

Economic Challenge and Capacity Management

Increasing Value and ROI from Capacity Management - A Maturity Model

David Wagner
Solution Labs Inc

January 29, 2009



Contents

Overview	3
The Business Need and Challenge	3
Service Capacity Risk	4
The Challenge of Investment.....	4
The Challenge of Change.....	4
Lifecycle Capacity Management.....	5
Capacity Management Maturity	6
Business Dimension - Technology, Service or Business?	6
Technical Dimension - Easy, Feasible Guesswork versus Time consuming, difficult Accuracy?	6
Total Capacity Management ROI and Value	7
The Performance Analysis and Capacity Management Maturity Value Spectrum.....	7
Level 1. Simple Metrics, Manual Reporting - Low ROI on investment in Capacity Management	8
Description	8
Level 2. Simple Metrics, Automated Reporting - Low/Mid ROI on investment in Capacity Management.....	9
Description	9
Level 3. Application/Service Oriented Metrics, Manual Reporting and Analysis - Mid ROI on investment in Capacity Management.....	9
Description	9
Level 4. Application/Service Oriented Metrics, Automated Reporting and Analysis - Mid/High ROI on investment in Capacity Management.....	10
Description	10
Level 5. Application/Service Oriented Metrics, Automated Reporting and Analysis, Predictive Modeling - Highest ROI on investment in Capacity Management.....	11
Description	11
Accelerating Value and ROI from Capacity Management.....	12
Tool Considerations	12
Conclusion	12
About The Author	13
About Solution Labs Inc.	14

Overview

During times of economic stress, pressures on IT can become seemingly insurmountable. Requirements to "ensure business services" and "optimize availability" - challenging enough when budgets were more open, become extremely challenging. Spending freezes or significant reductions in capital infrastructure, are common. However, just because your hardware or software acquisition budgets have been shrunk or frozen, doesn't mean the requirement to ensure business service uptime, availability and performance lessens.

These challenges make your ability to business justify Capacity Management more critical than ever. Whether you have formalized processes - and the tools to accompany them, or you simply recognize the need to "get started" with Capacity Management, it is critical that you be able to optimize the delivery of Capacity Management within your organization.

To do this you need to know where you currently stand in the maturity of your processes and deliverables. You need to elevate your process maturity. You need to optimize the value you deliver. You need to leverage existing staff resources. You need to expand your organizational scope and reach.

Finally, and most critically, you can't get started unless you are able to justify the benefits of optimized Capacity Management to the broader organization.

This whitepaper is intended to assist you in understanding the key dimensions to Capacity Management as they affect the business and as they are affected by technical choices and tradeoffs. This paper suggests a framework by which to conceptualize the business value and return on investment in Capacity Management.

Optimized Capacity Management means:

- Doing Performance Analysis and Capacity Management more quickly,
 - Delivering as high a business value as possible, to as many cross organizational stakeholders as possible by aligning to the business instead of technology,
 - Delivering higher levels of intelligence to more safely increase overall utilization levels,
 - Applying the discipline of Performance Analysis and Capacity Management across all, or as much as possible, of IT infrastructure to minimize service risks,
- and finally,
- Doing Performance Analysis and Capacity Management without requiring increases in staff resource levels

The Business Need and Challenge

Save Money without increasing Risks!

Too bad you can't just press an "Magic Button"! Capacity Management is all about the fine balance between the quantity and type of IT resources that are required to sustain acceptable business service throughput and responsiveness - and the costs associated with acquiring,

deploying and managing them. Too much capacity, and financial resources are wasted. Too little capacity, and business services become unavailable.

In a sense, Capacity Management is really an IT insurance policy, and must be evaluated accordingly. Whereas life insurance is about how to not have death negatively impact the living (as much as possible), Capacity Management is about how to not have IT resource capacity negatively impact Services.

We call this new concept in Capacity Management: Service Capacity Risk.

Service Capacity Risk

Service Capacity Risk is the exposure of having insufficient hardware resource capacity (CPU, I/O, Memory, etc.) to sustain required service levels (response time, transaction throughput, batch processing, etc.) or to meet regulatory and/or compliance mandates and guidelines.

