4=Prompt   F12=Cancel
```

Fill in the text and use F4 (Prompt) to obtain the correct performance member. Use the *SLTITV option for start time so that on the next display, you can select the appropriate intervals to start and stop.

7. The following display is shown:

```
                              Select Time Interval

Library  . . . . . . . :   QPFRDATA      Performance member . . :   HTTP

Type option, press Enter.  Select first and last interval.
  1=Select

                            ---Transaction---   --CPU Util---   I/Os per
Opt   Date       Time        Count   Rsp Time   Total   Inter   Sync   A
 1    01/14/98   16:26:58       0        .0       6       0       3
      01/14/98   16:27:43       1        .0       4       0       6










F3=Exit   F12=Cancel   F15=Sort by interval        F16=Sort by count
F17=Sort by rsp time   F18=Sort by total CPU util   F19=Sort by total I/
```

Type a 1 to select the start and stop times for the performance data that you want to use as the time to model. This should match a busy period of activity.

8. Select Classify jobs into workloads as shown in the following display:

```
                                 Classify Jobs

Select one of the following:

     1. Use default job classification
     2. Classify jobs into workloads
     3. Use existing job classifications















Selection
     2

F3=Exit   F12=Cancel
```

9. Choose the Job name category, which is category number 3.
10. On the next display, press F9 to display values from performance data.
11. We are now going to create two workloads (HTTPSERV and QDEFAULT). To do this, first type in a workload name. Then scroll down to select the jobs for that workload. You only need to do it for HTTPSERV because all the remaining jobs automatically go into QDEFAULT. In our model, we want to grow QHTTPSERV and QDEFAULT.

```
                         Assign Jobs to Workloads

Workload . . . . . . . . . . . . . . . . .

Type options, press Enter.  Unassigned jobs become part of workload QDEF
  1=Assign to above workload   2=Unassign

                              Number of       CPU       I/O
Opt  Workload    User ID    Transactions    Seconds    Count
     HTTPSERV                       0        10.203     4883
                 QGATE              0          .000        0
                 QMSF               0          .000        0
                 QPGMR              0          .258      222
                 QSNADS             0          .000        0
                 QSPLJOB            0          .000        0
                 QSYS               0          .700       84
                 QTCP               0          .024        1
     HTTPSERV    QTMHHTTP           0        41.113     3603
                 QUSER              0          .668      211
                 SCHIMU             0          .000        0

F3=Exit   F12=Cancel      F15=Sort by workload      F16=Sort by user id
F17=Sort by transactions   F18=Sort by CPU seconds   F19=Sort by I/O cou
```

Press Enter when you are finished and the following display is shown.

```
                         Specify Paging Behaviors

Type choices, press Enter.

               Paging Behavior
Workload        (F4 for list)
QDEFAULT          *GENERIC
HTTPSERV          *GENERIC

F3=Exit   F4=Prompt   F12=Cancel
```

12. Accept the defaults on this display.
13. Enter the number of hits per hour in the rightmost column and change the transaction type of HTTPSERV workload to *NONE as shown in the following display:

```
                     Define Non-Interactive Transactions

Job classification category  . . . . . . :   User ID


Type choices, press Enter.

               ---Activity Counted as Transaction---     Total Transacti
Workload          Type                 Quantity        when Type = *N
QDEFAULT          *LGLIO                 100.0                      0
HTTPSERV          *NONE                  100.0                  12000










Type:  *LGLIO, *CMNIO, *CPUSEC, *PRINT, *NONE

F3=Exit   F12=Cancel
```

14. Save your newly acquired Job Classification data in a file, just in case you want to reuse it.

15. Type in data similar to the next display screen and press Enter to build your model. Use the WRKSBMJOB command to determine when your job has completed, or use F5 to see when your model is built.

16. When your model has been built, use option 5 to work with your model:

```
                           Work with BEST/1 Models

Library . . . . .   QGPL         Name

Type options, press Enter.
  1=Create   3=Copy   4=Delete   5=Work with   6=Print   7=Rename

Opt  Model       Text                                  Date      Tim

 5   HTTPSERV    Measured HTTP serving                 01/14/98  17:









Command
===>
F3=Exit   F4=Prompt   F5=Refresh        F9=Retrieve   F12=Cancel
F15=Sort by model     F16=Sort by text   F19=Sort by date and time
The model has been created
```

The following display is shown:

```
                              Work with BEST/1 Model

Performance data . . . :   QPFRDATA (HTTPSERV)
Model/Text . . . . . . :   HTTPSERV    Measured HTTP serving


Select one of the following:

     1. Work with workloads
     2. Specify objectives and active jobs

     5. Analyze current model
     6. Analyze current model and give recommendations
     7. Specify workload growth and analyze model

    10. Configuration menu
    11. Work with results

                                                                          M
Selection or command
===>
F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel   F15=Save current model
F17=Analyze using ANZBESTMDL       F22=Calibrate model   F24=More keys
```

17. Using option 5, analyze the model. On the following display, take option 5 to display the Analysis Summary:

```
                              Work with BEST/1 Model

Performance data . . . :   QPFRDATA (HTTPSERV)
Model/Text . . . . . . :   HTTPSERV    Measured HTTP serving


Select one of the following:

     1. Work with workloads
     2. Specify objectives and active jobs

     5. Analyze current model
     6. Analyze current model and give recommendations
     7. Specify workload growth and analyze model

    10. Configuration menu
    11. Work with results

                                                                          M
Selection or command
===> 5
F3=Exit   F4=Prompt   F9=Retrieve   F12=Cancel   F15=Save current model
F17=Analyze using ANZBESTMDL       F22=Calibrate model   F24=More keys
```

18. The following display is shown:

```
                              Work with Results

Printed report text . . . . . .   Measured HTTP serving



Type options, press Enter.
  5=Display   6=Print

Opt     Report Name
        Measured and Predicted Comparison
 5      Analysis Summary
        Recommendations
        Workload Report
        ASP and Disk Arm Report
        Disk IOP and Disk Arm Report
        Main Storage Pool Report
        Communications Resources Report
        All of the above


F3=Exit   F12=Cancel   F14=Select saved results   F15=Save current resul
F18=Graph current results   F19=Append saved results   F24=More keys
Model has been analyzed
```

Use option 5 on the Work with Results display to see the Measured and Predicted Comparison display. Make sure that the CPU, disk I/Os, and non-interactive transactions on this display are approximately equal (see the following display).

```
                      Measured and Predicted Comparison

                                         Measured      Predicted
Total CPU util . . . . . . . . . . . :        6.3            6.3
Disk IOP util  . . . . . . . . . . . :        3.3            1.8
Disk arm util  . . . . . . . . . . . :        1.5            2.1
Disk IOs per second  . . . . . . . . :       16.9           16.8

LAN IOP util . . . . . . . . . . . . :        4.5             .0
LAN line util  . . . . . . . . . . . :         .0             .0
WAN IOP util . . . . . . . . . . . . :        1.6            2.0
WAN line util  . . . . . . . . . . . :         .0             .0

Interactive:
  CPU util . . . . . . . . . . . . . :         .0             .0
  Int rsp time (seconds) . . . . . . :         .0             .0
  Transactions per hour  . . . . . . :          1              0
Non-interactive thruput  . . . . . . :      24061          24066


Performance estimates -- Press help to see disclaimer.
F3=Exit   F6=Print   F9=Work with spooled files   F12=Cancel
```

19. Press F12 until you return to the display shown in step 16 (Work with BEST/1 Models) and select option 7.

20. Use F11 on the following display to enable the Growth by Workload feature. Now type in some growth details by workload and then press Enter to analyze the model:

```
                    Specify Growth of Workload Activity

 Type information, press Enter to analyze model.
   Determine new configuration . . . . . . . . . .  Y    Y=Yes, N=No
   Periods to analyze  . . . . . . . . . . . . . .  1    1 - 10

   Period 1  . . . . . . . . . . . . .  Period 1      Name
   Period 2  . . . . . . . . . . . . .  Period 2      Name
   Period 3  . . . . . . . . . . . . .  Period 3      Name
   Period 4  . . . . . . . . . . . . .  Period 4      Name
   Period 5  . . . . . . . . . . . . .  Period 5      Name

             ------Percent Change in Workload Activity-------
 Workload    Period 1  Period 2  Period 3  Period 4  Period 5
 HTTPSERV        400       200       100        50       0.0
 QDEFAULT        400       200       100        50       0.0




 F3=Exit   F11=Specify total growth   F12=Cancel   F13=Display periods 6
 F17=Analyze using ANZBESTMDL
```

We assume compounding growth rates of 400% in the second year, then 200%, 100%, 50%.

21. Look at the analysis summary option shown in the following display. Note that the modeller has automatically selected new models for you as required.

```
                          Display Analysis Summary

 CPU Model / release level  . . . . . :   50S 2121     V4R1M0
 Main Storage . . . . . . . . . . . . :   256         MB

                                          Quantity    Predicted Util
 CPU  . . . . . . . . . . . . . . . . :      1            26.7
 Disk IOPs  . . . . . . . . . . . . . :      2             7.5
 Disk ctls  . . . . . . . . . . . . . :      3              .7
 Disk arms  . . . . . . . . . . . . . :      6             9.6
 Local WS ctls  . . . . . . . . . . . :      1              .0

                                                                  M
                                                 Interactive   Non-intera
 CPU utilization % . . . . . . . . . . . . . . :  .0            26.7
 Transactions per Hour . . . . . . . . . . . . :  0             120330
 Local response time (seconds) . . . . . . . . :  .0            .0
 LAN response time (seconds) . . . . . . . . . :  .0            .5
 WAN response time (seconds) . . . . . . . . . :  .0            .0

 Performance estimates -- Press help to see disclaimer.
 F3=Exit   F10=Re-analyze      F12=Cancel   F15=Configuration menu
 F17=Analyze multiple points   F18=Specify objectives   F24=More keys
```

22. Now that you know your way around BEST/1, we manually change the model to a newer server model with a similar configuration to your previous, and see how the server model performs. Use F15=Configuration menu on this display as a fast path to get to the configuration menu.

23. Now go back to the workload display and change the automatic update option from Y (yes) to N (no). The following display is shown:

```
                     Specify Growth of Workload Activity

Type information, press Enter to analyze model.
  Determine new configuration . . . . . . . . . . .   N    Y=Yes, N=No
  Periods to analyze  . . . . . . . . . . . . . . .   1    1 - 10

  Period 1  . . . . . . . . . . . . . .  Period 1      Name
  Period 2  . . . . . . . . . . . . . .  Period 2      Name
  Period 3  . . . . . . . . . . . . . .  Period 3      Name
  Period 4  . . . . . . . . . . . . . .  Period 4      Name
  Period 5  . . . . . . . . . . . . . .  Period 5      Name

            ------Percent Change in Workload Activity-------
Workload    Period 1  Period 2  Period 3  Period 4  Period 5
HTTPSERV      400.0     200.0     100.0      50.0        .0
QDEFAULT      400.0     200.0     100.0      50.0        .0





F3=Exit   F11=Specify total growth   F12=Cancel   F13=Display periods 6
F17=Analyze using ANZBESTMDL
```

Now review your results to make sure that the new AS/400 system can handle the projected workloads.

Chapter 8. SAP R/3 on AS/400 System

SAP-AG (Systems, Applications and Products) R/3 software is based on client/server architecture. It is designed as an "open system" for use on operating systems from a variety of vendors. The SAP R/3 application product range provides support for enterprise-wide resource planning (ERP) solutions and includes many specific solutions for industries such as:

- Aerospace and defense
- Automotive
- Banking
- Chemicals
- Engineering and construction
- Healthcare
- Insurance
- Oil and gas
- Public sector
- Retail
- Telecommunications
- Utilities

R/3 on the AS/400 system can run as one of the following implementations:

- **Two-tier Implementation** (centralized):

  This has a PC for a presentation layer and an AS/400 RISC-based system for application processing and database support.

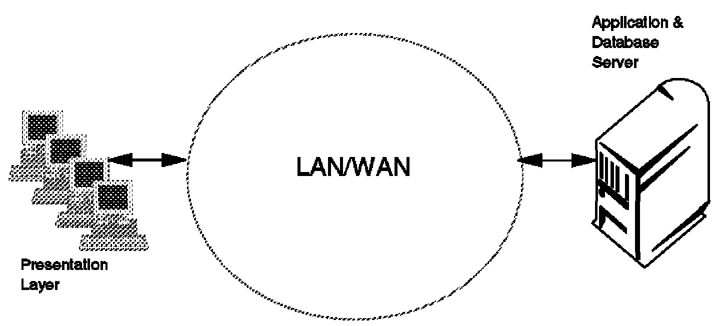


Figure 53. SAP R/3 Two-Tier Implementation

- **Three-tier Implementation** (client/server):

© Copyright IBM Corp. 1998

This has a PC for a presentation layer, one or more AS/400 RISC-based systems for application processing (the client), and a single AS/400 RISC-based system functioning as a database server (the server). The application servers and the database servers must be linked through fiber optics hardware and Optimover services.

In a three-tier implementation of SAP R/3, a capacity planning project must include the application servers as well as the database server.

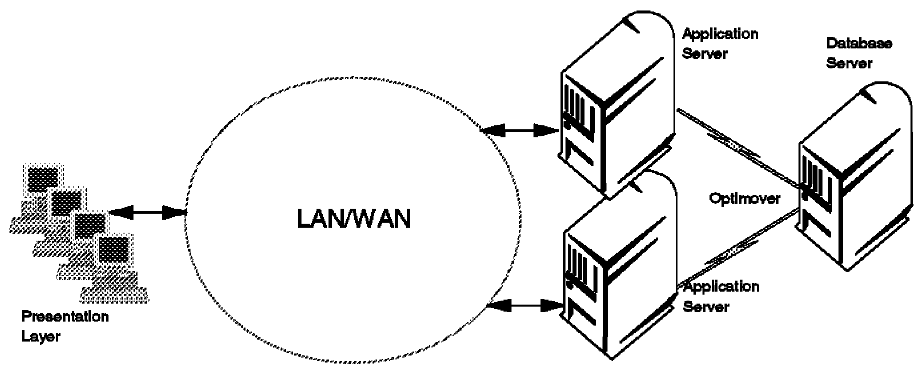


Figure 54. SAP R/3 Three-Tier Implementation

The connection between the presentation layer and the host AS/400 systems can be through a LAN or WAN.

For more information on the SAP application, please contact SAP-AG or your local SAP representative. For more information on the implementation of SAP R/3 on the AS/400 system, pleases refer to *SAP R/3 Implementation for AS/400*, SG24-4672.

If you intend collecting performance and workload information on the AS/400 system and the SAP R/3 application, you need to have:

- Access to a 5250 session on the AS/400 system
- Access to a workstation with SAPGUI
- An AS/400 user profile and password
- An SAP R/3 user profile and password
- An SAP R/3 client number

8.1 AS/400 Performance Data Collection

Use the standard AS/400 STRPFRMON command to collect AS/400 performance data. Enter a valid library name to collect the performance data and change the default time interval to five minutes. If you are uncertain of the period over which performance data is to be collected, change the number of hours data is to be collected to include the most likely period. Also make certain that you do not collect trace data. Trace data is not required for BEST/1 modelling and only increases system activity and the volume of data collected.

```
                     Start Performance Monitor (STRPFRMON)

Type choices, press Enter.

Member . . . . . . . . . . . . .   .*GEN.   Name, *GEN
Library  . . . . . . . . . . . .   .YOUR LIB.Name
Text 'description' . . . . . . .   .Your Company Name.

