BACK TO THE
TABLE OF CONTENTS

A CAPACITY PLANNING APPROACH FOR THE NEWCOMER: WHERE DO I START?

Daniel F. Schwarz, University of Wisconsin Hospital and Clinics
(dan.Schwarz@hosp.wisc.edu)
Sue Rohr, Information Resources, Inc.
(sue.rohr@infores.com)

This paper provides a starting point for a new Capacity planner/Performance analyst. Eight "rules" are presented that should provide some guidance on how to approach this new job. A bibliography of documents that should be read by every CP/PA is included.

Introduction.
Last week your boss called you into her office and told you that you were the new capacity planner and/or performance analyst (CP/PA) for the organization. Perhaps this was a job that you sought. Perhaps it was one that you had thrust upon you. You had been a systems programmer or a systems administrator. You were pretty good at what you did. Now you have some totally new responsibilities. What do you do now? Where do you start? That is the topic of this paper, to answer the question: "What do I do on Monday?"

Here are some rules that should help you get oriented to your new job. They are:

Rule 1: Make sure you understand the question.
Rule 2: Communicate both up and down on the organization chart.
Rule 3: You are overhead.
Rule 4: Manage your performance data.
Rule 5: Document everything!
Rule 6: Use the right tool for the job.
Rule 7: There is help available.
Rule 8: Use reporting techniques geared to your audience.

As you read these sections, keep in mind that no two CP/PA's will come up with the same set of rules. Everyone does it differently. Everyone charts his/her own course. A lot of this job is science. But much of it is art.

Rule 1: Make sure you understand the question.

Capacity planning and performance analysis are two different jobs. They have some things in common, but they have some big differences too. It is critical that you know which one you are. People often confuse them because many enterprises combine the two functions into one position.

A Capacity Planner is asked to answer the question: when will our resources be insufficient to handle the demands of our business? A similar question is: how many resources must we purchase in order to perform these new functions?

> "Capacity planning is the process of identifying a system configuration to provide satisfactory performance for projected future workloads." [SACK94]

A Performance Analyst answers the question: What changed? Why did we meet our service goals Tuesday but not on Wednesday?

> "Performance management is the process of configuring the current system to provide satisfactory performance for current workloads." [SACK94]

What they have in common is that they use the same data and they rely on some of the same statistical metrics in doing the analysis. Both jobs require knowledge about how many resources are absorbed by existing workloads. Capacity Planners need to know how many resources are absorbed by each business function in order to project when the workload will demand more than is available. Performance Analysts need to know how many resources are absorbed by workloads when they are running normally. Knowing this will allow the PA to identify when they are not running normally.

Data must be collected in order to answer both questions. The data that needs to be collected is that which associates consumption of resources with the work itself. On a mainframe the data is provided by SMF and RMF from IBM or CMF from BMC. On open platforms there are other data sources such as PACCT, SAR, IOSTAT and VMSTAT that provide similar information.

Rule 2: Communicate both up and down on the organization chart.

In many cases CP/PA's were systems programmers, systems administrators in their previous lives. Systems people can often survive as solitary beasts, installing software packages, applying patches and sending e-mails to vendor help desks and rarely having to see actual people. CP/PA's cannot. As a CP/PA you will need to sell your findings to management and you will have to rely on good relations with applications developers in order to get the information you need to develop your projections.

Expect to meet with three people or groups on a regular basis. In fact, insist upon it.

First, identify the people in your organization who makes purchase decisions for the Information Systems department. These are the people you need to meet with on (at least) a monthly basis. Pick a small group of data that will at a glance deliver a snapshot view of where your organization is and where it is going with respect to resource consumption. You have two goals with this meeting.

1. **Eliminate Surprises**. Make sure that this manager is aware that there is (or is not) growth in your environment and that at some point new equipment will be required. Giving as much notice as possible will make budgeting possible.

2. **Track Resource Consumption.** Convince this manager to trust you as the source of what resources are being consumed and by whom.

