TEAMQUEST

IT OPTIMIZATION INSIGHTS

Getting Started:
A Manager's Guide to Implementing Capacity Management

White Paper

BY RON POTTER

TeamQuest specializes in IT Service Optimization

TEAMQUEST

Getting Started:
A Manager's Guide to Implementing Capacity Management

You have just been tasked by senior management to develop capacity management capabilities for your organization. You are very familiar with managing day-to-day IT infrastructure activities, however when it comes to Capacity Management, you are one of the many uninformed professionals.

Although Capacity Management has been in existence in one form or another for over 40 years, only a minority of shops have chosen to implement it. With constantly lowering prices, many IT shops have chosen to throw hardware at performance and capacity problems rather than spend the time to manage what they have. In the past, that approach has worked, however times and management challenges have changed and costs increased; demanding a different approach. Capacity Management is one solution.

This guide is designed to help those managers who are now looking to assemble a Capacity Management organization and put in place the processes and best practices to make it successful.

What is Capacity Management?

Capacity Management definitions can be found in many places on the Internet, including many IT Infrastructure Library (ITIL) sites. The definitions can be both wordy and confusing. We will leave it to those sources for the definitive explanation.

For a brief, high level definition, Capacity Management is a set of work processes associated with the provisioning and management of IT infrastructure resources, such as servers, printers and telecommunications devices, used to support business processes in a cost effective manner. Capacity Management work processes include monitoring, reporting, tuning, planning and predictive modeling.

Capacity Management has three distinct aspects:

Resource Capacity Management

Resource Capacity Management looks at the use and performance of individual infrastructure resources. These resources could be servers, telecommunication lines, printers, scanners, routers and PCs. This is the lowest level approach. The work only addresses each individual unit's performance and capacity positions. Although very effective in managing individual units, this approach has shortcomings because of its narrow view.

Adding capacity on one infrastructure unit can have unforeseen implications on others. For example, upgrading a telecommunications line can permit more transactions to be performed during a set period of time, potentially flooding a server with unexpected transaction volumes. This possibility can result in poor performing

transactions or even worse, cause the server to shut down unexpectedly due to overwork.

Service Capacity Management

Service Capacity Management takes Resource Capacity Management to the next level. Individual resource information is aggregated into pools as it relates to IT services. Printing, email, telephone, and online transaction processing are just a few examples of IT services. This approach looks at capacity positions across the IT infrastructure components supporting and sustaining IT services.

For example, an online transaction processing service could have server, telecommunications, desktop and printing components, just to name a few. Managing at a service view ensures that upgrading any one component does not adversely impact the others. The shortcoming of this approach is that although those in IT understand the work and its value to the organization, the business cannot relate to the technical terminology. As a result, there are usually communication disconnects between technical and business staff.

Business Capacity Management

This is the highest level approach to Capacity Management. Instead of aggregating usage and performance into IT services, we aggregate into business processes. This approach looks at capacity positions across the IT infrastructure components supporting and sustaining Business services. Sales order entry, claims adjudication, customer relationship management and general ledger are just a few examples of business processes.

For example, a Sales Order Entry system could have telephone, server, telecommunications, database, desktop and printing components, just to name a few. There is a lot of value in employing a business-oriented approach. Results are expressed in business terms so all understand. Conversations are transformed from ones of

expense to ones of business investment. Reaching this level should be any organizations long term goal.

Art or Science?

The arguments over whether Capacity Management is an art or science have been raging for years. Having been employed in a capacity management role for many years, I tend to think it is a combination of both. The science comes into play when collecting and analyzing data, and transforming it into information. Scientific methodologies ensure the integrity of the data and that analysis processes are consistent and repeatable.

Art comes into play when reviewing the results. Only experience and intuition can accurately judge whether a result is reasonable. Real life can stray from predictions and good judgment can detect those possibilities early, before plans are solidified and put into motion.

Managing Capacity Management

Contrary to popular belief, you don't need a doctorate in mathematics to manage a Capacity Management team. Understanding queuing theory is a plus, but also not required. You do need to understand the work that must be performed. Your challenges will be to develop credibility in your work and convince the business that their participation in the process is important to business success. You will need to understand business processes, IT processes and how they interact. However, your job is more involved with relationship building and communications throughout the organization.