Time interval (in minutes) . . .   .5.    5, 10, 15, 20, 25, 30, 35
Stops data collection  . . . . .   *ELAPSED      *ELAPSED, *TIME, *NOMAX
Days from current day  . . . . .   0             0-9
Hour . . . . . . . . . . . . . .   .2.    0-999
Minutes  . . . . . . . . . . . .   0             0-99
Data type  . . . . . . . . . . .   *ALL          *ALL, *SYS
Trace type . . . . . . . . . . .   .*NONE.*NONE, *ALL
Dump the trace . . . . . . . . .   *YES          *YES, *NO
Job trace interval . . . . . . .   .5            .5 - 9.9 seconds
Job types  . . . . . . . . . . .   *DFT          *NONE, *DFT, *ASJ, *BCH...
               + for more values
Start database monitor . . . . .   *NO           *YES, *NO
                                                                       More....
 F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
 F24=More keys
```

8.2 SAP R/3 Performance Data Collection

The SAP R/3 application services include well-developed facilities to measure workload and response times. The unit of work used by SAP R/3 is referred to as a **dialog step**. In an interactive environment, a dialog step is equivalent to a screen change. In addition to this, SAP R/3 recognizes dialog steps for other workloads such as batch and update processing. The capacity planning unit of measure (CPUM) for AS/400 resource requirement estimation we use in this example is the SAP R/3 **dialog step**.

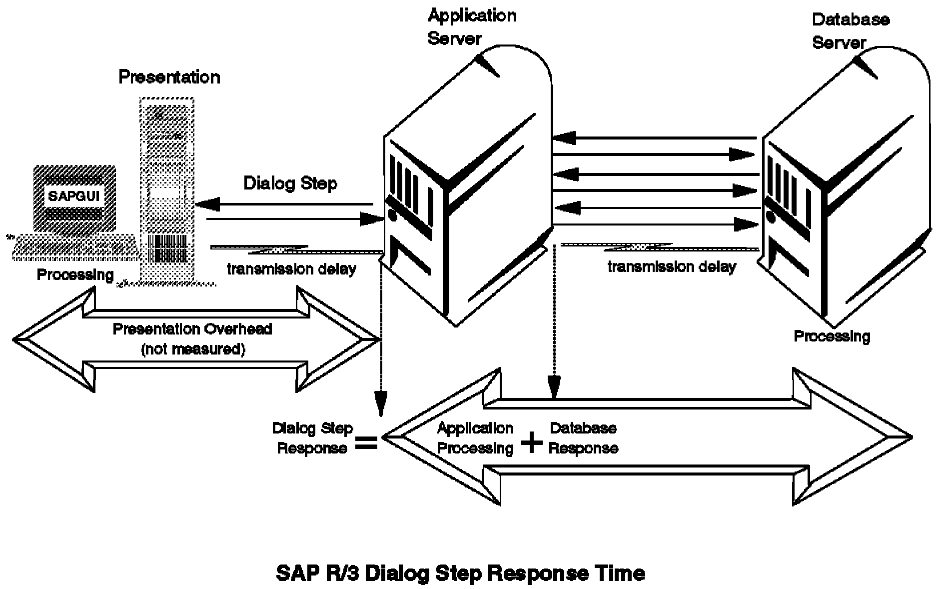


Figure 55. SAP R/3 Response Time

The database time, application processing time, and other delays are measured by the SAP application server using the program **SAPOSCOL**, which is started automatically by SAP R/3.

In the case of a two-tier (or centralized) system, the application server measures the CPU time and the database time on the same machine. In a three-tier system, the database time measured by the application server includes any AS/400 processing overhead (on the application server and database server), and any communication time over the Optimover services. Considering the high speeds of the fiber optic buses involved, the communication delay is insignificant.

The time between a dialog step request reaching the application server and the response exiting the application server is reported as the **dialog step response time**. This value can be displayed on the presentation display through SAPGUI.

However, there is no facility to measure any delay between the application server and the presentation layer on the PC. Thus, any references to SAP R/3 response time refers to the AS/400 servers (application and database) only and does not include any client processing time or communications delays.

SAP R/3 workload and response time information is retrieved through SAP R/3 Computer Center Management System (CCMS) functions, which are outlined in the following sections.

8.2.1 SAP R/3 Work Processes

Sign on to the SAP application and enter the transaction **SM50**. The following display shows the work processes that are active on the specific SAP R/3 instance of the application server you are signed on to.

Note: Only the SAP Work Process list and the associated WRKPID commands may be completed during the performance data collection period. All other information referred to in Section 8.2.2, "SAP R/3 Workload Analysis" on page 137 **must be collected after the data collection period**.

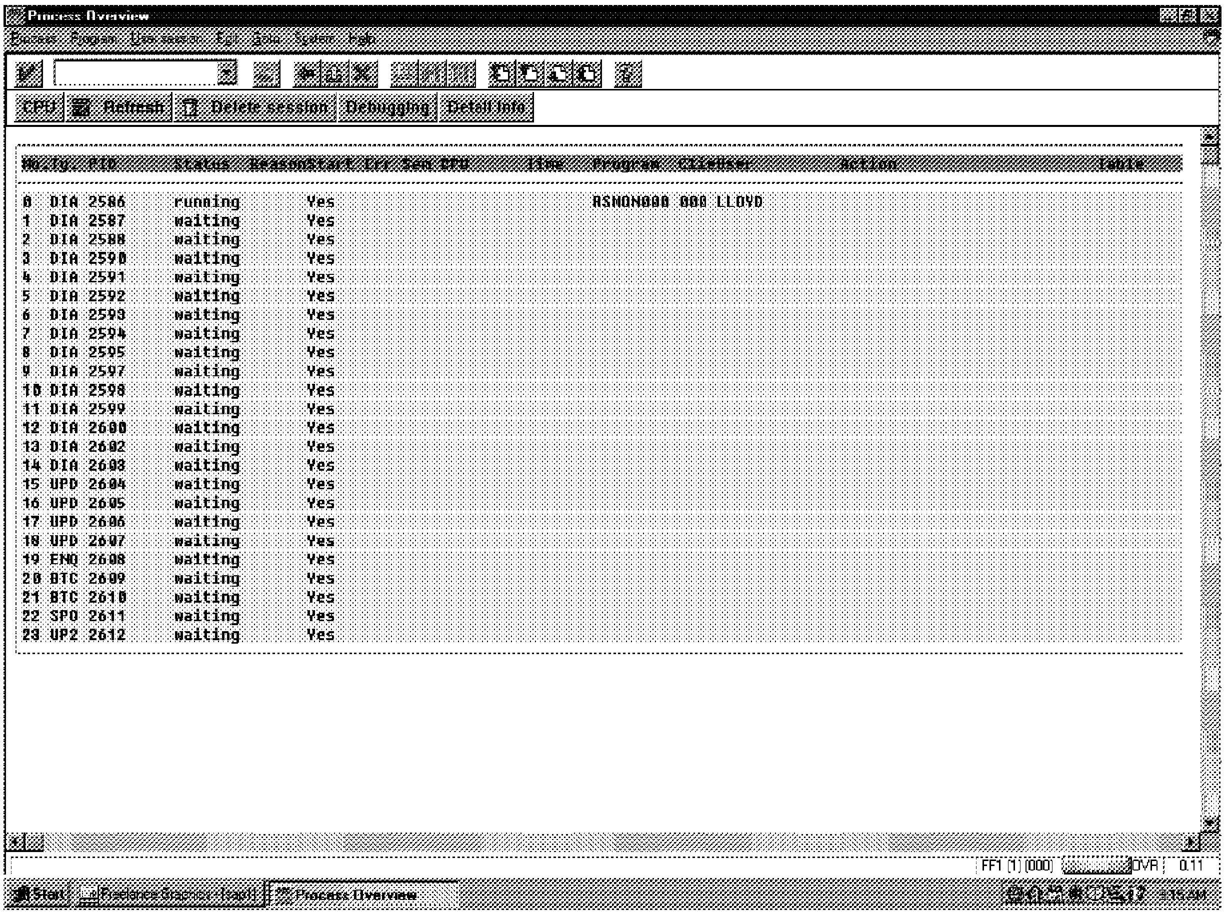


Figure 56. SM50 - Work Processes

The following figures show you how to save the SAP R/3 performance data sets to a file on your PC:

- Figure 57 on page 134
- Figure 58 on page 135
- Figure 59 on page 136

Select **System -> List -> Save -> Local File** with the mouse as indicated in Figure 57 and press Enter.

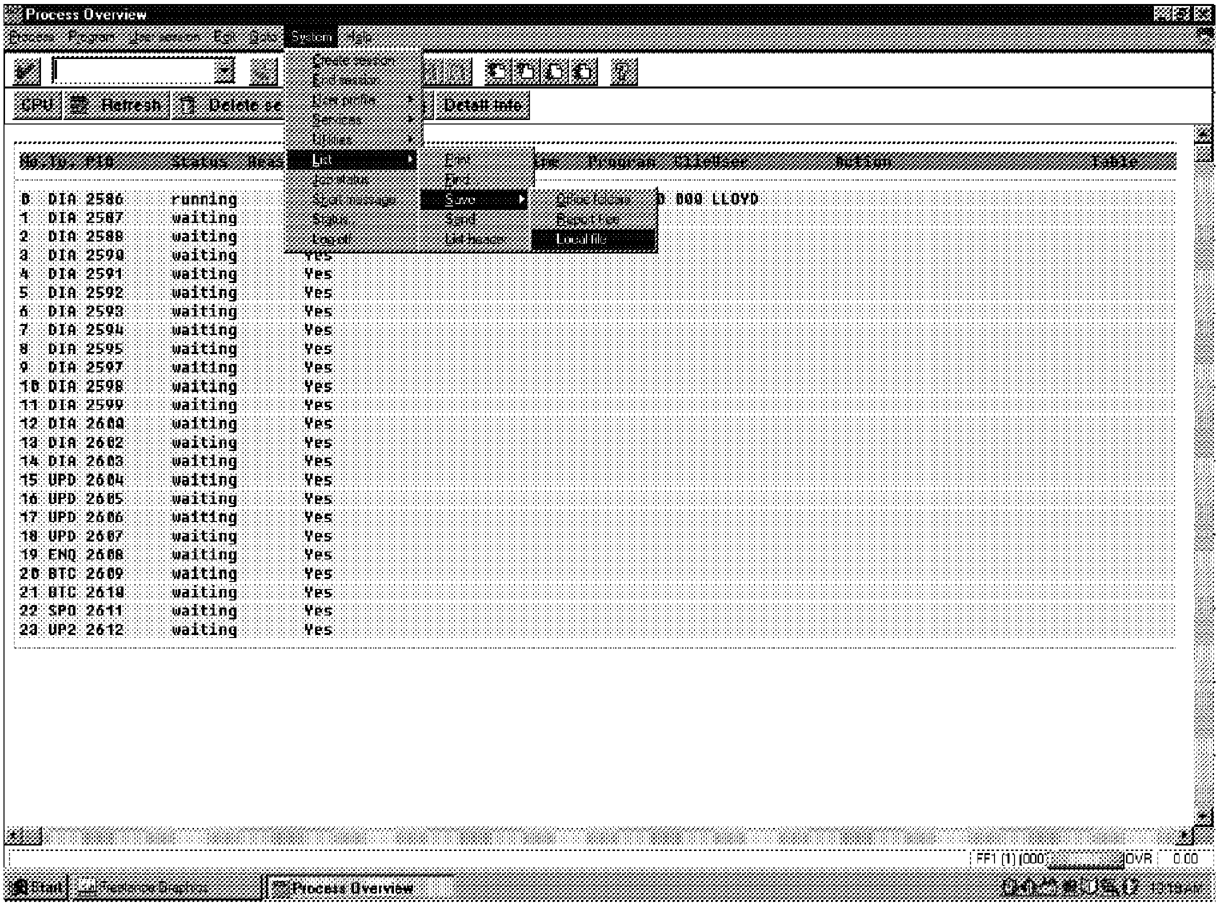


Figure 57. SM50 - Work Processes (Save 1 of 3)

Select **unconverted** as indicated in Figure 58 on page 135.

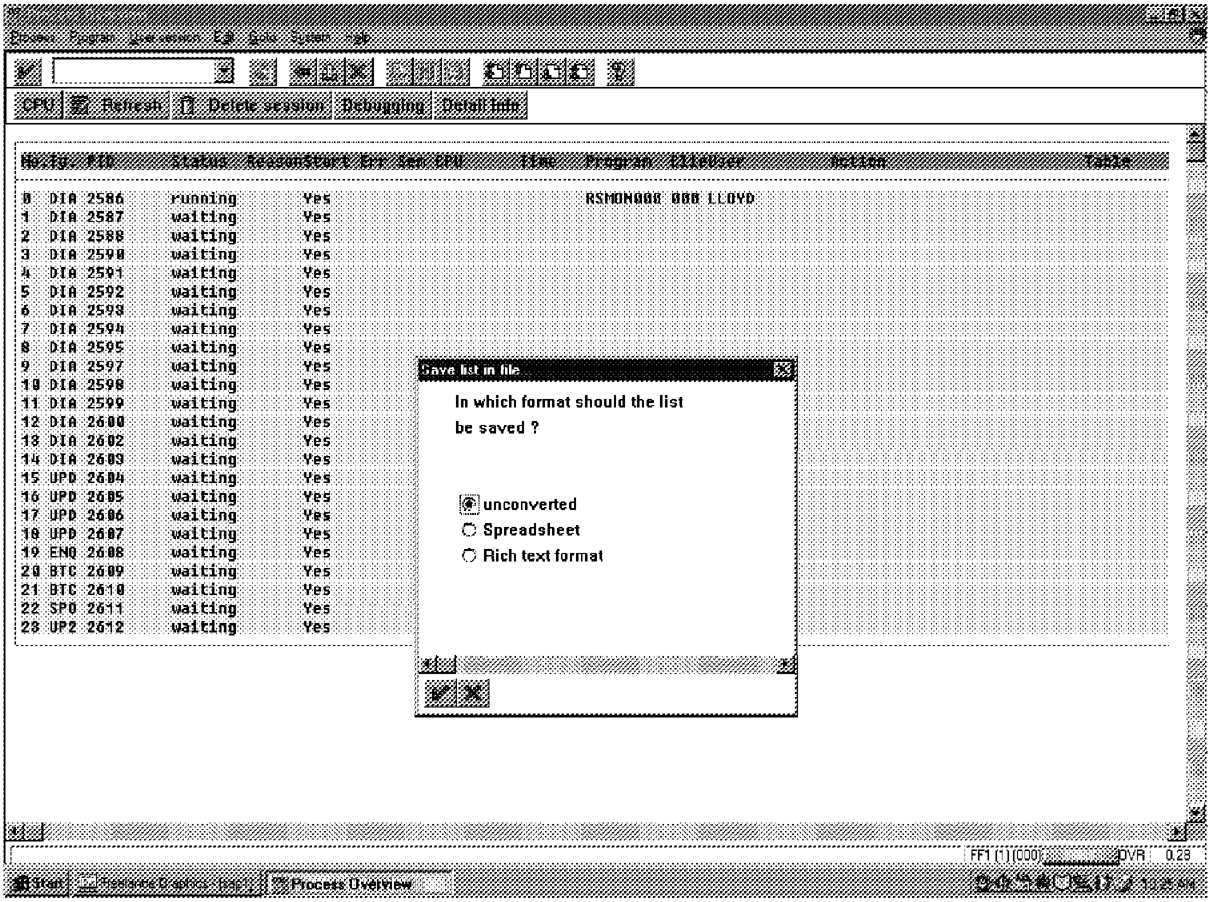


Figure 58. SM50 - Work Processes (Save 2 of 3)

Enter a name for the PC file you want the information written to.

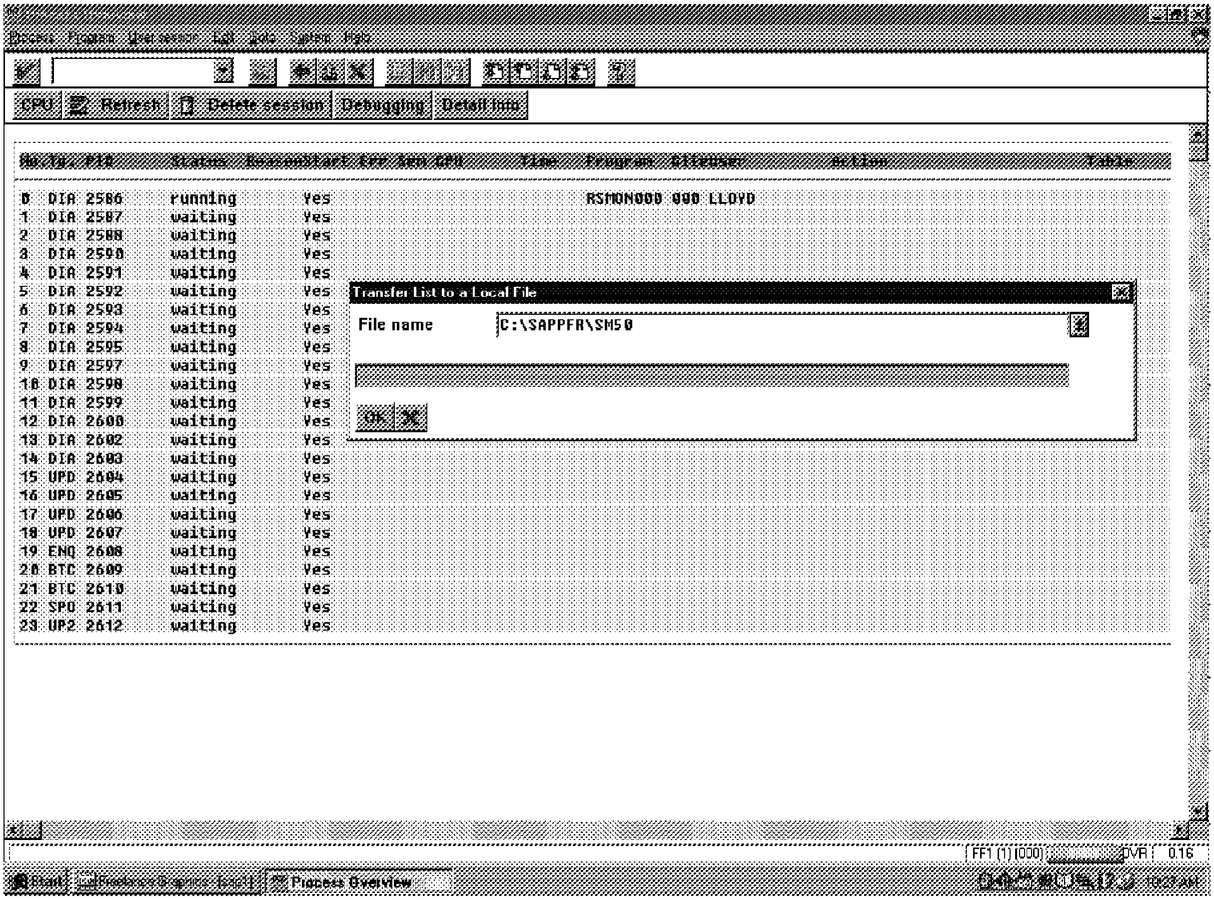


Figure 59. SM50 - Work Processes (Save 3 of 3)

Sign on to a 5250 Session on the AS/400 application server and use the WRKPID command in the SAP kernel library (R3<rel>OPT) where <rel> represents the installed SAP R/3 release. This provides the AS/400 job number corresponding to the Process Identifiers (PIDs) displayed by SAP transaction SM50.

```
                        Work with Job by PID (WRKPID)

 Type choices, press Enter.

 Process ID . . . . . . . . . . .   .389.  Character value
 Prompt command . . . . . . . . .   *YES          *YES, *NO, Y, N
 xxxJOB command . . . . . . . . .   .DSP.  Character value





  F3=Exit   F4=Prompt   F5=Refresh   F12=Cancel   F13=How to use this display
  F24=More keys
```

When the DSPJOB display is shown, note the AS/400 job number. You need these to allocate the various AS/400 jobs to the SAP R/3 workloads when you are creating the BEST/1 model.

```
                         Display Job (DSPJOB)

 Type choices, press Enter.

 Job name . . . . . . . . . . . . > DISP          Name, *
   User . . . . . . . . . . . . . >   SAP00       Name
   Number . . . . . . . . . . . . >  .004379.  000000-999999
 Output . . . . . . . . . . . . .   *             *, *PRINT
 Option . . . . . . . . . . . . .   *SELECT       *SELECT, *STSA, *DFNA..