New technologies such as dynamic and automated resource provisioning and reconfiguration - whether physical, virtual, or both - offer the opportunity to rapidly respond to changing requirements for resource capacity needs, with the *potential* to significantly reduce Service Capacity Risk. However, automatically deploying resources without understanding what is needed, when, and most importantly why, simply means you are making potentially bad decisions much more rapidly! You are simply wasting money, or putting business services at risk much more quickly.

The only way Service Capacity Risk can safely be understood and fully managed is through life-cycle Capacity Management processes and deliverables. This means understanding what resources have been used (versus business demand), what resource are currently being used (for current business demand), and what resources will be needed and under what circumstances (for projected, potential business demand). This must be done continuously against all resources associated with each/every important business service.

Quite simply, you must have "intelligence behind automation".

There are two high-level challenges associated with Service Capacity Risk. If you always had the ability to purchase unlimited resources, you could minimize Service Capacity Risk. But, you wouldn't be very efficient. And, if resource configurations, business demands, application environments and the like never changed, you could minimize Service Capacity Risks. But, especially during times of economic challenge, neither of these luxuries hold true.

The Challenge of Investment

When fiscal resources are tight, capital spending on incremental resource capacity can become difficult or impossible. Freezing hardware resource acquisition has the potential to save significant capital expenses (and much directly associated operational expense such as licenses, staff costs, power and cooling, etc.), but is only a viable strategy if organizations can be 100% certain they are not putting themselves at increased "Service Capacity Risk".

Without Capacity Management, this can only be done if there is no change in the business or IT environment, to do otherwise is simply guessing.

The Challenge of Change

Every time there is a change in the overall "System" (of applications, business demand, and/or underlying technical infrastructure configurations used to support business services and applications), there is a potential impact on the ability of the system to deliver valuable business results in a timely and productive fashion.

In the name of increasing levels of cost and operational efficiency, most organizations are now implementing consolidation (physical, geographic and/or logical) and/or virtualization projects to increase overall technical infrastructure utilization levels, and thereby reduce their per-business-transaction costs associated with IT.

Whether or not these projects include the latest in automated, dynamically provisioned IT resources, they significantly increase the magnitude and rate of change underlying the "system". As such, they intrinsically increase Service Capacity Risk.

Mergers, acquisitions, divestitures, and other externally imposed organizational changes often result in significant changes to business processes as well as IT infrastructures.

Finally, ongoing business pressures to roll-out new revenue producing applications and services results in additional change; change that accrues whether or not there is increase in business demand from existing services.

Lifecycle Capacity Management

The only way to address external business challenges of investment and change, and simultaneously manage Service Capacity Risk is through the discipline of Lifecycle Capacity Management.

Lifecycle Capacity Management can be summarized as a continuous (improvement) process spanning historical, current, and projected/planned future behaviors of business services and the IT resources that support them. This lifecycle process has the following fundamental steps:

- **Metric measurement:** measurement and recording of IT resource utilization rates as they are demanded and consumed
- **Metric analysis:** evaluation of metrics to determine if/when problems might occur and to understand typical and atypical behaviors and requirements
- **Planning:** evaluating potential future behaviors and requirements
- **Continuous reporting:** dissemination of information gleaned from measurement, analysis and planning activities to any/all relevant stakeholders

Implementation of some, or none, of these phases of the Capacity Management Lifecycle impacts the overall organizational maturity in the discipline, and is an obvious critical factor in the total Capacity Management value realized by the entire organization. Obviously, if an organization is only measuring performance, with limited to no analysis, reporting or planning, they cannot hope to yield much, if any value from Capacity Management; they aren't fully doing it.

Importantly, each of these stages in the Capacity Management Lifecycle may also be conducted at varying levels of maturity. It is at this level where we focus this white paper.