You are the gatherer of business planning information to support capacity planning efforts. You are also the bearer of news resulting from the work – sometimes it's good and sometimes it's bad. That being said, in medium and large organizations you can have more financial impact on the organization than in any other IT department, plus your ability to relate IT services to Business processes will be invaluable to the organization.

The Work of Capacity Management

Good Capacity Management requires performance management and capacity planning together with business-aligned financial analytics. Those higher-level functions are enabled by necessary foundational tasks.

Collecting and Storing Data

There is an old business adage that says "If you can't measure it, you can't manage it." In order to manage IT infrastructure components, we need usage and performance data. Real-time data is good for the moment, but to be able to correlate it to other time periods, we need to store it for future analysis. Operating systems report the data in a variety of ways. You may choose to build collection agents or buy them. In either case, data can be gathered and stored. Server utilization, disk storage reads/writes, telecommunications delay times and database statistics are all examples of data you may choose to collect.

Monitoring

Monitoring permits you to see how a particular infrastructure component or application is performing at that moment in time. Monitoring can help you diagnose a customer complaint of poor performance or if used proactively, help you find performance problems before they negatively impact your customers. Monitoring can be accomplished through the issuance of certain operating system commands or through the use of software and hardware tools.

Reporting

Gathering data is not enough. For the work to be of value, you need to be able to communicate the results to management so more informed decisions can be made. To generate reports from your collected data, you can employ very simple methods using basic spreadsheet and text editing tools or professionally developed reporting tools with extensive graphics, automated report generation and report annotating capabilities. Reports should be generated at regular intervals.

Where large numbers of components exist, exception reporting should be considered to reduce the amounts of information senior leaders need to digest. For example, it is extremely unlikely that you can manage 10,000 servers by reviewing reporting on each one. However, if your reporting only produces information on those servers that deviate from capacity and performance plans, the numbers to review an substantially lower. In reviewing just the exceptions, you put your focus on the servers needing attention and take it away from those operating within expected limits.

Reports are also important because they market the value of your work. Remember that without reports, no one knows nor understands what you do and its value to the organization.

Analysis

Data by itself has no intrinsic value to the organization. To be of value, it needs to be analyzed, correlated and converted into information with which leaders can take actionable steps. Analysis is the process of turning our data into information. Telecommunications bandwidth, server memory, data channel busy and server utilization are just a few of the resource types that are frequently analyzed.

Within IT, there are two commonly used analysis types – reactive and proactive. Reactive analysis is used when there are issues in service delivery. The goal is to find the cause of the issue, determine ways to

mitigate it and do the work necessary to restore service to previous levels. Proactive analysis employs some of the same processes but its goal is to find potential or impending issues so they can be mitigated before impacting customers and business staff. Reactive analysis will always be with us but the more proactive work that is accomplished, the fewer reactive events will occur.

Performance Management

Better performing applications use fewer resources. Tuning applications to perform better improves service while reducing costs — a truly win-win situation. Software and hardware infrastructure components are delivered with configuration parameters directed for the "average" data center. If these are not tweaked to reflect your actual operating conditions, you may not get the most from your investment. The tuning process ensures the proper configuration parameters are in place.

In addition, application programmers are evaluated by the number of lines of code they generate and for supplying required functionality. In most cases, application performance is not a concern. This is not a criticism of application programmers. It is just the way most shops work. Rather than slow down the programmers, management feels it is easier for tuning experts to go back through applications after they are developed and if necessary, find ways to streamline them.

Capacity Planning

This is the planning function of Capacity Management. Capacity Planning takes business planning information, translates it into IT infrastructure resource needs and predicts what resources need to be put in place at a particular time to satisfy the business needs.

Capacity planning also ensures that the costs associated with the service are within the business' ability to pay. If they are not, capacity planning looks at alternatives that best fit the need. Reducing service levels could be an alternative where support costs are too expensive.