F3=Exit   F4=Prompt   F5=Refresh   F10=Additional parameters   F12=Cancel
F13=How to use this display        F24=More keys
```

8.2.2 SAP R/3 Workload Analysis

Enter the SAP R/3 transaction **ST03**, which shows a window similar to Figure 60.

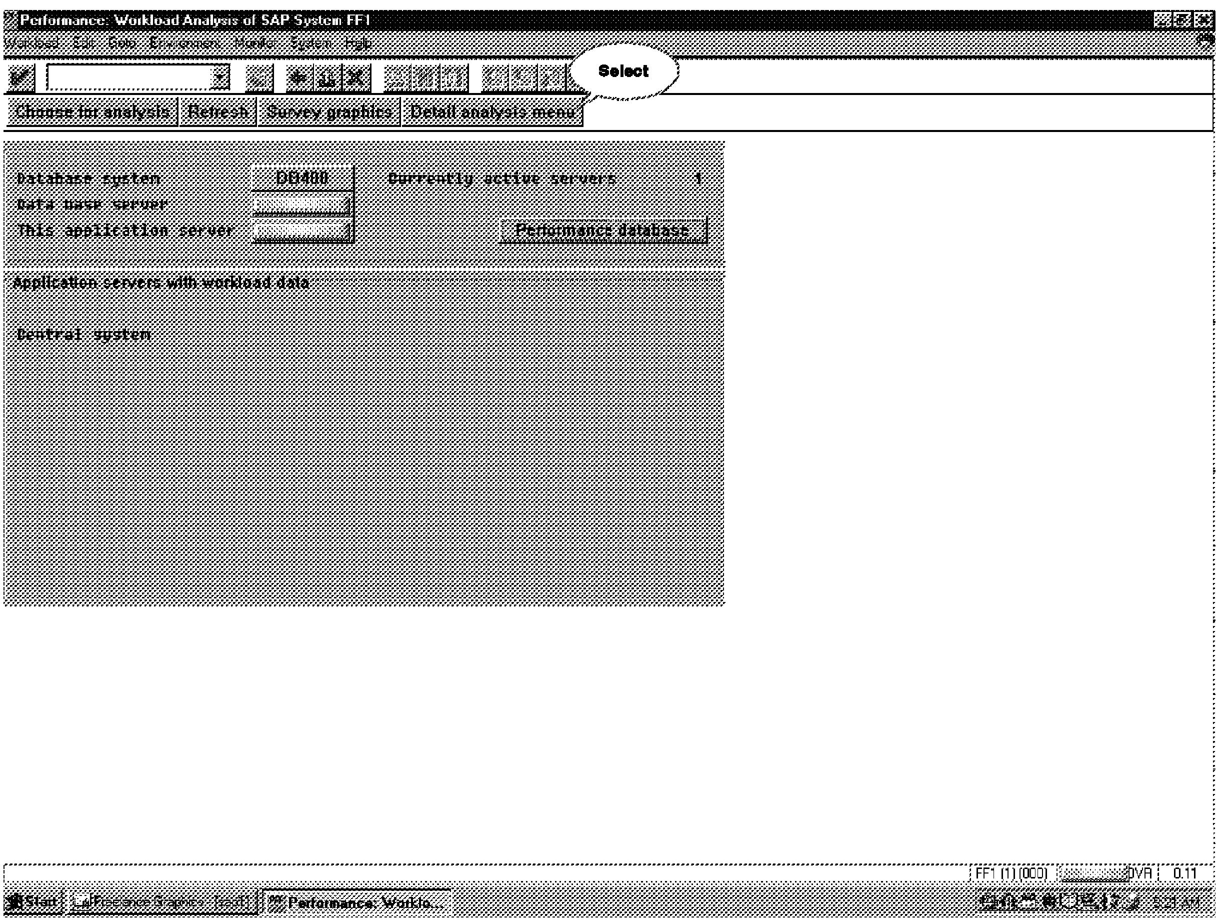


Figure 60. ST03 - Workload Analysis (Server Selection)

From the initial ST03 window, select **Detail analysis menu** with the mouse. On Figure 61, select **Last minutes workload**, which shows a window to set the time interval. Select **Other selection** in this window.

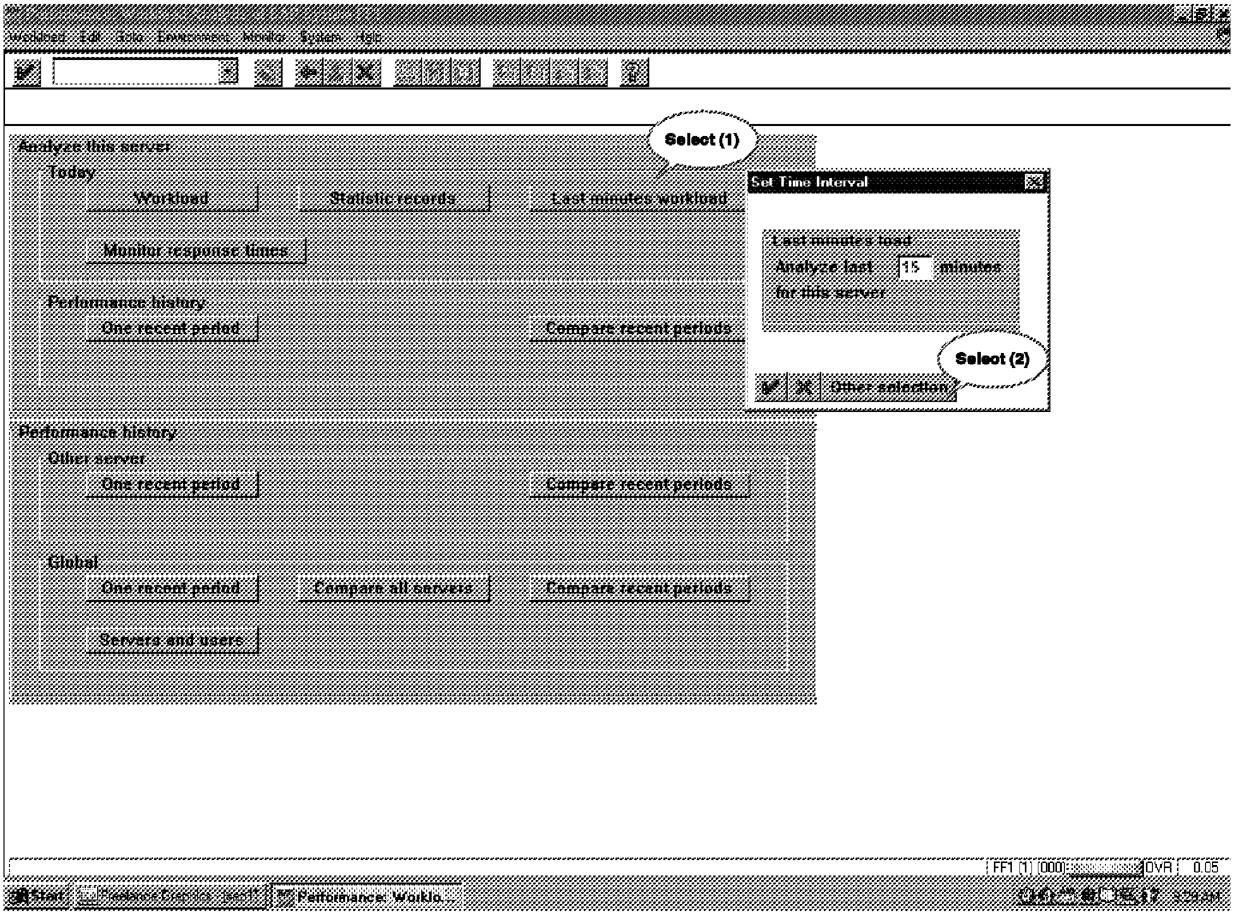


Figure 61. ST03 - Workload Analysis (Time Interval)

When Figure 62 is displayed, enter the date, start, and end times that correspond to the data collection period specified on the AS/400 performance data collection command.

Ensure that you enter **time profile resolution** to correspond to the **time interval** specified in the STRPFRMON command. You also must de-select the option **task profile by work process**.

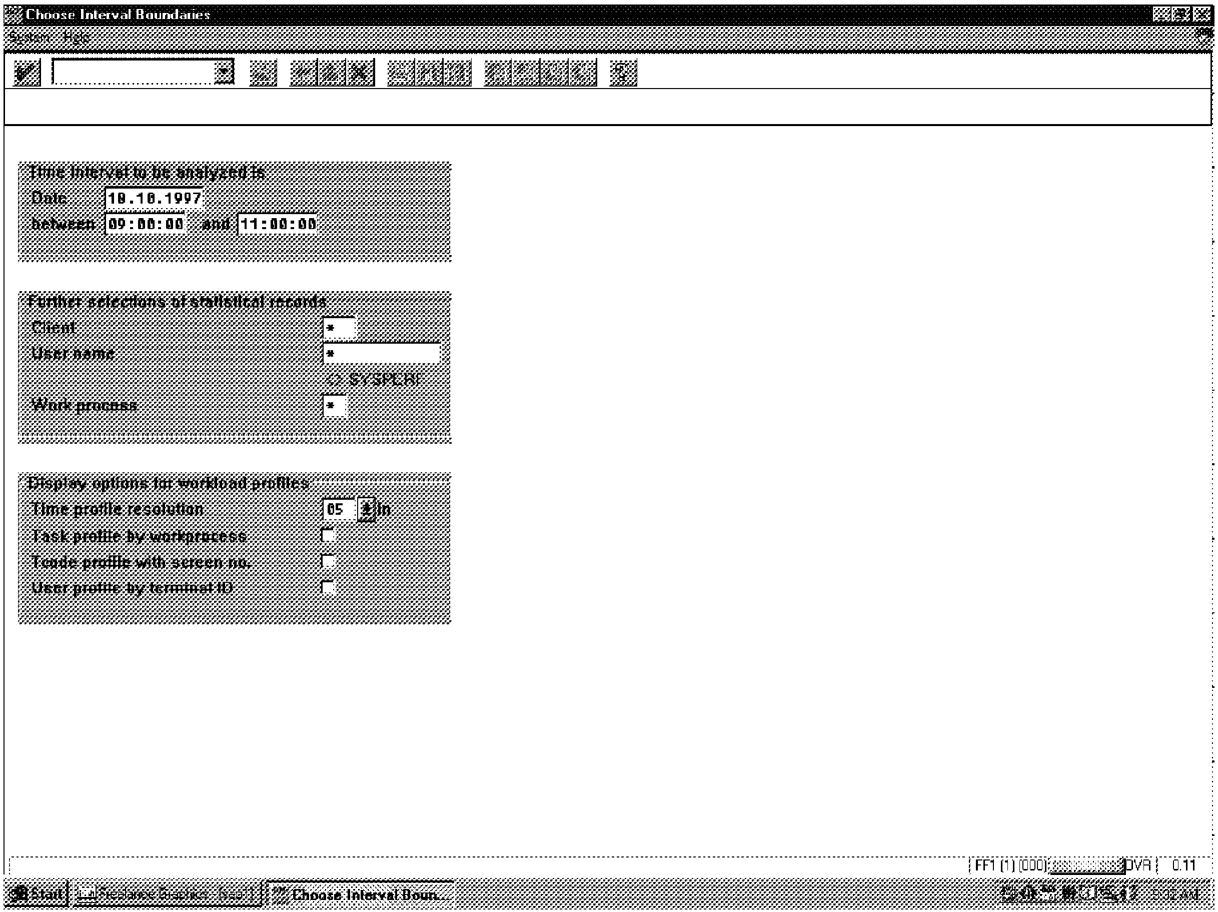


Figure 62. ST03 - Workload Analysis (Interval Boundaries)

8.2.2.1 SAP R/3 Workload by Time Profile

When you press Enter, a window similar to Figure 63 is displayed indicating a summary analysis of the **Total** workload.

Select **Time Profile** with the mouse to display the workload analysis for the total workload by time intervals.

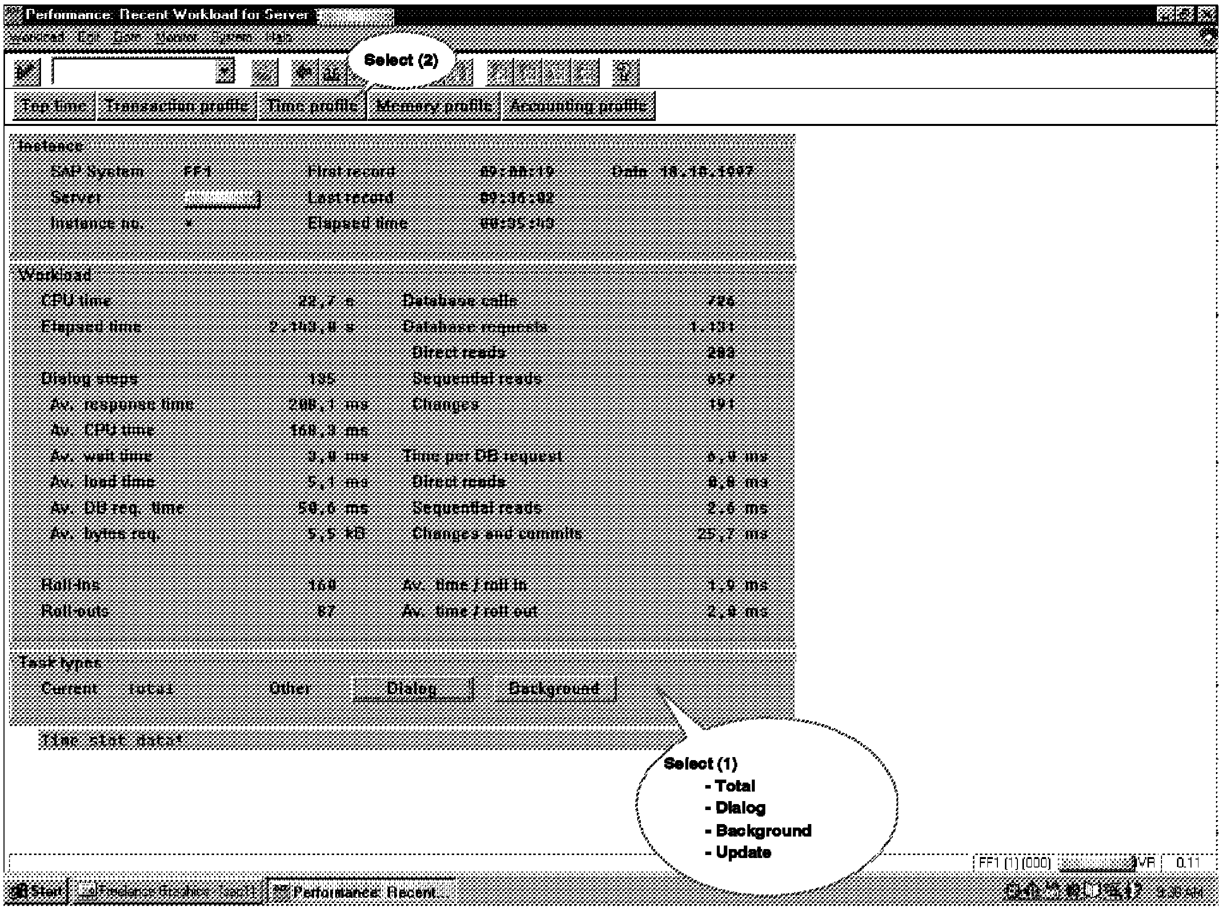


Figure 63. ST03 - Workload Analysis (Summary)

The next window is similar to Figure 64 on page 141. Select **System -> List -> Save -> Local file** and enter the name of the PC file to store the data.

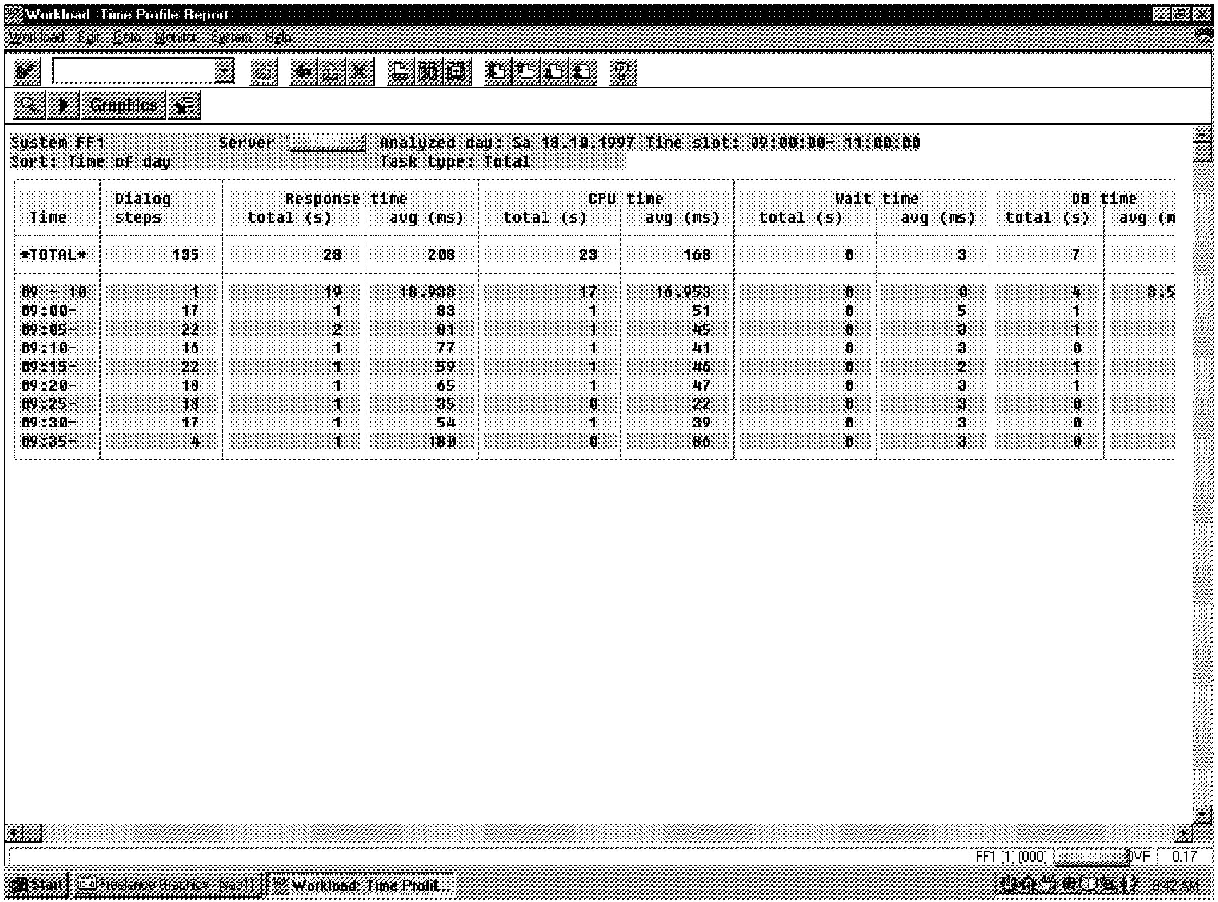


Figure 64. ST03 - Time Profile Report

Return to the Workload Analysis (Summary) window and repeat the process to record the Time Profile data for the following workloads:

- Dialog
- Background
- Update

8.2.2.2 SAP R/3 Workload Overview

Return to the Workload Analysis window and select **Goto -> Summary reports -> Workload overview**.

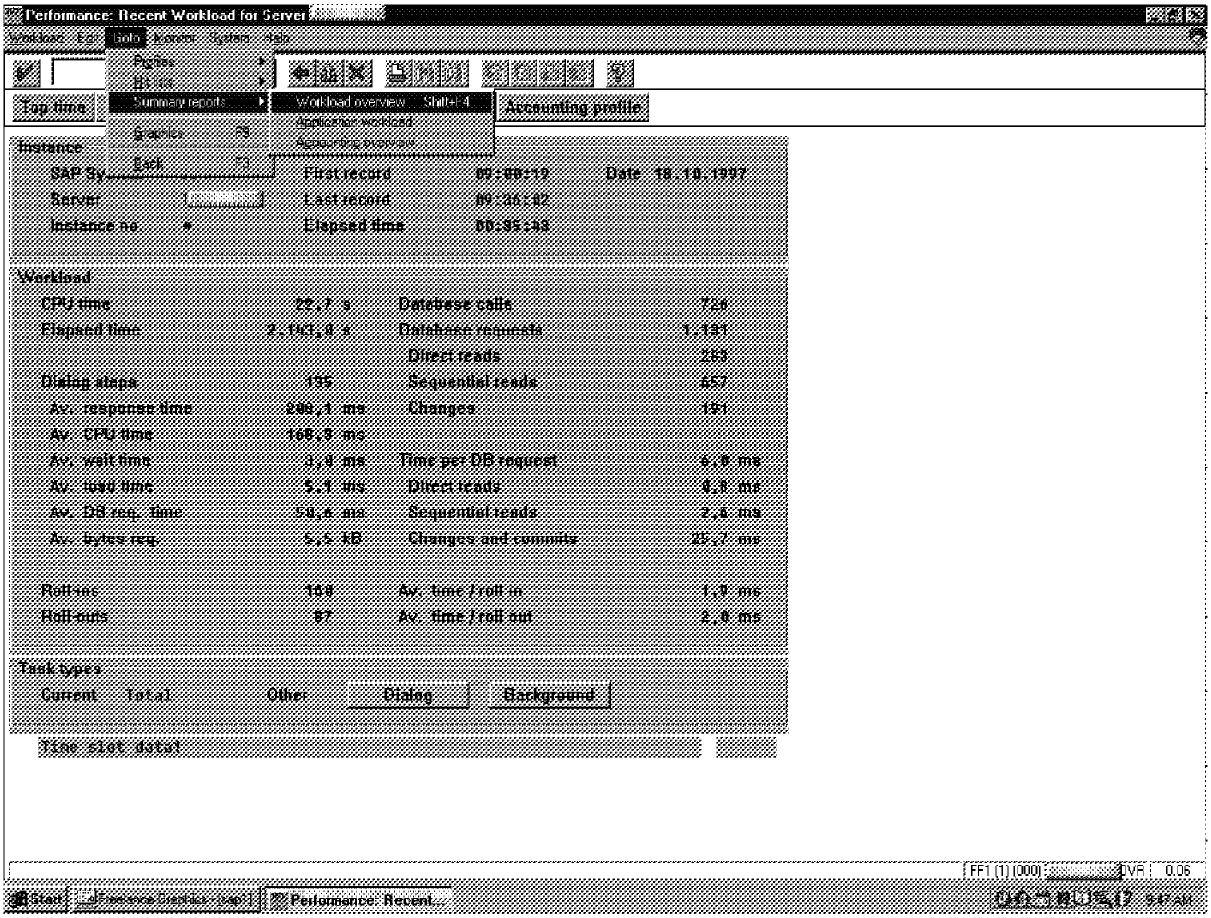
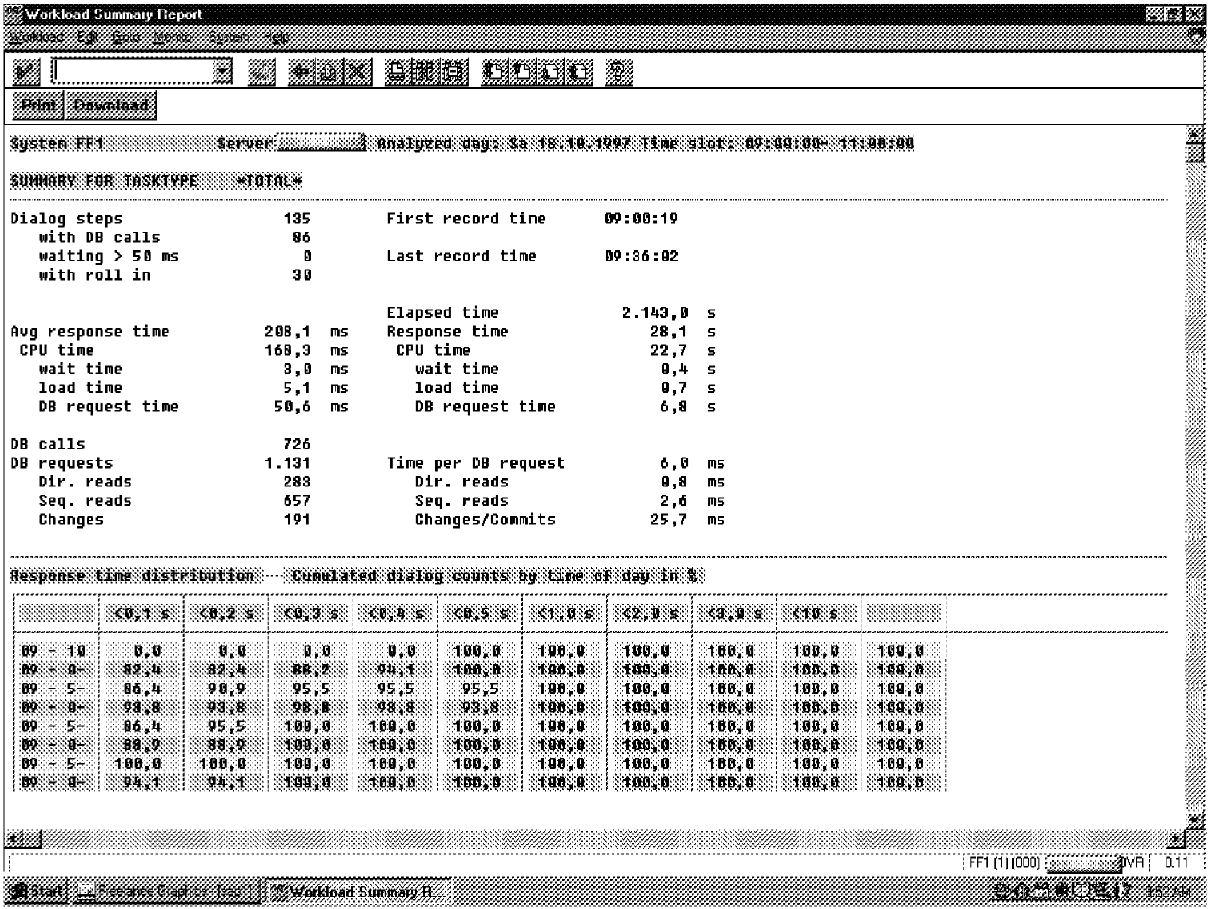


Figure 65. ST03 - Workload Overview

Figure 66 shows a summary report of the response times and resource utilizations for the selected period. Use **System -> List -> Save -> Local file** and enter the name of the PC file to store the data.



Workload Summary Report

System FF1 Server Analyzed day: Sa 18.10.1997 Time slot: 09:00:00- 11:00:00

SUMMARY FOR TASKTYPE *TOTAL*

Dialog steps	135		First record time	09:00:19	
with DB calls	86				
waiting > 50 ms	0		Last record time	09:36:02	
with roll in	30				
			Elapsed time	2.143,0	s
Avg response time	208,1	ms	Response time	28,1	s
CPU time	168,3	ms	CPU time	22,7	s
wait time	3,0	ms	wait time	0,4	s
load time	5,1	ms	load time	0,7	s
DB request time	50,6	ms	DB request time	6,8	s
DB calls	726				
DB requests	1.131		Time per DB request	6,0	ms
Dir. reads	283		Dir. reads	0,8	ms
Seq. reads	657		Seq. reads	2,6	ms
Changes	191		Changes/Commits	25,7	ms

Response time distribution --- Cumulated dialog counts by time of day in %

	<0,1 s	<0,2 s	<0,3 s	<0,4 s	<0,5 s	<1,0 s	<2,0 s	<3,0 s	<10 s	
09 - 10	0,0	0,0	0,0	0,0	100,0	100,0	100,0	100,0	100,0	100,0
09 - 0-	82,4	82,4	88,2	94,1	100,0	100,0	100,0	100,0	100,0	100,0
09 - 5-	86,4	90,9	95,5	95,5	95,5	100,0	100,0	100,0	100,0	100,0
09 - 0-	93,8	93,8	93,8	93,8	93,8	100,0	100,0	100,0	100,0	100,0
09 - 5-	86,4	95,5	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
09 - 0-	88,9	88,9	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
09 - 5-	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
09 - 0-	94,1	94,1	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Figure 66. ST03 - Workload Overview Report

8.2.2.3 SAP R/3 Workload Overview (Application)

If you can isolate the AS/400 performance data by SAP R/3 application modules (for examples, using separate SAP R/3 instances for each module), you can obtain SAP R/3 data by application.

From the Workload overview window, select SUMMARY REPORTS -> APPLICATION WORKLOAD for each of the following types:

- Total
- Dialog
- Background
- Update

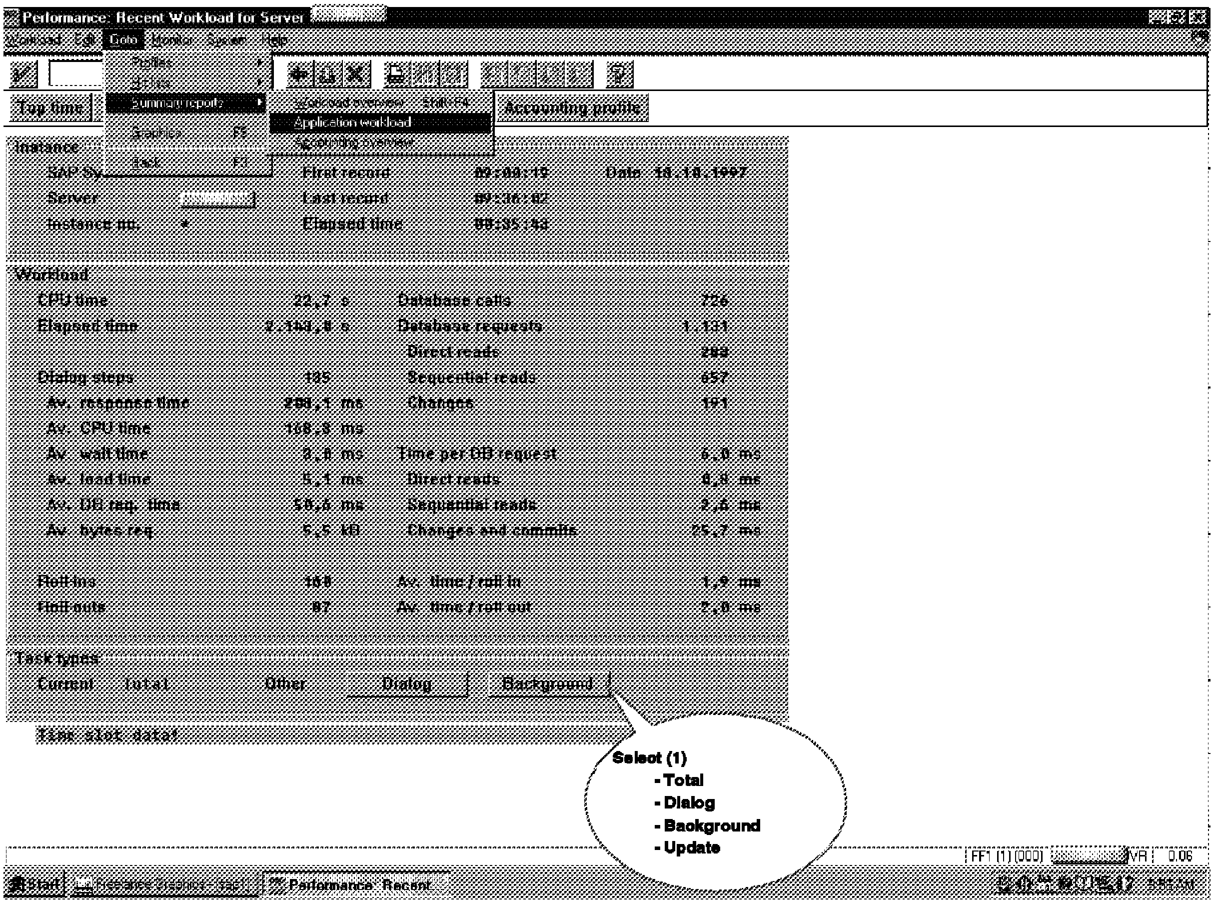


Figure 67. ST03 - Application Workload

Use **System -> List -> Save -> Local file** and enter the name of the PC file to store the data.

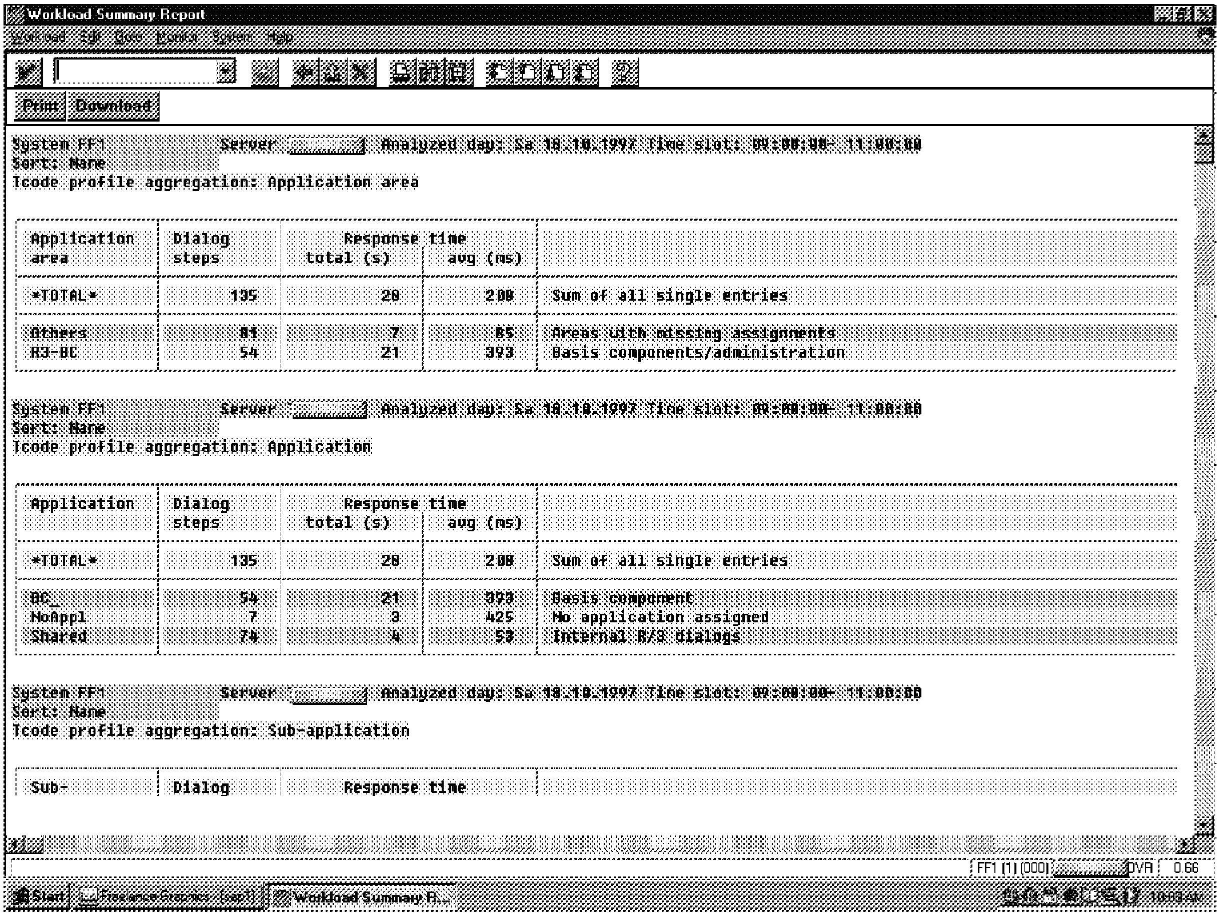


Figure 68. ST03 - Application Workload Report

8.3 SAP R/3 Modelling with BEST/1

In the following section, we show an example of using BEST/1 to predict AS/400 resource utilization with increased workload and the expected server response times.

AS/400 performance data was collected at five-minute intervals. At the end of the measured period, the SAP R/3 transactions discussed in the previous section were run and the SAP R/3 performance data downloaded to PC files and printed.

8.3.1 Centralized SAP R/3 Implementation

This section presents an example of modelling a centralized implementation of SAP R/3 where the application and database server functions are performed on a single AS/400 system.

8.3.1.1 SAP R/3 Data

The following reports present the key SAP R/3 reports used in the model:

Measured Data

It is not within the scope of this presentation to analyze or reconcile the information from the SAP R/3 reports. For example, the background dialogs mainly have sub-second response times, but about five steps that span multiple periods have large response times resulting in a high average. Also, the sum of the average CPU and database times for the interactive (and update, also) dialog is greater than the average response time!

Note: The Work Process report in Figure 69 on page 147 includes the AS/400 Job numbers that have been introduced manually using the WRKPID command on the AS/400 system.

8.3.1.2 SM50 Work Process Report