Second, you also need to meet regularly with the applications development staff. You need to have as much warning as possible of new applications and systems, and changes to existing applications and systems. The applications staff knows when new applications and changes to existing applications will come on line. The applications staff will make it possible for you to instrument the new or changed applications in order to gather resource consumption metrics early in the development and implementation phases of the process.

Third, you must meet the person who forecasts your organization's business growth. Perhaps this is the CFO or the head of the accounting department. This person holds the keys to your forecasts of growth in the systems you are responsible for. Get to know this person very well.

As a systems programmer we never wanted our names to be known because only bad things could be associated with us. As a CP/PA that is no longer true. Visibility is now an important asset for you.

Rule 3: You are overhead.

Make sure you provide a benefit equal to or greater than your cost [RALS1]. There are some organizations where management truly believes in the value of CP/PA. In those organizations tools can be purchased and there is no struggle to prove ones worth. There are other organizations that expect to be shown that the costs of CP/PA to be returned in savings on hardware and software plus the added benefits that come from appropriately sized computer equipment on an annual basis. It is easy to see how the cost could be enormous. A typical CP/PA could cost $150,000/year in salary and benefits. 10 large NT servers can be purchased for that cost. Be aware that you will have to prove your worth to your employer.

In some cases it is easy. In e-commerce web applications a poorly performing web server could cost a company millions of dollars in lost business. If the computer is servicing staff on the phone with customers poor performance could cause customers to flee and unnecessary stress for staff. A good CP/PA can turn a disaster into a great victory.

In a case where traditional non-web data processing goes on, it may be harder to quantify the value of a good capacity plan because the end-users might be internal employees of your company. They might not threaten to take their business to a competitor if computer service is poor. In that case it might not be obvious that the costs of the business of lost productivity and wasted staff time are real. Those are your challenges.

Rule 4: Manage your performance data [PILC1].

Use good data processing techniques to minimize the cost of your processing. CP/PA generates huge amounts of data. Keeping that data in a form that allows data to be extracted is difficult and requires a great deal of planning.

One tip is to measure when it matters. Turning on measurement instruments at critical times and off again when those intervals have passed is one way to reduce the volume of data. Of course, it might not be obvious when those times are. Is your overnight batch critical? Is your e-commerce web site getting pounded from Hong Kong in the middle of the night? It is often a chicken and egg decision. You don't know when to measure until you measure.

Once you have determined your critical times you will need to create standard reporting periods. Be aware, however that a standard reporting period must be normalized so that lengths of time are consistent. For example, comparing February to May is likely to mislead if no adjustments are made to account for different number of days in those months.

There are three types of data that you need to gather. First is the metric you will use for evaluation against the service level objectives. This might be response times for various functions, the number of functions that are performed or the amount of availability of the system. You will use this information to determine if your systems are meeting their service goals and these are the values you will be projecting into the future in a capacity planning exercise.

The second type of data needed is the counts of business activity that is supported by the computer systems. Things such as customer contacts, orders, and invoices are business activities that might generate computer activity in your company.

The third type of data will be all the computer stuff, CPU consumption, I/O rates, memory usage etc. These values will be correlated with the business measurements to determine how many computer things each business thing your company performs consumes.

Combine that knowledge with:

- Information about future system and application changes
- Anticipated changes in business volume
- Technology to provide capacity requirements.

Now you have what is needed to perform capacity planning.

But there is a lot of data here, and keeping it in a form that allows access and analysis in a convenient form is the tricky part.

Rule 5: Document everything!

Keep track of what data you use for your analysis. Keep track of which procedures were used for each analysis you did. This data will come in handy down the road when you have to repeat the analysis or explain why different results were achieved with the same data.

As you present your data to various audiences there will be some who just want to see the results, and some who want to know every bloody detail of how your analysis was done. You need to be prepared to satisfy the most demanding of audiences.

Keep a daily diary in which you record details of what you did and how you did it for each analysis. This is comparable to your institution's change management system of tracking changes.

Rule 6: Use the right tool for the job.