Capacity planners look at historical growth. They determine periods of time representative of normal business operations and analyze usage and performance information during those time periods. Using this data they develop a point-in-time "picture" of infrastructure. This is called a "Baseline." All growth projections, whether positive or negative, are applied to the Baseline to develop predictions of future infrastructure needs. For example, a call center application may be evaluated from 10 a.m. to Noon on Mondays and Tuesdays because those call volumes are representative of the levels that the business wants to provide good service.

Capacity planning work looks at all aspects of the IT infrastructure — addressing more than hardware and software. Capacity planners need to understand impacts to facilities, both IT and business. They also need to understand the capacity of the staff to perform the work to add or reduce infrastructure components. Without those considerations, projects can fail because there are not enough people to do the work or there is insufficient floor space to house the new equipment. Mistakes such as those can be costly in a fast-moving business environment.

Capacity planning time frames should address immediate needs and also look several years out. The immediate view should be aligned with the budget cycle. The business and senior IT leaders are already managing to the budget cycle so understanding of the implications of the planning work is more easily understood. A longer term view provides additional input to the decision making process. If long term direction varies from immediate needs, a different approach may be taken than if short and long term directions are similar.

Predictive Modeling/Simulation

Many organizations use trending to predict future IT infrastructure component performance. Trending involves graphing capacity positions for a period in the past, drawing a line that depicts the growth trend, then project the line into the future based upon business growth plans.

The problem with this approach is that computer systems do not perform in a linear manner. As a threshold capacity of a device is approached, requests for that resource start to queue (wait for service). In a very short time, transaction queuing causes performance to degrade at a rapid rate as depicted in the graph below.

In order to more accurately predict computer system performance, Analytical Modeling or Event Simulation techniques must be employed. Event Simulation techniques are the most accurate, but are also the most time consuming. Event Simulation does what the name implies; it uses computer software to simulate the running of all the work on the infrastructure components. In order to do this work, modelers need to identify each individual transaction, process and/or network packet and enter it into the model. Depending on the size of the model, it could take weeks or months to complete the work.

On the other hand, analytical modeling uses mathematical equations to predict future performance. It can be very accurate as well; results usually fall within 5% of actual conditions. Because the software has mathematical background, you can generate many more models in the same period of time. It is also much easier to run a variety of what-if scenarios. Most people who model employ analytical modeling software.

Financials

Capacity Management needs to be concerned with financials. You can have the best provisioning plan in the world but it is of little value if the business cannot afford to implement it. Money is the language of business so all works need to be done with that aspect in mind. There are several aspects of financials that you need to consider:

- **Budget**
  This is the financial process associated with planning expenditures and capital improvements over the planning period — usually one year and aligned with the organization's business cycle.

- **Total Cost of Ownership (TCO)**
  This is a methodology to determine the actual costs of implementing and maintaining a specific IT project over a specific period of time, usually five years. The process looks at all aspects of costs such as acquisition, installation, ongoing maintenance, facilities, organizational overhead, and staffing costs. With short- and long-term project costs better understood, management can make more informed decisions on whether to proceed with a project.
- **Chargeback**
  Business has already learned that something free is abused. Almost every organization has controls on office supplies and travel, making each manager responsible for what they use. IT resources are no different. Chargeback systems permit IT resource usage to be attributed to a particular user or cost center. By asking managers to plan and account for usage, more effective use of computer systems has been experienced.
- **Expressing the value of your work**
  In order to communicate the value of your team's work, you need to express results in financial terms. If you don't, management will not understand your value. For example, will business better understand the value of what you do if you say that you removed 200 servers from the infrastructure through consolidation or that you saved the organization $7 million per year by combining servers?

What Do I Need to Get Started?

Now that you have a high level of understanding of the work, we need the organizational resources to perform the work. As in any business endeavor, people, processes, data, and tools are needed to ensure success. The other unspoken needs are senior and business management buy-in and ongoing support. Since Capacity management has so many touch points across the entire organization, lack of management support will result in limited value to the organization or outright failure.

Management Support

Management support consists of more than just getting the blessing to proceed with the work and encouraging cooperation where there is none. Business strategy, planning information, and business process flows are also needed. Relationships and dialogues are needed to better understand business needs so that the right IT resources are in the right place when the business needs them. Building relationships also ensures you get involved early-on in a project, permitting you time to review and plan instead of reacting to a project at the last minute.