```
                                          |AS/400 Job Numbers
                                          |Entered using WRKPID
|-----------------------------------------|--------------------------------------------------------------------------------|
|No.Ty. PID      Status  ReasonStart Err|Sem CPU      Time   Program  Cl. User        Action                     Table   |
|----------------------------------------V--------------------------------------------------------------------------------|
|0  DIA 39470    running       Yes  (134499)                 RSMON000 084 IBM                                              |
|1  DIA 39471    running       Yes  (134500)                 SAPMSYST 084 XXXX9999    Roll In                              |
|2  DIA 39472    waiting       Yes  (134501)                                                                               |
|3  DIA 39473    waiting       Yes  (134502)                                                                               |
|4  DIA 39474    running       Yes  (134503)                 ZSDP0002 084 YYYY9999                                         |
|5  DIA 39475    waiting       Yes  (134504)                                                                               |
|6  DIA 39476    waiting       Yes  (134505)                                                                               |
|7  DIA 39477    waiting       Yes  (134506)                                                                               |
|8  DIA 39478    waiting       Yes  (134507)                                                                               |
|9  DIA 39479    waiting       Yes  (134508)                                                                               |
|10 DIA 39480    waiting       Yes  (134509)                                                                               |
|11 DIA 39481    waiting       Yes  (134510)                                                                               |
|12 DIA 39482    waiting       Yes  (134511)                                                                               |
|13 DIA 39483    waiting       Yes  (134512)                                                                               |
|14 DIA 39485    waiting       Yes  (134514)                                                                               |
|15 DIA 39486    waiting       Yes  (134515)                                                                               |
|16 DIA 39487    waiting       Yes  (134516)                                                                               |
|17 DIA 39488    waiting       Yes  (134517)                                                                               |
|18 DIA 39489    waiting       Yes  (134518)                                                                               |
|19 DIA 39490    waiting       Yes  (134519)                                                                               |
|                                                                                                                          |
|20 UPD 39491    waiting       Yes  (134520)                                                                               |
|21 UPD 39492    waiting       Yes  (134521)                                                                               |
|22 UPD 39493    waiting       Yes  (134522)                                                                               |
|23 UPD 39494    waiting       Yes  (134523)                                                                               |
|                                                                                                                          |
|24 ENQ 39495    waiting       Yes  (134524)                                                                               |
|                                                                                                                          |
|25 BTC 39496    waiting       Yes  (134525)                                                                               |
|26 BTC 39497    waiting       Yes  (134526)                                                                               |
|27 BTC 39498    waiting       Yes  (134527)                                                                               |
|28 BTC 39499    waiting       Yes  (134528)                                                                               |
|29 BTC 39500    waiting       Yes  (134529)                                                                               |
|30 BTC 39501    waiting       Yes  (134530)                                                                               |
|31 BTC 39502    waiting       Yes  (134531)                                                                               |
|32 BTC 39503    waiting       Yes  (134532)                                                                               |
|                                                                                                                          |
|33 SPO 39504    waiting       Yes  (134533)                                                                               |
|                                                                                                                          |
|34 UP2 39505    waiting       Yes  (134534)                                                                               |
|35 UP2 39506    waiting       Yes  (134535)                                                                               |
|36 UP2 39507    waiting       Yes  (134536)                                                                               |
|37 UP2 39508    waiting       Yes  (134537)                                                                               |
----------------------------------------------------------------------------------------------------------------------------
```

Figure 69. SM50 - Work Processes with AS/400 Job Numbers

8.3.1.3 SM03 Interactive Workload Report

System XXX   Server XXXXXXXX   Analyzed day: Mo 24.11.1997 Time slot: 14:00:00- 16:00:00
Sort: Time of day   Task type: Dialog

Time	Dialog steps	Response time total (s)	Response time avg (ms)	CPU time total (s)	CPU time avg (ms)	Wait time total (s)	Wait time avg (ms)	DB time total (s)	DB time avg (ms)
TOTAL	17.201	17.636	1.025	9.441	549	62	4	9.900	576
14:00-	1.012	1.571	1.552	658	650	12	12	783	774
14:05-	774	1.671	2.158	442	571	6	8	701	905
14:10-	799	1.589	1.989	881	1.103	4	4	778	974
14:15-	910	2.737	3.007	2.134	2.345	2	2	2.535	2.785
14:20-	682	815	1.195	311	456	2	3	413	605
14:25-	603	548	908	384	636	1	2	348	577
14:30-	670	569	849	220	329	6	8	257	384
14:35-	548	397	724	146	267	1	2	127	231
14:40-	648	551	850	237	366	1	2	197	303
14:45-	598	475	794	212	354	1	2	237	396
14:50-	515	407	790	158	307	1	2	198	383
14:55-	658	270	411	159	242	1	1	138	210
15:00-	859	481	560	298	347	8	9	168	196
15:05-	979	319	326	163	166	2	2	161	165
15:10-	748	265	355	134	179	1	1	136	182
15:15-	472	287	609	167	354	1	2	144	306
15:20-	422	518	1.228	261	618	2	4	269	638
15:25-	631	408	647	276	437	2	3	172	272
15:30-	822	625	761	263	320	3	4	289	351
15:35-	738	590	799	139	189	1	2	156	211
15:40-	967	1.780	1.841	1.451	1.501	2	2	1.284	1.328
15:45-	724	293	404	114	158	1	2	152	209
15:50-	824	238	288	132	161	1	1	135	164
15:55-	596	232	389	100	168	1	2	122	205
16:00-	2	1	604	1	259	0	1	1	490

Figure 70. ST03 - Interactive Work Report

8.3.1.4 SM03 Batch Workload Report

System XXX Server XXXXXXXXX Analyzed day: Mo 24.11.1997 Time slot: 14:00:00- 16:00:00
Sort: Time of day Task type: Background

Time	Dialog steps	Response time total (s)	Response time avg (ms)	CPU time total (s)	CPU time avg (ms)	Wait time total (s)	Wait time avg (ms)	DB time total (s)	DB time avg (ms)
TOTAL	107	1.070	10.001	106	987	6	56	228	2.130
14 - 15	3	717	238.962	75	24.839	0	0	148	49.411
14:00-	8	54	6.749	3	341	3	371	31	3.872
14:05-	6	2	311	1	94	0	25	2	259
14:10-	3	0	147	0	74	0	10	0	122
14:15-	4	1	339	0	98	0	17	1	266
14:20-	4	1	191	0	74	0	52	1	166
14:25-	4	1	290	0	97	0	20	1	216
14:30-	5	1	224	0	68	0	10	1	200
14:35-	3	0	100	0	68	0	6	0	76
14:40-	4	0	110	0	69	0	15	0	83
14:45-	4	1	130	0	95	0	13	0	101
14:50-	4	1	191	0	71	0	7	0	76
14:55-	4	1	187	0	70	0	19	0	105
15 - 16	2	262	130.969	19	9.733	0	0	24	12.025
15:00-	8	19	2.404	2	231	2	199	12	1.491
15:05-	3	0	143	0	71	0	13	0	106
15:10-	4	1	140	0	70	0	12	0	106
15:15-	4	1	131	0	90	0	7	0	80
15:20-	4	1	268	0	74	0	35	1	243
15:25-	4	1	189	0	80	0	10	1	149
15:30-	4	0	122	0	75	0	24	0	87
15:35-	4	2	618	1	133	0	18	1	237
15:40-	4	1	143	0	73	0	13	0	117
15:45-	4	1	180	0	72	0	11	1	151
15:50-	4	0	101	0	72	0	10	0	76
15:55-	2	1	320	0	111	0	9	0	91

Figure 71. ST03 - Batch Workload Report

8.3.1.5 SM03 Update Workload Report

System XXX Server XXXXXXX Analyzed day: Mo 24.11.1997 Time slot: 14:00:00- 16:00:00
Sort: Time of day Task type: Update

Time	Dialog steps	Response time total (s)	Response time avg (ms)	CPU time total (s)	CPU time avg (ms)	Wait time total (s)	Wait time avg (ms)	DB time total (s)	DB time avg (ms)
TOTAL	2.677	1.835	685	1.098	410	15	6	1.235	461
14:00-	234	187	799	63	269	4	16	137	585
14:05-	198	238	1.202	94	477	2	11	156	785
14:10-	100	46	459	26	257	1	8	29	293
14:15-	88	63	719	50	568	0	4	46	521
14:20-	95	73	773	47	494	0	3	50	523
14:25-	89	58	647	44	499	0	4	38	429
14:30-	160	157	983	116	728	1	9	112	697
14:35-	178	97	547	82	458	1	4	57	318
14:40-	118	54	462	39	329	1	5	35	292
14:45-	170	122	720	98	577	0	3	83	486
14:50-	132	69	521	57	430	0	2	43	329
14:55-	63	34	542	26	416	0	2	20	322
15:00-	218	138	634	75	343	1	5	93	427
15:05-	122	46	375	35	286	0	3	25	205
15:10-	67	42	626	29	437	0	3	27	406
15:15-	35	18	509	12	343	0	2	11	317
15:20-	130	61	469	41	317	0	4	41	315
15:25-	64	36	561	22	345	0	5	23	358
15:30-	51	39	756	23	443	0	5	29	571
15:35-	56	40	720	17	301	0	5	29	513
15:40-	71	36	506	21	296	0	5	22	309
15:45-	63	23	370	14	223	0	4	13	200
15:50-	51	30	589	17	341	0	3	18	355
15:55-	124	127	1.020	49	397	0	3	100	805

Figure 72. ST03 - Update Workload Report

8.3.1.6 SM03 Total Workload Report