There are many tools on all platforms that will generate, collect and analyze this data for you. Some of these tools you might have to purchase, others you might already have available to you. One of your first tasks will be to take an inventory of these tools and determine which might be of value to you. These tools will vary a great deal in terms of how they work and what they will do for you. A CP/PA can spend a lot of time evaluating and understanding tools. This is time well spent because you will learn a lot about CP/PA by doing so.

There are many popular tools. The functions these tools will perform are discussed below. If your shop doesn't have a tool to perform one of these functions, you will have to find a way to perform the function some other way or justify the cost of the tool.

Basic Tools

- Detailed measurement data on jobs, job steps, started tasks, on-lines, and system task overhead covering:
  - CPU time for all levels
  - DISK performance
  - Tape
  - Printers
  - Memory
  - Queue times

- A Report Writer tool to extract specifics from your measurement data
- Spread sheet with graphics
- A tool to compute basic statistics such as
  - CPU time for all levels
  - Regressions
  - Correlations
  - Q analysis (Little's Law etc.)
  - Means, Averages, Standard Deviations etc

Advanced tools

- Modeling
- Simulation
- End to end response time

There are four types of tools that are used to evaluate the capacity and performance of computer systems.

1. **Ruler and spreadsheet method**. Downloading the data to a spreadsheet and using the spreadsheet's power to tabulate, graph and analyze by applying functions or macros to the data is the least expensive method of doing CP/PA. It allows the analyst to gain an intimate feel for the data and it is quick and easy for some types of data. It has some limitations however. It is not a technique that will do any type of sophisticated modeling and handling huge amounts of data or data from many servers may overwhelm the analyst.

2. **Analytic model**. There are several vendors that sell analytic modeling software specifically designed to do CP/PA for computer systems. They vary in their sophistication but they all treat a computer as a series of queues. Each queue, processor, I/O etc., has a service time for each request and a computed wait time. An application is considered to be a certain number of requests to each queue for each unit of work that is performed. Then the system is modeled with the parameters specified by the analyst. Results can be very accurate and can be applied to a large number of applications and systems. They can provide the analyst with great insight. The packages vary in price, but price is sometimes an obstacle.

3. **Simulation**. Simulation is the most accurate and most expensive tool there is. It is used when there is no room for error, such as when evaluating software to be used in landing a man on the moon.

4. **Trace.** Trace has less value in capacity planning. It is more useful in performance analysis. Traces will allow analysts to look at an application at the level of the computer instructions to analyze where the computer is spending its time.

Rule 7: There is help available. [RALS1]

You are going to need some help and you are going to need to read a lot to learn about this job. Develop a mentor-like relationship with the person who had the job before you if possible. In addition, CMG and the Regions are an excellent way to meet people who can help you get started in the job and will allow you to develop a network that will help you throughout your career.

There are several web sites and list servers that are filled with useful information on CP/PA. CMG offers a list service on CP/PA issues. Sign up at www.cmg.org. Also recommended is the MXG list, MXG-L. You can sign up at www.mxg.com. This list is specifically oriented towards MXG issues, but they are tolerant of all performance and capacity issues.

The following is a suggested bibliography of books you will want to read. (This is an inventory of Richard Ralston's bookshelf):

CMG Proceedings & Transactions

IBM Manuals & Redbooks

Other Vendor manuals and publications

MXG Documentation, now found in the MXG source code PDS.

The Art of Computer Systems Performance Analysis, Raj Jain, John Wiley and Sons, Inc, 1991, ISBN 0-471-50336-3

Performance Engineering of Software Systems, Connie U. Smith, Ph.D., Addison-Wesley Publishing Co., 1990, ISBN 0-201-53769-9

The Visual Display of Quantitative Information, Edward R. Tufte, Graphics Press, ISBN 0-9613-9214-2

MVS Performance Management: OS/390 Edition: With MVS/ESA SP Version 5, Stephen L. Samson, McGraw Hill, 1997, ISBN 0-07-057700-5