People

Staffing will depend upon the size of the IT organization. In smaller sites, a few generalists are in order. Larger sites should rely on specialization as more detailed work can be completed in shorter periods of time. Senior people are preferred because things in the

real world rarely act in the same manner as they do in the college laboratory.

I once hired a mathematician as a capacity planner. He had a great understanding of the mathematics behind the work and created wonderful statistical documents. The problem was that he didn't have any experience with the equipment or software so couldn't judge the reasonableness of his work. Remember that we previously mentioned that this work is part science, experience and intuition. You need all the parts to be successful.

For the most part, I have found that system programmers and system administrators make the best capacity planners, especially those with considerable operational experience. It doesn't seem to make a difference if they have a college degree. Most are intelligent and can quickly absorb the mathematics and processes. The biggest challenge will be to get them out of the details. Since capacity planners deal with averages, working at too fine a level of detail can bog down analyses and have little or no improvement on the precision of the work.

Training is an important part of keeping staff current and effective. Technology changes at a frightful rate and your staff needs to keep abreast of it. Newer technology may provide efficiencies that can substantially improve operations and lower costs. If you and your people do not continue to improve your knowledge, you cannot hope to keep your business ahead of its competitors.

Processes

TeamQuest ITSO and ITIL rely on processes to deliver meaningful, repeatable and reliable results. Processes ensure you come up with the same answer every time you analyze the same pieces of data. Meaningful, repeatable and reliable results generate credibility; something capacity management needs to survive.

Processes describe the work that needs to be accomplished. They identify the inputs and outputs; information coming into the process

step that needs to be acted upon and the deliverables as a result of the work. Processes identify the recipients of the results of the work performed. Well written processes ensure the people executing them completely understand their roles and responsibilities. They also permit people to understand the importance of their work and how their work supports business success.

Sample ITIL processes can be obtained in ITIL books. Sample ITSO processes can be obtained from TeamQuest. Many consultancies also can provide samples. You may be able to find sample processes on the Internet. In all cases, be prepared to tailor them to your needs. Few organizations exactly match the sample scenarios.

Data

The type and amount of data you need to collect depends upon the mission of your capacity management organization. Commonly collected and stored data includes performance-related, transactional and usage. If you only do servers, you will collect data such as:

- Processor busy
- Disk read/writes
- Memory statistics
- Job/process execution statistics
- Database related statistics

Networking capacity planners commonly look at bandwidths, transmission delay, router, firewall and packet statistics. The types of data you will collect will depend on your capacity management goals. In most organizations, too much data is collected initially; then reduced and refined as the capacity management organization matures.

Data should be captured in the same intervals across all infrastructure components for which you are responsible. It is difficult to analyze the interactions between a transaction server and a database server if one is collecting at one-minute intervals and the other at 5-minute

intervals. Spikes seen at one-minute intervals will be unnoticeable on the 5-minute intervals, making work correlation difficult if not impossible.

All infrastructure devices should be synchronized to the same clock; usually GMT/Zulu or a government sponsored atomic clock such as the one at the United States National Institute of Standards and Technology (NIST) or National Physical Laboratory in the UK. We do this to ensure log and collection times are consistent across the organization. It is difficult to perform an end-to-end transactional analysis if each of the components employed by that transaction have different times, especially where large, fast computers are involved that can process many millions of transactions each second.

Archiving of data is very dependent on the organization. Retention times depend on the amount of change in the environment and business cycles. For example, if applications and business processes change frequently, storing data for long periods of time may be counter-productive as last year's data will have little or no relevance to current operations. Business cycles also determine retention periods. A business cycle with one busy period that lasts one or two months per year may wish to keep 3-4 years of data. Another organization with monthly peaks may choose to save only 13-18 months of data.

Tools

Tools permit you to trade cost for productivity. Yes, you can do the work with spreadsheets. It will be difficult and your people will spend more time gathering and processing data than they will be analyzing it. Modern tools automatically collect, organize and archive data plus provide analysis and modeling capabilities. They free up your staff's time from mundane, repetitive tasks to permit more important analysis, planning and prediction work. I can't tell you the right answer for your organization. You know your cost and business drivers better than me. However, I have formed a number of capacity management teams over my career and my best success has come from using

tools to manage the data and free my people's time to do the analysis necessary to improve service and reduce costs.