```
System XXX            Server XXXXXXXXX Analyzed day: Mo 24.11.1997 Time slot: 14:00:00- 16:00:00
Sort: Time of day                      Task type: Total
```

Time	Dialog steps	Response time total (s)	Response time avg (ms)	CPU time total (s)	CPU time avg (ms)	Wait time total (s)	Wait time avg (ms)	DB time total (s)	DB time avg (ms)
TOTAL	20.488	20.765	1.014	10.709	523	107	5	11.420	557
14 - 15	3	717	238.962	75	24.839	0	0	148	49.411
14:00-	1.273	1.823	1.432	726	570	20	16	954	749
14:05-	1.049	1.969	1.877	548	522	15	15	866	825
14:10-	917	1.640	1.789	909	991	5	5	810	883
14:15-	1.031	2.817	2.732	2.189	2.123	3	3	2.587	2.510
14:20-	795	893	1.124	360	452	3	4	465	585
14:25-	711	611	859	430	605	2	3	389	546
14:30-	856	736	859	340	397	7	9	372	435
14:35-	742	497	670	229	309	2	2	184	248
14:40-	782	608	778	277	354	2	3	232	297
14:45-	785	600	764	311	396	2	2	320	408
14:50-	662	478	722	216	326	1	2	242	365
14:55-	736	306	416	186	253	1	2	159	217
15 - 16	2	262	130.969	19	9.733	0	0	24	12.025
15:00-	1.116	653	585	379	340	11	10	276	247
15:05-	1.118	369	330	199	178	2	2	188	168
15:10-	837	315	376	165	197	1	1	166	198
15:15-	547	327	598	185	338	3	5	165	301
15:20-	616	610	991	313	507	10	16	317	514
15:25-	720	452	628	301	418	3	4	197	274
15:30-	888	666	750	287	323	4	4	319	359
15:35-	810	635	784	158	195	2	2	186	230
15:40-	1.054	1.818	1.725	1.473	1.398	2	2	1.307	1.240
15:45-	804	322	401	131	163	2	2	167	208
15:50-	901	277	308	153	170	2	3	155	172
15:55-	731	362	495	151	206	2	2	224	306
16:00-	2	1	604	1	259	0	1	1	490

Figure 73. ST03 - Total Workload Report

8.3.1.7 Create SAP R/3 Centralized Model

1. From the **Work with BEST/1 Models** menu, we select *option 1 - Create a model* and specify a model name.

2. From the **Create BEST/1 Model** menu, we select *option 1 - Create from performance data*.

3. At the **Create BEST/1 Model from Performance Data** menu, we enter the highlighted values in the following display.

```
                 Create BEST/1 Model from Performance Data

Model . . . . . . . . . . . . . . :   CENTRAL

Type choices, press Enter.  Use *SLTHOUR to select an hour-long time period or
  use *SLTITV to select first and last interval of a one to two hour
  time period.  The time period selected should be representative of your peak
  processing activity.

Text . . . . . . . . . . . . . .  SAP R/3 Centralized System

Performance data:
  Member . . . . . . . . . . . .  Q972221400   Name, F4 for list
  Library  . . . . . . . . . . .  SAPJOBS      Name

Start time . . . . . . . . . . .  *SLTITV      Time, *FIRST, *SLTHOUR, *SLTITV
Start date . . . . . . . . . .    *FIRST       Date, *FIRST

Stop time  . . . . . . . . . .    *LAST        Time, *LAST
Stop date  . . . . . . . . . .    *LAST        Date, *LAST

F3=Exit   F4=Prompt   F12=Cancel
```

4. At the **Select Time Interval** display, we select a two-hour period corresponding to the SAP R/3 reports printed on the previous pages.

5. At the **Classify Jobs** display, we select *option 2 - Classify Jobs into workloads*.

6. At the **Specify Job Classification Category** display, we select *option 5 - Job number* because from the SAP R/3 Work Process Report (Figure 69 on page 147), we know which AS/400 jobs correspond to the various SAP R/3 Work Process types.

7. At the **Edit Job Classifications** display, we select *F9 - to display values from performance data* and allocate the AS/400 jobs to the following workloads:

   - DIALOG (interactive)
   - UPDATE (update)
   - BATCH (batch)
   - ENQ_SPL (enqueue/spool)
   - QDEFAULT (all other jobs)

8. At the **Specify Paging Behaviors** display, we accept the default value of *GENERIC for all workloads.

9. At the **Define Non-Interactive Transactions** display, we enter the values as highlighted using the dialog step counts in the reports divided by two because the reports covered a **two-hour** period and the value to be entered into the model is **per hour**.

   - Figure 70 on page 148 for the DIALOG workload
   - Figure 72 on page 150 for the UPDATE workload
   - Figure 71 on page 149 for the BATCH workload

```
                        Define Non-Interactive Transactions

 Job classification category  . . . . . . :   Job Number


 Type choices, press Enter.

               ---Activity Counted as Transaction---     Total Transactions
 Workload          Type                   Quantity        when Type = *NONE
 QDEFAULT          *LGLIO                  100.0                    0
 ENQ_SPL           *LGLIO                  100.0                    0
 UPDATE           .*NONE                   100.0               .[illegible]
 BATCH            .*NONE                   100.0               .[illegible]
 DIALOG           .*NONE                   100.0               .[illegible]

                                                                         Bottom

Type:  *LGLIO, *CMNIO, *CPUSEC, *PRINT, *NONE

F3=Exit   F12=Cancel
```

10. At the **Save Job Classification Member** display, we enter the values for the member, library, and a description.

11. At the **Confirm Creation of BEST/1 Model** display, we press Enter to proceed with model creation.

12. When the model creation is complete, it is shown on the **Work with BEST/1 Models** display.

8.3.1.8 Validate SAP R/3 Centralized Model

After the BEST/1 model has been created, it must be validated against actual measured data. A part of this is done within BEST/1 options.

1. From the **Work with BEST/1 Models** menu, we select *option 5 - Work with* next to the model we created.

2. From the **Work with BEST/1 Model** menu, we select *option 5 - Analyze current model*, which shows the following **Work with Results** display after the model is analyzed.

3. We select option 5 to display **Measured and Predicted Comparison**:

```
                                  Work with Results

 Printed report text . . . . . .   SAP R/3 Centralized System



 Type options, press Enter.
   5=Display   6=Print

 Opt     Report Name
 .5      Measured and Predicted Comparison
 __      Analysis Summary
 __      Recommendations
 __      Workload Report
 __      ASP and Disk Arm Report
 __      Disk IOP and Disk Arm Report
 __      Main Storage Pool Report
 __      Communications Resources Report
 __      All of the above

                                                                     Bottom
 F3=Exit   F12=Cancel   F14=Select saved results   F15=Save current results
 F18=Graph current results   F19=Append saved results   F24=More keys
 Model has been analyzed
```

4. We see the comparison between the AS/400 measured resource utilization and the predictions computed by the BEST/1 model.

```
                        Measured and Predicted Comparison

                                               Measured     Predicted
Total CPU util . . . . . . . . . . . . :          41.1          41.1
Disk IOP util  . . . . . . . . . . . . :          12.2          11.6
Disk arm util  . . . . . . . . . . . . :           3.6           3.7
Disk IOs per second  . . . . . . . . . :         215.4         215.5

LAN IOP util . . . . . . . . . . . . . :           7.0            .0
LAN line util  . . . . . . . . . . . . :            .0            .0
WAN IOP util . . . . . . . . . . . . . :           7.7           5.9
WAN line util  . . . . . . . . . . . . :            .5            .1

Interactive:
  CPU util . . . . . . . . . . . . . . :            .2            .2
  Int rsp time (seconds) . . . . . . . :            .4            .3
  Transactions per hour  . . . . . . . :           384           384
Non-interactive thruput  . . . . . . . :         11926         12074

Performance estimates -- Press help to see disclaimer.
F3=Exit   F6=Print   F9=Work with spooled files   F12=Cancel
F17=Calibrate response time
```

Note that the two sets of values are similar except for the communications workload. However, as we only intend to get utilizations and response times **at the sever**, we opt to ignore any discrepancy in the communications resources.

5. In addition to the resource utilization, we must ensure that the response times measured by SAP R/3 correspond to the values predicted by BEST/1.

   We return to the **Work with Results** display by pressing **F12** and select option 5 against **Workload Report**:

```
                            Work with Results

Printed report text . . . . . .   SAP R/3 Centralized System



Type options, press Enter.
  5=Display   6=Print

Opt     Report Name
__      Measured and Predicted Comparison
__      Analysis Summary
__      Recommendations
5_      Workload Report
__      ASP and Disk Arm Report
__      Disk IOP and Disk Arm Report
__      Main Storage Pool Report
__      Communications Resources Report
__      All of the above

                                                                      Bottom
F3=Exit   F12=Cancel   F14=Select saved results   F15=Save current results
F18=Graph current results   F19=Append saved results   F24=More keys
Model has been analyzed
```

6. The following display is examined for differences between the predicted values shown on this display and the values measured by SAP R/3 transaction **ST03**.