How to Lie with Charts, Gerald E. Jones, Sybex, 1995, ISBN 0-7821-1723-6

Capacity Planning and Performance Modeling From Mainframes to Client-Server Systems, Daniel A. Menasce, Virgilio A.F. Almeida, & Larry W. Dowdy, Prentise Hall, 1994, ISBN 0-13-789546-1

Capacity Planning for Web Performance Metrics, Models, & Methods, Daniel A. Menasce & Virgilio A.F. Almeida, Prentice Hall, 1998, ISBN 0-13-693-822-1

MVS Answer Book, David J Sacks. John Wiley & Sons Inc., 1994, ISBN 0 471-60821-1

Windows 2000 Performance Guide by Mark Friedman, Odysseas Pentakalos O'Reilly & Associates, January 2002; ISBN 1-565-92466-5.

Rule 8: Use reporting techniques geared to your audience.

A large amount of your time on the job will consist of looking at the output from reports of various types of computer activity. After you have looked at the data enough you will become very familiar with it. At some point an epiphany may happen and you will have gained a fundamental bit of understanding of your computer environment. To share this knowledge with an audience without having the ability to guide them through the painstaking process of analysis that you went through is a difficult task because of time constraints and the level of detail required.

First, remember to gear your presentation to the technical level of your audience. Some audiences are very technical and want to see all the details of your analysis. If you fail to provide sufficient detail, your credibility will suffer. Other audiences are **not interested in details**; they just want to see the bottom line. If you have too much detail in your presentation to them, their eyes will glaze over **and they will miss** the points you are trying to make.

Keep tables to two or three rows and columns. Use color charts and graphs if at all possible. For capacity, you want to show a graph with monthly data including the following items:

1. **Total Available CPU Hours.** This is total shop capacity during peak processing times. Capacity can be in MIPS or Maximum Hours possible in the shop (Total hours in peak time slot times number of CP's in the machine). If you have multiple machines, normalize on the smallest. Then add each machine into the Total Capacity number.
2. **A Pain Threshold.** Take Total Available CPU Hours and subtract Operating and Subsystem Overhead. This is what your application workload can actually achieve before users start complaining. In UNIX you may have to collect war stories from management or users to determine the right percentage of overhead for your shop. Mainframe people can calculate the number from RMF data. Keep it simple and make it a single percent of total capacity for each configuration.
3. **Normalized Total Hours Used.** Actual hours used by all applications normalized for monthly and Holiday differences. You may also want to drop vertical lines from the 'normalized total hours used" line, marking significant

application changes that increased or decreased usage.

4. **A trend line**. Base your trend line on "normalized total hours used". Take it out at least 18 months or as far as management wants. But, remember you need as many points in the past as you project. Always tell your audience your confidence level. For example, if you do a trend on six months of data and project 18 months, make sure everyone knows the trend can change significantly as more data is gathered (even if nothing changes in your shop).

Note that each upgrade or decrease in capacity will cause an update to this graph. Your log of changes will be a blessing here.

Now you need to answer the boss's questions. "Why do I need capacity? Who is causing this?" For this you need a breakdown of usage by each major business application. Pie charts are wonderful for this and of course you'll need the detail for the really big application. Often there is no really big application you can pinpoint.

Proving growth is not always easy. To convince management there really is growth you may have to keep 13 or 24 months of data. With 13 months of data on each application, you can show a growth rate in hours for each. With 24 months you can chart last year over this year for applications. Having the 13-months of data can pinpoint the applications you want to graph.

Finally, remember there are as many ways to do this as there are people doing it. All the rules are the same, but the implementation is peculiar to the type of business and the individual presenting the data. Remember you start with the best AVAILABLE data and go forward. GOOD LUCK.

Bibliography

[RALS1]: This rule comes from Richard Ralston.

[PILC1]: This rule comes from John Pilch.

[SACK94]: *MVS Answer Book* p112
David J Sacks. John Wiley & Sons Inc. © 1994, ISBN 0 471-60821-1

Tutorial on Integrated Performance and Capacity Management Process
Robert G. Gutjar, CMG99

RMF Performance Management Guide, SC-28-1951-04, Chapter 1.