Common tasks performed by capacity management tools are data collection and archival, analysis, modeling, reporting. Data collection and archival relieves a lot of time from staff as there is initial setup and then besides a few tweaks now and then, it basically runs itself. Analysis tools permit you to correlate events and drill down into details to find causes of performance problems or to identify performance improvement opportunities. Modeling tools permit you to perform very accurate what-if scenarios to predict future requirements; permitting you to more effectively use and manage IT resources. Reporting tools can help you set up automatically generated reports. You develop the report once, set up a publishing schedule and it does the work. You only get involved with exceptions or changes.

Easing Into It

Now that we know the work that needs to be done and what is needed to do it, where do we start? Again, each organization has its own set of priorities and goals so where to start is dependent on them. Here are a few areas where successful organizations have started..

- **Mission critical resources.**
  These transactions and jobs are critical to business success so any improvements in service or reduction in costs gets immediate attention from senior IT and business leaders. Starting with these can more quickly develop credibility in your work; making it much easier to gain cooperation in other areas.

- **Low hanging fruit.**
  Your technical staff probably already knows which applications and services are poor-performing. In many cases, they already know the solution; just have not been able to garner support for corrective activities. Choosing this place to start permits you to more quickly address performance problems. Being able to quickly resolve some nagging problems with the business will go a long way in developing credibility in your organization.
- **More frequently used transactions or jobs.**
  Savings from small, frequently used transactions can have bigger impact than a single large saving on a seldom used job. I have seen simple performance tweaking of a heavily used transaction drop a large server's utilization from 32 processors to 12. We were able to satisfy several years' growth on that server plus provision several new applications, saving considerable hardware and software expense.

What's Next?

Keeping focus on the start-up tasks – learning the basics, getting your processes in place and maturing them – is important; however you need to keep an eye to the future. Once your capacity management organization starts to stabilize, there are several things to consider to take it to the next level..

I can't stress enough about the importance of building better business relationships. Better relationships means they come to you in the planning stages of a new project. It means they come to you to alert you to the new marketing campaign which, if successful, could over-run your servers. Relationships also mean that your business partners

understand the value of your work to the success of the business, making it much easier to sell the needed work.

Another method to reduce IT infrastructure usage is to influence behavior through chargeback. You probably already have the information needed to identify which business units consume which IT resources. Using a chargeback systems holds business units accountable for their use of IT services. Just as what happened with office supplies and travel expenses, assigning departmental accountability has in most cases resulted in reductions in usage.

Finally, let's talk a little about Sales and Marketing. You are doing a great job of improving performance and reducing costs but if you don't tell anyone, all that great work may go for naught. To paraphrase a Forrester analyst, "Shout your successes to the rafters." In addition, start to put yourself in position to change conversations from that of expense to that of business investment. Showing graphs of server growth or Total Cost of Ownership spreadsheets induce visions of cost. If you can express results in terms of IT cost per increment of business work, such as an order or workflow item, the business can see the value to the investments, especially if the unit costs decline as a result of economies of scale.

In Closing

Good luck with your work to establish a capacity management organization. I think you will find it as rewarding as I have. I hope this guide has helped you understand the work from a high level and reduced the uncertainty of beginning.

ABOUT THE AUTHOR

Ron Potter is the Best Practices manager for TeamQuest Corporation. Ron's background includes more than 20 years in the IT industry, spearheading a successful ITIL implementation with a Fortune 500 insurance company, and discussing ITIL topics as a presenter at several conferences and trade shows.

TEAMQUEST

connect with TeamQuest: itsoblog

WORLDWIDE HEADQUARTERS

UNITED STATES
TeamQuest Corporation
One TeamQuest Way
Clear Lake, Iowa USA 50428

OTHER LOCATIONS

SWEDEN
GERMANY
UNITED KINGDOM
MEXICO
HONG KONG

With resellers in many additional countries.

CONTACT US

info@teamquest.com
teamquest.com/about-us/contact-us/