```
                              Display Workload Report

 Period:   Analysis
                       CPU     Thruput  -------Response Times (Secs)-------
 Workload    Type     Util    per Hour  Internal   Local     LAN      WAN
 QDEFAULT      1        .2         384       .3       .0       .3       .0
 BATCH         2        .7   [redacted]   .[redacted]      2.7       .0       .0
 DIALOG        2      32.1   .[redacted]   .[redacted]      0.6       .0       .0
 ENQ_SPL       2        .2         109       .5       .5       .0       .0
 QDEFAULT      2       4.0        1818       .6       .0       .6       .0
 UPDATE        2       4.0   .[redacted]   .[redacted]       .6       .0       .0

                                                                      Bottom
Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-analyze   F11=Response time detail      F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

We press **F11** to see the details of response time:

```
                              Display Workload Report

 Period:   Analysis
                                Total   --------Rsp Time Secs spent in--------
 Workload    Type   Connect  Rsp Time     CPU     I/O    Pool    Comm   Other
 QDEFAULT      1    *LAN           .3      .1      .2      .0      .0      .0
 BATCH         2    *LOCAL  .[redacted]  .[redacted]  .[redacted]      .0      .0     7.3
 DIALOG        2    *LOCAL  .[redacted]  .[redacted]  .[redacted]      .0      .0      .4
 ENQ_SPL       2    *LOCAL         .5      .3      .3      .0      .0      .0
 QDEFAULT      2    *LAN           .6      .2      .3      .0      .0      .0
 UPDATE        2    *LOCAL  .[redacted]  .[redacted]  .[redacted]      .0      .0      .1


                                                                       Bottom
Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-analyze   F11=Workload summary          F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

8.3.1.9 Calibrate SAP R/3 Centralized Model

It is at this point that some judgement must be exercised in the points at which the adjustments are to be made.

Considering the different AS/400 facilities that are used in the measurements (native AS/400 Performance Tools uses STRPFRMON while SAP R/3 uses the SAPOSCOL application), some differences can be expected to occur. Also, there can be some application anomalies that may have occurred during the measurement.

Manual Calibration

If there were any major differences in the system resource utilizations, we have to make adjustments through **Manual Calibration** as discussed in Section 4.5, “Validate the BEST/1 Model” on page 48.

Note: The BEST/1 manual - *BEST/1 Capacity Planning Tool*, SC41-5341, discusses the calibration adjustments required when modelling "server jobs."

1. The information in reports in Figure 70 on page 148, Figure 71 on page 149, and Figure 72 on page 150 and the data from the preceding displays are summarized here for convenience and indicate that some calibration adjustments are necessary.

	Dialog			Update			Batch		
	CPU	Disk	Resp	CPU	Disk	Resp	CPU	Disk	Resp
SAP R/3	0.549	0.566	1.025	0.410	0.461	0.685	0.987	2.130	10.001
BEST/1	0.4	0.2	0.6	0.4	0.1	0.6	1.8	0.9	2.7
Calibrate	-	-	0.42	-	-	.08	-	-	7.30

2. This information is represented graphically in Figure 74.

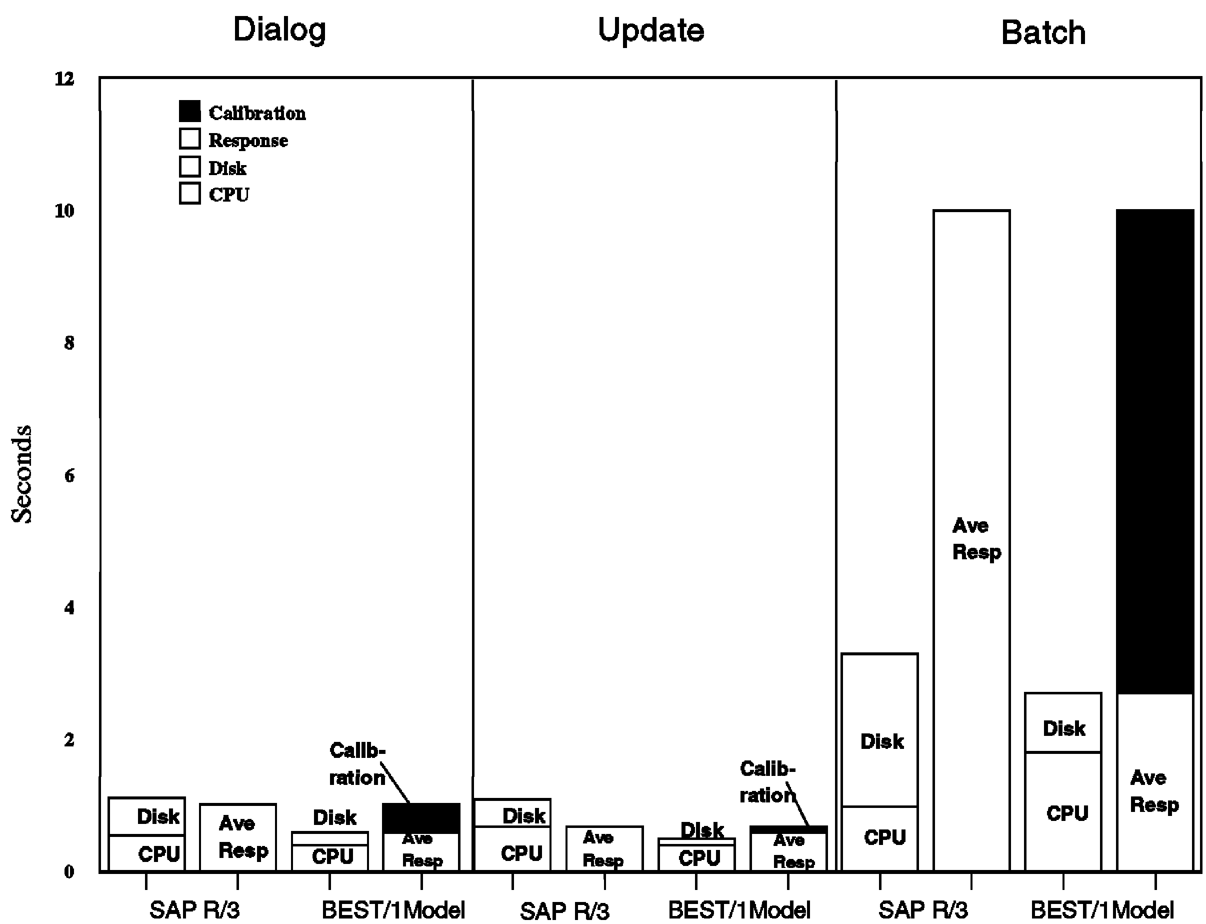


Figure 74. Comparison of Measured versus Predicted Response Times

The dark portion of the "stacked column" for BEST/1 average response time represents the amount of calibration to be included in the model. Apart from the batch workload, the values for the interactive and update workload are quite small.

3. The overall CPU and disk resource utilization between the measured and predicted values agree. Thus, it is not prudent to make adjustments to CPU or disk values in the **transaction profiles** of the workloads. Therefore, we opt to calibrate the three workloads using **Additional delays** within the respective **Functions**.

```
                                 Change Function

 Workload  . . . . :   DIALOG      Measured from SAPDBS (Q972721400)
 Function  . . . . :   DIALOG

 Change fields, press Enter.
   Function text . . . . . . . . . . .   Function of DIALOG
   Key/Think time  . . . . . . . . . .   N/A            Seconds
   Additional delays . . . . . . . . .   .4             Seconds

 Transaction   Pool                   Transactions      CPU Time         Total
     Type       ID      Priority      per Function       (Secs)           I/Os
      2          2        19               8602.00        20.813          35.4
      2          1         0                  1.00       557.740       62251.2


                                                                            Bottom
Transaction Type: 1=Interactive, 2=Non-interactive

F3=Exit   F6=Work with transactions   F12=Cancel
```

4. Repeat the previous step for the UPDATE and BATCH workloads.

5. After completing the calibration, we return to the **Work with BEST/1 Model** display and select option 5 to **Analyze current model** and **Display Workload Report**.

```
                            Display Workload Report

Period:   Analysis
                     CPU     Thruput  -------Response Times (Secs)-------
Workload    Type    Util    per Hour  Internal    Local      LAN      WAN
QDEFAULT     1         .2        384       .3        .0       .3       .0
BATCH        2      .  .7     .   61   . 11.0      10.0       .0       .0
DIALOG       2      .32.1     . 8602   .  1.0       1.0       .0       .0
ENQ_SPL      2         .2        109       .5        .5       .0       .0
QDEFAULT     2        4.0       1818       .6        .0       .6       .0
UPDATE       2      . 4.0     . 1339   .   .7        .7       .0       .0

                                                                       Bottom
Type: 1=Interactive, 2=Non-interactive, 3=*BATCHJOB
Performance estimates -- Press help to see disclaimer.
F3=Exit   F10=Re-analyze   F11=Response time detail      F12=Cancel
F15=Configuration menu     F17=Analyze multiple points   F24=More keys
```

Acceptable BEST/1 Model

We now have a BEST/1 Model that is consistent with the measured values with respect to:

- AS/400 resource utilization
- Dialog steps processed per hour
- Average response times per dialog step

6. We can now proceed with estimating the impact of growth in workloads and the effect of making changes to the hardware configuration.

8.3.1.10 Growth Prediction with SAP R/3 Centralized Model

1. From the **Work with BEST/1 Models** display, select the required model using option 5.

2. From the **Work with BEST/1 Model** display of the selected model, select option 7 **Specify workload growth and analyze model**.

3. In this example, we chose not to have BEST/1 determine the configuration, and we specify 10 periods of growth of 10% each.

```
                       Specify Growth of Workload Activity

 Type information, press Enter to analyze model.
   Determine new configuration . . . . . . . . . .  N      Y=Yes, N=No
   Periods to analyze  . . . . . . . . . . . . . .  10     1 - 10

   Period 1  . . . . . . . . . . . . .  Period 1      Name
   Period 2  . . . . . . . . . . . . .  Period 2      Name
   Period 3  . . . . . . . . . . . . .  Period 3      Name
   Period 4  . . . . . . . . . . . . .  Period 4      Name
   Period 5  . . . . . . . . . . . . .  Period 5      Name

             ------Percent Change in Workload Activity-------
 Workload    Period 1  Period 2  Period 3  Period 4  Period 5
 *ALL              .0      10.0      10.0      10.0      10.0


                                                                   Bottom
 F3=Exit   F11=Specify growth by workload   F12=Cancel
 F13=Display periods 6 to 10   F17=Analyze using ANZBESTMDL
```

4. The following "composite" display using the **Display analysis** option shows the projected impact of workload growth on the current AS/400 configuration.

```
                             Display Analysis Summary

                           Stor     CPU  -Disk IOPs--  -Disk Ctls--  -Disk Arms--
 Period     CPU Model      (MB)    Util   Nbr    Util   Nbr    Util   Nbr    Util
 Period 1   53S 2157       2048    41.1     3    11.6    34      .4    34     3.7
 Period 2   53S 2157       2048    45.2     3    12.8    34      .4    34     4.0
 Period 3   53S 2157       2048    49.7     3    14.1    34      .4    34     4.4
 Period 4   53S 2157       2048    54.7     3    15.5    34      .5    34     4.9
 Period 5   53S 2157       2048    60.2     3    17.0    34      .5    34     5.4
 Period 6   53S 2157       2048    66.2     3    18.7    34      .6    34     5.9
 Period 7   53S 2157       2048    72.8     3    20.7    34      .6    34     6.5
 Period 8   53S 2157       2048    80.1     3    22.7    34      .7    34     7.2
 Period 9   53S 2157       2048    88.1     3    25.2    34      .8    34     7.9
 Period10   53S 2157       2048    97.0     3    27.7    34      .8    34     8.7
                                                                          Bottom
```

5. The following "composite" display shows the projected impact of workload growth on the current average response times. We pressed **F11** to view the **non-interactive** information.

```
            --Non-Inter Rsp Time--  -----Non-Inter------  Release
 Period      Local     LAN     WAN  CPU Util    Trans/Hr   Level
 Period 1      1.0      .6      .0     40.9       12074  V3R7M0
 Period 2      1.0      .6      .0     45.0       13281  V3R7M0
 Period 3      1.0      .6      .0     49.5       14610  V3R7M0
 Period 4      1.0      .6      .0     54.5       16070  V3R7M0
 Period 5      1.0      .6      .0     59.9       17678  V3R7M0
 Period 6      1.1      .7      .0     65.9       19445  V3R7M0
 Period 7      1.1      .7      .0     72.5       21390  V3R7M0
 Period 8      1.3      .9      .0     79.8       23529  V3R7M0
 Period 9      1.6     1.2      .0     87.7       25882  V3R7M0
 Period10      4.0     7.2      .0     96.5       28470  V3R7M0
                                                                          Bottom
F3=Exit   F10=Re-analyze   F11=Alternative view        F12=Cancel
F15=Configuration menu     F17=Analyze multiple points  F24=More keys
```

The previous information from the Display Analysis displays is used in Figure 75 on page 159. While there is a linear increase in the utilization of CPU, disk arms, and disk IOPs, the relative rate of increase (slope) is greatest with the CPU in this case. This indicates that the CPU is the first resource to reach saturation. The response time curve reaches its "knee" at approximately 25 000 ds per hour after which response degraded significantly.

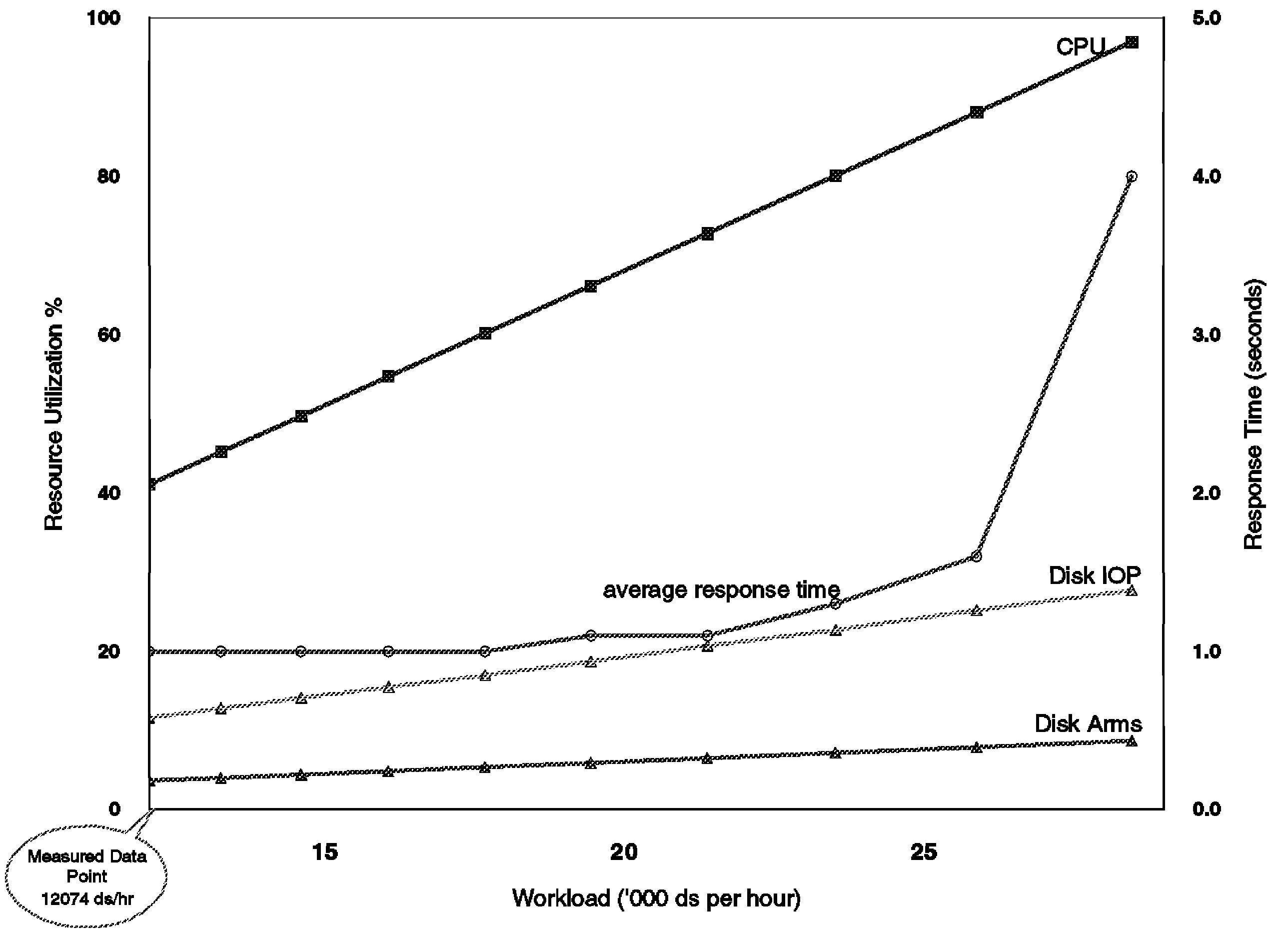


Figure 75. BEST/1 AS/400 Resources and Response Time Estimates

6. Additional estimates from BEST/1 (by changing the AS/400 configuration in the BEST/1 model) for CPU utilization and average response time for AS/400 model S20-2166 and S30-2259 are summarized in the following graph (details of the BEST/1 output are not shown).

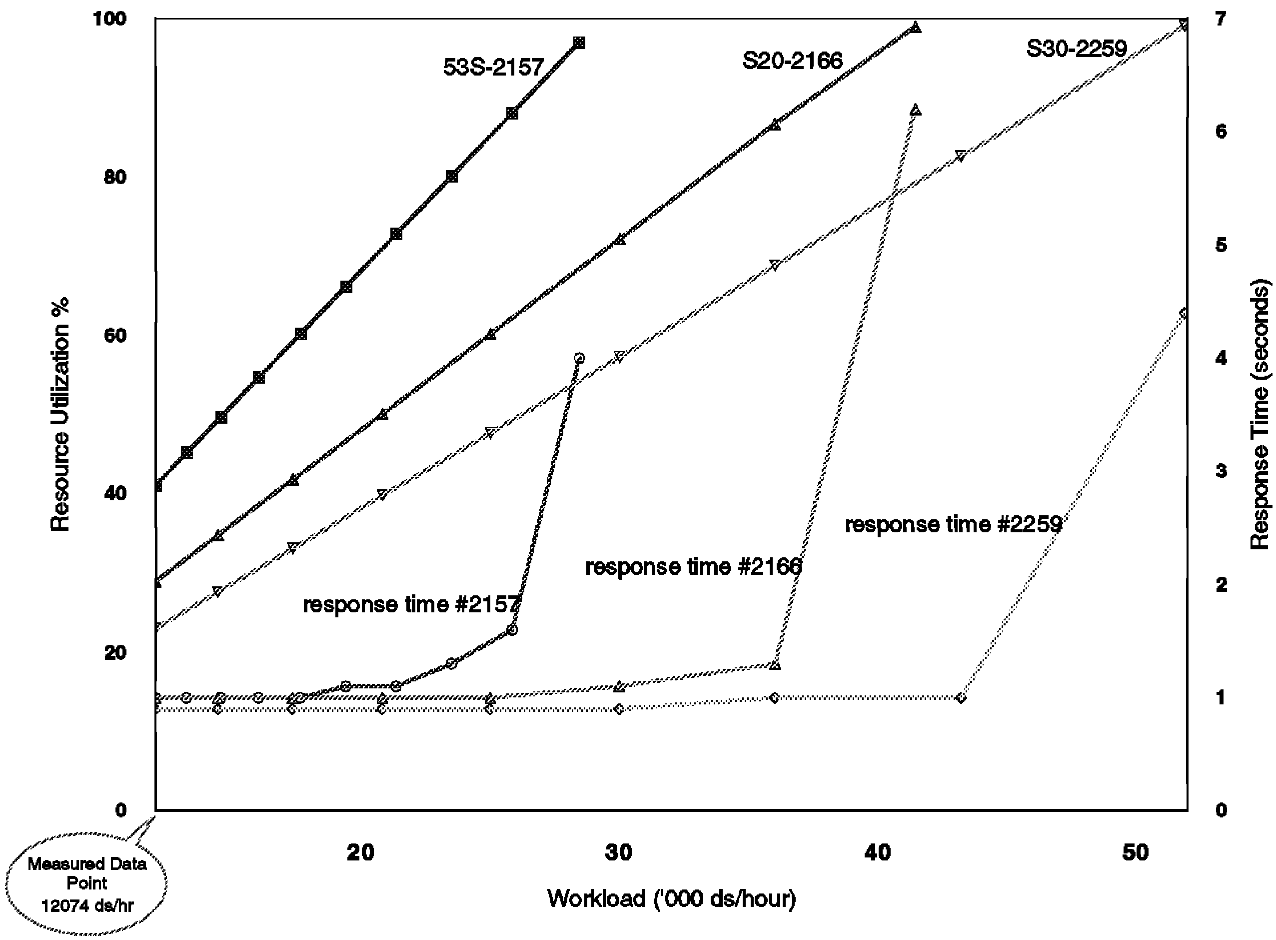


Figure 76. BEST/1 AS/400 Response Time Estimates versus CPU % on Selected Models

The graph highlights the increased workloads that can be supported by the various processors before the "knee" of the response time curve is reached.

7. The following "composite" display using the **Workload Report** option shows the effect on each of the server workloads in terms of response time and dialog steps.

Display Workload Report

Period: Period 1

Workload	Type	CPU Util	Thruput per Hour	Response Times (Secs) Internal	Local	LAN	WAN
QDEFAULT	1	.2	384	.3	.0	.3	.0
BATCH	2	.7	54	10.0	10.0	.0	.0
DIALOG	2	32.1	8602	1.0	1.0	.0	.0
ENQ_SPL	2	.2	109	.5	.5	.0	.0
QDEFAULT	2	4.0	1818	.6	.0	.6	.0
UPDATE	2	4.0	1339	.7	.7	.0	. 0

Period: Period 2

Workload	Type	CPU Util	Thruput per Hour	Response Times (Secs) Internal	Local	LAN	WAN
QDEFAULT	1	.2	423	.3	.0	.3	.0
BATCH	2	.7	59	10.1	10.1	.0	.0
DIALOG	2	35.3	9462	1.0	1.0	.0	.0
ENQ_SPL	2	.2	120	.6	.6	.0	.0
QDEFAULT	2	4.4	1999	.6	.0	.6	.0
UPDATE	2	4.4	1473	.7	.7	.0	.0

Period: Period 3

Workload	Type	CPU Util	Thruput per Hour	Response Times (Secs) Internal	Local	LAN	WAN
QDEFAULT	1	.2	465	.3	.0	.3	.0
BATCH	2	.8	65	10.2	10.2	.0	.0
DIALOG	2	38.9	10408	1.0	1.0	.0	.0
ENQ_SPL	2	.3	131	.6	.6	.0	.0
QDEFAULT	2	4.8	2199	.6	.0	.6	.0
UPDATE	2	4.8	1620	.7	.7	.0	.0

Period: Period 4

Workload	Type	CPU Util	Thruput per Hour	Response Times (Secs) Internal	Local	LAN	WAN
QDEFAULT	1	.2	512	.3	.0	.3	.0
BATCH	2	.9	72	10.3	10.3	.0	.0
DIALOG	2	42.7	11449	1.0	1.0	.0	.0
ENQ_SPL	2	.3	145	.6	.6	.0	.0
QDEFAULT	2	5.3	2419	.6	.0	.6	.0
UPDATE	2	5.3	1782	.8	.8	.0	.0

Period: Period 5

Workload	Type	CPU Util	Thruput per Hour	Response Times (Secs) Internal	Local	LAN	WAN
QDEFAULT	1	.3	563	.3	.0	.3	.0
BATCH	2	1.0	79	10.6	10.6	.0	.0
DIALOG	2	47.0	12594	1.0	1.0	.0	.0
ENQ_SPL	2	.3	159	.6	.6	.0	.0
QDEFAULT	2	5.8	2661	.6	.0	.6	.0
UPDATE	2	5.8	1960	.8	.8	.0	.0

Period: Period 6

Workload	Type	CPU Util	Thruput per Hour	Internal	Local	LAN	WAN
				-------Response Times (Secs)-------			
QDEFAULT	1	.3	619	.4	.0	.4	.0
BATCH	2	1.1	87	11.0	11.0	.0	.0
DIALOG	2	51.7	13854	1.1	1.1	.0	.0
ENQ_SPL	2	.3	175	.7	.7	.0	.0
QDEFAULT	2	6.4	2927	.7	.0	.7	.0
UPDATE	2	6.4	2156	.9	.9	.0	.0

Period: Period 7

Workload	Type	CPU Util	Thruput per Hour	Internal	Local	LAN	WAN
				-------Response Times (Secs)-------			
QDEFAULT	1	.3	681	.5	.0	.5	.0
BATCH	2	1.2	96	11.6	11.6	.0	.0
DIALOG	2	56.9	15239	1.1	1.1	.0	.0
ENQ_SPL	2	.4	192	.8	.8	.0	.0
QDEFAULT	2	7.1	3220	.7	.0	.7	.0
UPDATE	2	7.0	2372	1.1	1.1	.0	.0

Period: Period 8

Workload	Type	CPU Util	Thruput per Hour	Internal	Local	LAN	WAN
				-------Response Times (Secs)-------			
QDEFAULT	1	.3	749	.6	.0	.6	.0
BATCH	2	1.3	105	13.0	13.0	.0	.0
DIALOG	2	62.6	16763	1.2	1.2	.0	.0
ENQ_SPL	2	.4	212	1.0	1.0	.0	.0
QDEFAULT	2	7.8	3542	.9	.0	.9	.0
UPDATE	2	7.7	2609	1.4	1.4	.0	.0

Period: Period 9

Workload	Type	CPU Util	Thruput per Hour	Internal	Local	LAN	WAN
				-------Response Times (Secs)-------			
QDEFAULT	1	.4	824	1.0	.0	1.0	.0
BATCH	2	1.4	116	18.6	18.6	.0	.0
DIALOG	2	68.8	18439	1.3	1.3	.0	.0
ENQ_SPL	2	.4	233	1.9	1.9	.0	.0
QDEFAULT	2	8.5	3896	1.2	.0	1.2	.0
UPDATE	2	8.5	2870	2.8	2.8	.0	.0

Period: Period10

Workload	Type	CPU Util	Thruput per Hour	Internal	Local	LAN	WAN
				-------Response Times (Secs)-------			
QDEFAULT	1	.4	907	3.0	.0	3.0	.0
BATCH	2	1.6	127	61.3	61.3	.0	.0
DIALOG	2	75.7	20283	2.3	2.3	.0	.0
ENQ_SPL	2	.5	256	8.9	8.9	.0	.0
QDEFAULT	2	9.4	4286	7.2	.0	7.2	.0
UPDATE	2	9.3	3157	12.8	12.8	.0	.0

8. The following graph tracks the change in response time and CPU utilization of the interactive (DIALOG) workload with increasing workload volumes (10% period-on-period).

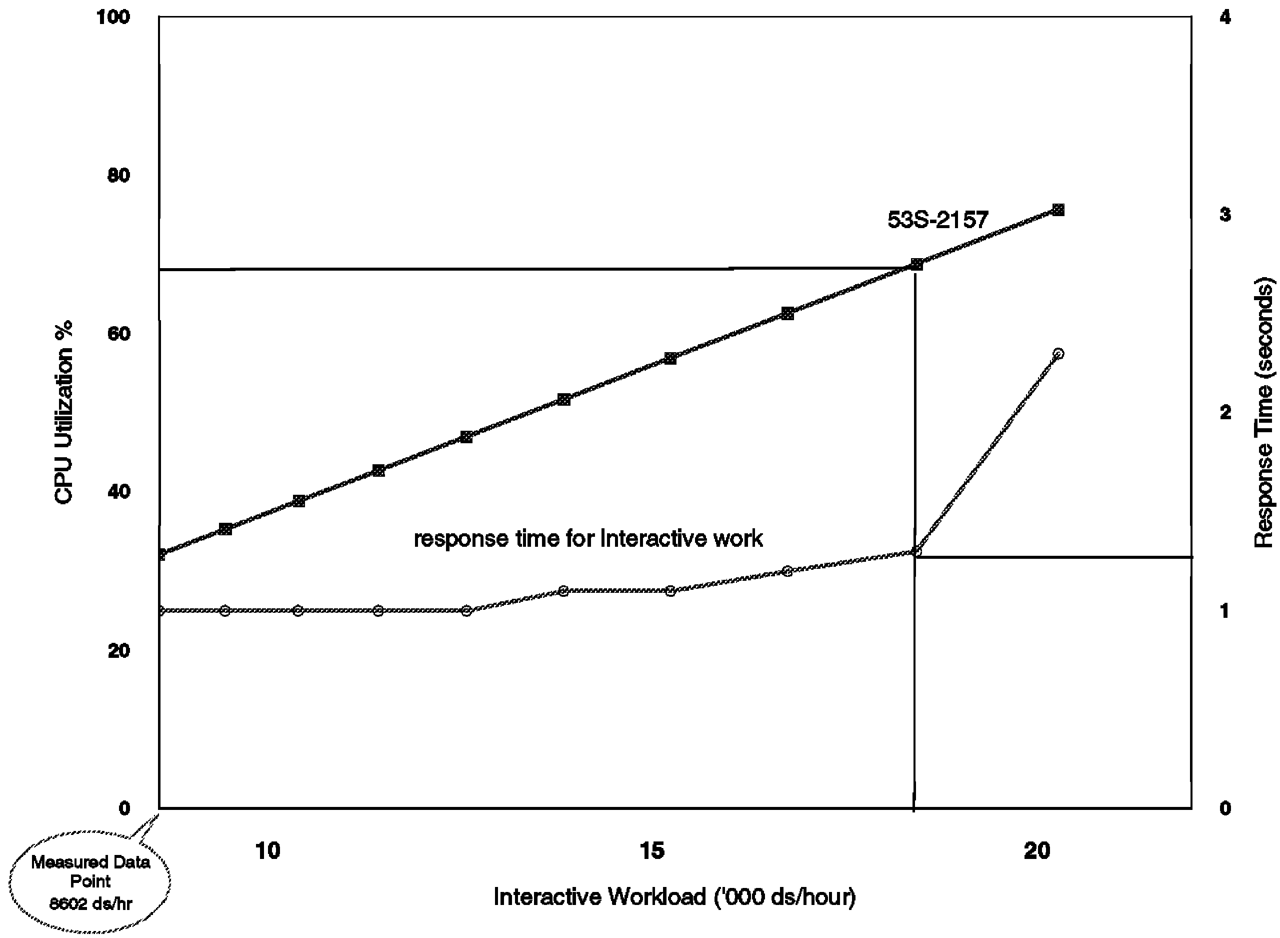


Figure 77. BEST/1 Interactive Response/CPU Utilization versus Workload

8.3.2 Normalizing SAP R/3 Dialog Steps

In initial SAP R/3 sizing estimates, each of the module dialog steps are "normalized" to a basic unit corresponding to a standard Financial Module dialog step (referred to as a **normalized FI dialog step**) using standard weighting factors based on experience.

However, the flexibility of SAP R/3 application architecture allows the application modules to be customized during implementation. This can result in the workload of any dialog step being changed (increased or decreased) in any particular customer situation. Also, the flexibility in usage may further change the resource utilization of a dialog step.

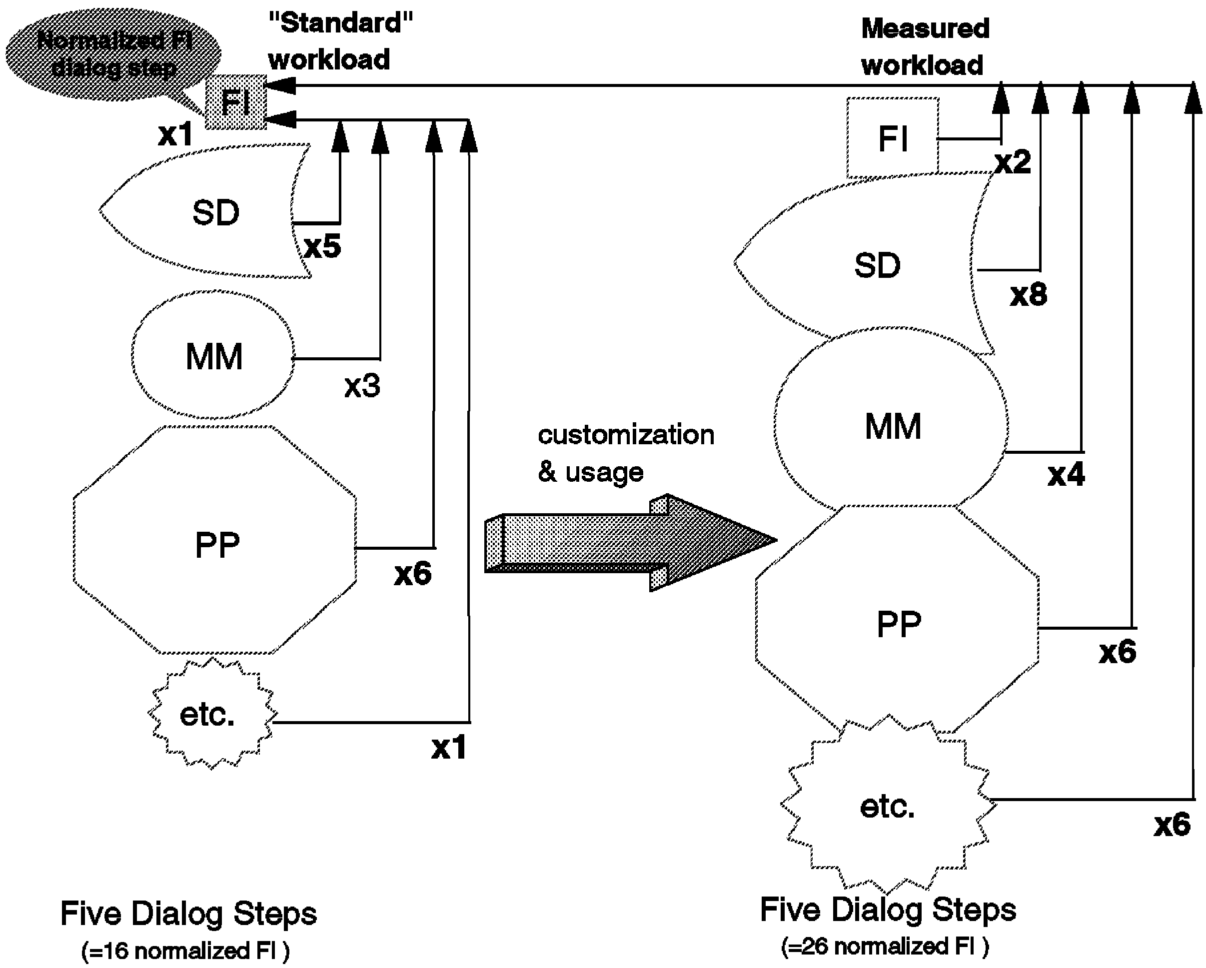


Figure 78. Normalizing Concept in SAP R/3

Figure 78 attempts to clarify the potential impact of customization and flexibility of usage on workload. Each "shape" represents a dialog step of a particular module such as Sales and Distribution (SD), Materials Management (MM), Production Planning (PP), and so on. Assume the approximate size of each object represents the resource usage by each of the dialogs. The numbers in the diagram represent the workload weighting based on relative complexity of a "standard" dialog step compared to a "standard" Financial (FI) dialog step.

In the "standard" environment on left-hand side, there are five dialog steps that equate to 16 normalized FI dialog steps.

In the hypothetical customized example on the right-hand side, the relative workload of the dialogs in relation to the "standard" Financial dialog is different than the "standard" workloads.

Thus, following customization, the normalization process results in 26 dialog steps, even though the total dialog steps is the same as the "standard" because the relative weight between the modules is different, and the relationship to a "standard" FI dialog is also different. The same reasons result in situations where the specific workload is lower than the average.

If you need to compare relative complexities of dialog steps, we advise you to examine the CPU seconds per dialog and the number of disk accesses resulting from it. Naturally, you need to give due consideration to the AS/400 processor capacity also. Ensure that extreme conditions (such as object locks, high memory faulting, and so on) do not result in abnormal usage values.

Appendix A. Working Paper on AS/400 Performance Management

This paper has been prepared to provide some guidance on how to manage the performance of an IBM AS/400 computer system. The paper has been written for installations of only one machine but the ideas can be easily extended and modified to cover multiple machine installations.

It has been assumed that the IBM Performance Tools/400 Manager Version (V3R2 5763-PT1, V3R7 5716-PT1, or V4R1 5769-PT1) is installed and that the reader is familiar with the general operations of the AS/400 system.

A.1 Objective

The objective of the paper is to set out a series of actions that should be applied to manage the performance of an IBM AS/400 Computer System.

A.2 AS/400 Performance Management

This can be considered under the following main headings:

A.2.1 Performance Data Collection

Set up the automatic performance data collection so that it is monitoring the system for all of the prime shift (say 08:00 to 20:00). The time interval should be set to 15 minutes and TRACE data should **not** be specified. (The Trace option should only be specified when a performance analysis is being undertaken.) The performance data collection should be set so that the data is written to a data library other than the Performance Tools library, QPFRDATA. In other words, the performance data collection should specify a library name such as DAYPERDATA.

We suggest that the daily performance data members are kept for up to two weeks. At the end of the two weeks, the extract of the historic data should be carried out, and the data members that are more than two weeks old can then be deleted from the system. See Section A.3.1, “Instructions for Archiving Performance Data” on page 169 for instructions on how to carry out these archiving steps.

A.2.2 Systems Tuning

Basic Systems Tuning: This should be done to ensure that the AS/400 main storage has been correctly allocated to the various storage pools. The steps involved are described in Chapter 14 of the *AS/400 Work Management Manual*, SC41-3306, but they can be summarized as follows:

1. Ensure that the non-DB faulting in the machine pool is less than 10.0 for "Good Performance".
2. Ensure that the total faulting in each of the other pools is less than the guideline figure (see the *Work Management manual*), for example, this figure is 50 for "Good Performance" on a 500/2141 AS/400 system.
3. Ensure that the total faulting in all pools on the system is less than the guideline figure (see the *Work Management* manual), for example, this is 75 for "Good Performance" on a 500/2141 AS/400 system.

© Copyright IBM Corp. 1998

4. Check that activity levels are set correctly by calculating the value of W-I/A-W transitions, and it should be less than 0.1, although it should not be zero.

All of these checks should be carried out at a "busy time" for the system (say 11:00 to 12:00 or 15:00 to 16:00) and should be based on WRKSYSSTS readings over about 5 minutes.

On-going Check on Tuning: This can most easily be done by running the "Advisor" facility once every month and reviewing its output.

A.2.3 Determine Workload Profile

Using standard Performance Tools/400 facilities, produce graphs for each performance member for two weeks of data. Examine the graphs and decide on:

- The busiest day of the week
- The busiest hour of the week

This procedure should be repeated once every six months or after any significant change in user working patterns.

A.2.4 Workload Monitoring

This can be done in two ways:

1. Produce the Performance Tools/400 "System Report" for the busiest hour of the busiest day of each week and complete the spread sheet tables shown in Section A.4, “Sample Spread Sheet” on page 169. This allows the utilization of the other aspects of the system to be monitored against guidelines. These guidelines are as follows:

Table 11. Key Performance Factors

High Priority CPU (for Single CPU System)	**< 70%**
Main storage (memory)	See faulting Guidelines
Disk Arm Utilization	< 50%
Remote Lines	< 30%

Note: * - Where the application code being used is complex, simply keeping the disk arm utilization below 50% may not be adequate due to the high number of SDIOs per transaction. We recommend, therefore, in addition to monitoring this parameter, that you carefully review the disk response time and make every attempt to keep this figure below 15 milliseconds per I/O.

2. Produce the following three historic graphs on an IPDS printer:
   - Transactions per hour (QIBMTNS)
   - Interactive Response Time (QIBMRSP)
   - Percentage of disk occupied (QIBMTOTDSK)

Individual readings are not as important as the "trends" that these two approaches show.

A.2.5 Capacity Planning

Once the workload trend has been established as previously described, the Best/1 Capacity Planning tool can be used to see how long it will be before the response times deteriorate to an unacceptable level. The Best/1 tool is an integral part of the Performance Tools/400 product. The Best/1 model should be reviewed every six months or whenever there is a dramatic change in workload.

A.3 Performance Data Archiving

The following sections provide instructions on managing the volume of performance data by archival.

A.3.1 Instructions for Archiving Performance Data

1. Type GO PERFORM to get to the "IBM Performance Tools /400" menu.
2. Select Option 9 to get to the "Performance Tools Graphics" menu.
3. Select Option 2 to get to the "Work with Historical Data" display.
4. Enter the name of library that performance data collection is writing data to at the "Library" prompt.
5. Type a "1" beside all members that are over one week old and press Enter to start a batch job to "Create Historical Data".
6. When batch jobs have completed (check by using "Work with Historical Data" to ensure that "YES" is in the column headed "Historical Data" for all selected members), it is safe to delete performance data members.

A.3.2 Instructions to Delete Performance Data Members

1. Type GO PERFORM to get to the "IBM Performance Tools /400" menu.
2. Select Option 6 to get to the "Configure and Manage Tools" menu.
3. Select Option 2 to get to the "Delete Performance Data" display.
4. Enter the name of the library that performance data collection is writing data to at the "Library" prompt.
5. Type a "4" beside the members that are over one week old and have successfully had their historic data extracted (see the previous instructions). Press Enter to start batch jobs that delete the performance data members but leave the historic data intact.

A.4 Sample Spread Sheet

Extracts from system reports produced for "peak hour" on the busiest day of the week can be used to populate a simple "Key Performance Factors" spread sheet or a slightly more detailed "Performance Factors" spread sheet shown in the following table.

Table 12 (Page 1 of 2). Key Performance Factors

Factors	**Comment**	**Page of System Report**	**Example**
Average Response Time (Secs)	General	1	1.09
Total CPU Utilization	General	1	83.3%

Table 12 (Page 2 of 2). Key Performance Factors			
Factors	**Comment**	**Page of System Report**	**Example**
Transactions/Hour	Applied Workload	2	6,854
Avg. DBIO/Transaction	Application Complexity	2	35
Total high priority (System + Inter) Guidelines<70%	CPU	3	77
Total faults in System - Guidelines < 75 for Model 500/2141	Memory	4	37.9
Response Time (Millisecs)	Disk	5	9.1

The "Key Performance Factors" spread sheet can be particularly useful for reporting to management on overall trends, while the "Performance Factors" spread sheet can help the staff responsible for the system operation to monitor and understand its performance.

Overall System Readings	Key Indicator	Page of System Report	Example - 15 Sept. 1997
Average Response Time (Secs)	*	1	1.09
Total CPU Utilization	*	1	83.3%
Batch LDBIO/Sec		1	

Interactive Workload Profile	Key Indicator	Page of System Report	Example - 15 Sept. 1997
Transactions/Hour	*	2	6,854
Average CPU Secs/Transaction		2	.31
Average SDIO/Transaction		2	19.0
Average DBIO/Transaction	*	2	35.0

CPU Breakdown	Key Indicator	Page of System Report	Example - 15 Sept. 1997
System		3	.9.4
Interactive		2	60.0
Total high priority (<21 - Guideline < 70%)	*	3	77.0
Low Priority Batch (50)		3	2.8

Main Storage	Key Indicator	Page of System Report	Example - 15 Sept. 1997
Non DB faults in MCH Pool - Guideline < 10		4	8.0

Main Storage	Key Indicator	Page of System Report	Example - 15 Sept. 1997
Total faults in System - Guideline <75 for model 500-2141		4	37.9

Disk Performance	Key Indicator	Page of System Report	Example - 15 Sept. 1997
Space Utilization - Guideline <85%		5	86.4
Arm Utilization - Guideline <50%		5	5.1
Arm Response Time (Millisecs)	*	5	9.1

Table 13. Performance Factors

Appendix B. Special Notices

This publication is intended to help Customers, IBM Business Partners and IBM Specialists to do capacity planning for server applications running on AS/400 systems or servers. The information in this publication is not intended as the specification of any programming interfaces that are provided by OS/400 and the AS/400 Performance Tools. See the PUBLICATIONS section of the IBM Programming Announcement for the AS/400 Performance Tools for more information about what publications are considered to be product documentation.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to the IBM Director of Licensing, IBM Corporation, 500 Columbus Avenue, Thornwood, NY 10594 USA.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The information about non-IBM ("vendor") products in this manual has been supplied by the vendor and IBM assumes no responsibility for its accuracy or completeness. 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.

Any performance data contained in this document was determined in a controlled environment, and therefore, the results that may be obtained in other operating environments may vary significantly. Users of this document should verify the applicable data for their specific environment.

Reference to PTF numbers that have not been released through the normal distribution process does not imply general availability. The purpose of including these reference numbers is to alert IBM customers to specific information relative to the implementation of the PTF when it becomes available to each customer according to the normal IBM PTF distribution process.

The following terms are trademarks of the International Business Machines Corporation in the United States and/or other countries:

APPN®	AS/400®
Client Access	DB2®
IBM®	IPDS
OS/400®	400®

The following terms are trademarks of other companies:

C-bus is a trademark of Corollary, Inc.

Java and HotJava are trademarks of Sun Microsystems, Incorporated.

Microsoft, Windows, Windows NT, and the Windows 95 logo are trademarks or registered trademarks of Microsoft Corporation.

PC Direct is a trademark of Ziff Communications Company and is used by IBM Corporation under license.

Pentium, MMX, ProShare, LANDesk, and ActionMedia are trademarks or registered trademarks of Intel Corporation in the U.S. and other countries.

UNIX is a registered trademark in the United States and other countries licensed exclusively through X/Open Company Limited.

Other company, product, and service names may be trademarks or service marks of others.

Appendix C. Related Publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

C.1 International Technical Support Organization Publications

For information on ordering these ITSO publications see "How to Get ITSO Redbooks" on page 177.

- *AS/400 Client/Server Performance using Windows Clients*, SG24-4526
- *AS/400 Performance Management V3R6/V3R7*, SG24-4735
- *AS/400 Communication Performance Investigation*, SG24-4669

C.2 Redbooks on CD-ROMs

Redbooks are also available on CD-ROMs. **Order a subscription** and receive updates 2-4 times a year at significant savings.

CD-ROM Title	Subscription Number	Collection Kit Number
System/390 Redbooks Collection	SBOF-7201	SK2T-2177
Networking and Systems Management Redbooks Collection	SBOF-7370	SK2T-6022
Transaction Processing and Data Management Redbook	SBOF-7240	SK2T-8038
Lotus Redbooks Collection	SBOF-6899	SK2T-8039
Tivoli Redbooks Collection	SBOF-6898	SK2T-8044
AS/400 Redbooks Collection	SBOF-7270	SK2T-2849
RS/6000 Redbooks Collection (HTML, BkMgr)	SBOF-7230	SK2T-8040
RS/6000 Redbooks Collection (PostScript)	SBOF-7205	SK2T-8041
RS/6000 Redbooks Collection (PDF Format)	SBOF-8700	SK2T-8043
Application Development Redbooks Collection	SBOF-7290	SK2T-8037

C.3 Other Publications

These publications are also relevant as further information sources:

- Version 4 Release 1 publications:
  - *Work Management APIs*, SC41-5878
  - *CL Reference*, SC41-5722
  - *Work Management*, SC41-5306
  - *Performance Tools for AS/400*, SC41-5340
  - *DB2 for OS/400 Database Programming*, SC41-5701
  - *Integrated File System Introduction*, SC41-5711
  - *Data Management*, SC41-5710
  - *DB2 for OS/400 SQL Programming*, SC41-5611
  - *Publication Reference Book*, SC41-5003
  - *System API Reference*, SC41-5801
- Version 3 Publications:
  - *OS/400 Work Management*, SC41-4306

© Copyright IBM Corp. 1998

– *OS/400 Server Concepts and Administration*, SC41-3740
– *AS/400 System Startup and Problem Handling*, SC41-4206
– *OptiConnect for AS/400*, SC41-4414
– *OptiMover for AS/400*, SC41-0626
– *Performance Tools/400*, SC41-4340
– *Performance Tools/400-Getting Started*, SC41-4343
– *DB2 for OS/400 Database Programming*, SC41-4701
– *OS/400 Integrated File System Introduction*, SC41-4711
– *OS/400 Data Management*, SC41-4710
– *AS/400 System Concepts*, GC41-9802
– *DB2 for OS/400 SQL Programming*, SC41-4611
– *AS/400 System Operation*, SC41-4203
– *Getting Started with AS/400*, SC41-4204
– *AS/400 Publication Reference Book*, SC41-4003
– *AS/400 System API Reference*, SC41-4801

How to Get ITSO Redbooks

This section explains how both customers and IBM employees can find out about ITSO redbooks, CD-ROMs, workshops, and residencies. A form for ordering books and CD-ROMs is also provided.

This information was current at the time of publication, but is continually subject to change. The latest information may be found at http://www.redbooks.ibm.com.

How IBM Employees Can Get ITSO Redbooks

Employees may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **PUBORDER** — to order hardcopies in United States
- **GOPHER link to the Internet** - type GOPHER.WTSCPOK.ITSO.IBM.COM
- **Tools disks**

  To get LIST3820s of redbooks, type one of the following commands:

  ```
  TOOLS SENDTO EHONE4 TOOLS2 REDPRINT GET SG24xxxx PACKAGE
  TOOLS SENDTO CANVM2 TOOLS REDPRINT GET SG24xxxx PACKAGE (Canadian users only)
  ```

  To get BookManager BOOKs of redbooks, type the following command:

  ```
  TOOLCAT REDBOOKS
  ```

  To get lists of redbooks, type one of the following commands:

  ```
  TOOLS SENDTO USDIST MKTTOOLS MKTTOOLS GET ITSOCAT TXT
  TOOLS SENDTO USDIST MKTTOOLS MKTTOOLS GET LISTSERV PACKAGE
  ```

  To register for information on workshops, residencies, and redbooks, type the following command:

  ```
  TOOLS SENDTO WTSCPOK TOOLS ZDISK GET ITSOREGI 1998
  ```

  For a list of product area specialists in the ITSO: type the following command:

  ```
  TOOLS SENDTO WTSCPOK TOOLS ZDISK GET ORGCARD PACKAGE
  ```

- **Redbooks Web Site on the World Wide Web**

  http://w3.itso.ibm.com/redbooks

- **IBM Direct Publications Catalog on the World Wide Web**

  http://www.elink.ibmlink.ibm.com/pbl/pbl

  IBM employees may obtain LIST3820s of redbooks from this page.

- **REDBOOKS category on INEWS**
- **Online** — send orders to: USIB6FPL at IBMMAIL or DKIBMBSH at IBMMAIL
- **Internet Listserver**

  With an Internet e-mail address, anyone can subscribe to an IBM Announcement Listserver. To initiate the service, send an e-mail note to announce@webster.ibmlink.ibm.com with the keyword subscribe in the body of the note (leave the subject line blank). A category form and detailed instructions will be sent to you.

Redpieces

For information so current it is still in the process of being written, look at "Redpieces" on the Redbooks Web Site (http://www.redbooks.ibm.com/redpieces.htm). Redpieces are redbooks in progress; not all redbooks become redpieces, and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

How Customers Can Get ITSO Redbooks

Customers may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **Online Orders** — send orders to:

	IBMMAIL	**Internet**
In United States:	usib6fpl at ibmmail	usib6fpl@ibmmail.com
In Canada:	caibmbkz at ibmmail	lmannix@vnet.ibm.com
Outside North America:	dkibmbsh at ibmmail	bookshop@dk.ibm.com

- **Telephone orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	(long distance charges apply)
(+45) 4810-1320 - Danish	(+45) 4810-1020 - German
(+45) 4810-1420 - Dutch	(+45) 4810-1620 - Italian
(+45) 4810-1540 - English	(+45) 4810-1270 - Norwegian
(+45) 4810-1670 - Finnish	(+45) 4810-1120 - Spanish
(+45) 4810-1220 - French	(+45) 4810-1170 - Swedish

- **Mail Orders** — send orders to:

IBM Publications	IBM Publications	IBM Direct Services
Publications Customer Support	144-4th Avenue, S.W.	Sortemosevej 21
P.O. Box 29570	Calgary, Alberta T2P 3N5	DK-3450 Allerød
Raleigh, NC 27626-0570	Canada	Denmark
USA		

- **Fax** — send orders to:

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	(+45) 48 14 2207 (long distance charge)

- **1-800-IBM-4FAX (United States)** or **(+1)001-408-256-5422 (Outside USA)** — ask for:

  Index # 4421 Abstracts of new redbooks
  Index # 4422 IBM redbooks
  Index # 4420 Redbooks for last six months

- **Direct Services** - send note to `softwareshop@vnet.ibm.com`

- **On the World Wide Web**

Redbooks Web Site	http://www.redbooks.ibm.com
IBM Direct Publications Catalog	http://www.elink.ibmlink.ibm.com/pbl/pbl

- **Internet Listserver**

  With an Internet e-mail address, anyone can subscribe to an IBM Announcement Listserver. To initiate the service, send an e-mail note to `announce@webster.ibmlink.ibm.com` with the keyword `subscribe` in the body of the note (leave the subject line blank).

Redpieces

For information so current it is still in the process of being written, look at "Redpieces" on the Redbooks Web Site (`http://www.redbooks.ibm.com/redpieces.htm`). Redpieces are redbooks in progress; not all redbooks become redpieces, and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

IBM Redbook Order Form

Please send me the following:

Title	Order Number	Quantity

First name Last name

Company

Address

City Postal code Country

Telephone number Telefax number VAT number

□ Invoice to customer number

□ Credit card number

Credit card expiration date Card issued to Signature

We accept American Express, Diners, Eurocard, Master Card, and Visa. Payment by credit card not available in all countries. Signature mandatory for credit card payment.

List of Abbreviations

APA all points addressable

IBM International Business Machines Corporation

ITSO International Technical Support Organization

PROFS Professional Office System

Index

© Copyright IBM Corp. 1998

T

U

V

W

ITSO Redbook Evaluation

AS/400 Server Capacity Planning
SG24-2159-00

Your feedback is very important to help us maintain the quality of ITSO redbooks. **Please complete this questionnaire and return it using one of the following methods:**

- Use the online evaluation form found at http://www.redbooks.ibm.com
- Fax this form to: USA International Access Code + 1 914 432 8264
- Send your comments in an Internet note to redbook@vnet.ibm.com

Please rate your overall satisfaction with this book using the scale:
(1 = very good, 2 = good, 3 = average, 4 = poor, 5 = very poor)

Overall Satisfaction ____________

Please answer the following questions:

Was this redbook published in time for your needs? Yes____ No____

If no, please explain:

____

____

____

____

What other redbooks would you like to see published?

____

____

____

Comments/Suggestions: (THANK YOU FOR YOUR FEEDBACK!)

____

____

____

____

____

© Copyright IBM Corp. 1998

AS/400 Server Capacity Planning **SG24-2159-00**

SG24-2159-00
Printed in the U.S.A.