Capacity Management Maturity

Maturity of Capacity Management processes and deliverables, at each and every stage of the Capacity Management Lifecycle has two key dimensions:

- Tradeoffs associated with Business alignment (how closely does it align?), and
- Tradeoffs associated with Accuracy (simple and fast versus complex and time consuming)

As an example, organizations may simply measure a few "raw" performance metrics which have no direct relationship to the business, for example "CPU Utilization" versus another measuring business transaction throughputs and financials associated with those transactions.

Business Dimension - Technology, Service or Business?

A key factor of overall maturity is not simply implementation of each and every phase of the lifecycle process. Also critical is the degree to which the activities and steps within each phase of the lifecycle correlate to business services or overall business needs versus simply "technology specific" approaches.

In ITIL V3 this spectrum is denoted as "Component", "Service" and "Business" Capacity Management (from technology/resource-centric up through the highest business-abstracted level respectively). It is significant to note that earlier versions of ITIL (Version 2 and earlier) referred to "Component" as "Resource" which in some ways was more specific and germane to the typical capacity management approaches and maturities of virtually every IT organization today. In that sense, the ITIL V3 redefinition of "Resource" to "Component" can be interpreted as denoting the substantially increased importance of focusing Capacity Management on far more than simply IT physical/virtual infrastructural resources.

Investment decisions can no longer be made on the basis of solely technical metrics and information. Business owners neither ask nor care about the answers to questions like "How many I/O's per second can our web-banking application sustain". Rather, they typically ask: "How much capacity do we need to run web-banking profitably, what will it cost us, and how can we optimize it?"

During "boom times" of investment, business stakeholders might "grin and bear it", accepting technical metrics to make investment decisions. During times of significant economic pressure, any previous inclination on their part to give in and make decisions solely based on technical metrics and information usually disappears.

Therefore, the criticality of this dimension is significantly higher in 2009 than in years past.

Technical Dimension - Easy, Feasible Guesswork versus Time consuming, difficult Accuracy?

There is a technical dimension relating to "simplicity versus accuracy and precision" which associates to the process of tracking resource utilization against business services. At one end of the this spectrum, it can be easily seen that it is quite simple to simply take whatever Resource (Component) metrics one might already have and do some simplistic reporting against them (over time), and thus fulfill the steps of lifecycle Capacity Management. While this is typically quite simple and feasible (nothing much additional to implement or purchase after all), it is demonstrably inaccurate, and results in, at best, grossly inefficient IT resource utilization rates,

and, at worst, significant Service Capacity outages. And, such approaches quite obviously fail on the Business Dimension as well.

At the other end of this spectrum, a process can be implemented whereby any and all resource deployment decisions are fully analyzed using complex business and technology analytic technologies and processes; factoring all aspects of business and technology change. Metrics are gathered which more closely align to Services and the Business. Processes are put in place to ensure all variables are compared, analyzed, and evaluated across both technical and business dimensions.

Precisely because of today's highly dynamic and virtualized technology environment and the typical scale involved, it is equally demonstrable that such overly complex and highly specific Capacity Management approaches are simply impossible to implement on any but a small fraction of the total underlying IT infrastructure; at least without massive investment in scarce and expensive technical talent. Investments in incremental staffing that are typically impossible to justify.

Most organizations today are implementing server consolidation and virtualization projects as a way to become more operationally efficient. But, if they do not do Capacity Management on all relevant infrastructure, they are right back to Service Capacity Risk. But now this Service Capacity Risk is significantly amplified to the degree to which they have conducted (or will conduct) server consolidation and virtualization projects, since they now have more Business Services sharing underlying - and ultimately capacity limited - resources!

After all, if you have more "eggs" and fewer "baskets", the importance of optimizing the performance of those "baskets" increases substantially.

This conundrum suggests a solution of automation. Automation to achieve a balance of "accurate enough", everywhere, and sufficiently aligned to Business Services. To find that balance point and optimize the value received from Capacity management, one must evaluate ones starting point, and identify the benefits that accrue from each maturity step in both Business and Technical dimensions.

Total Capacity Management ROI and Value

It should be no surprise that these progressions and tradeoffs directly correlate (in increasing fashion) to the total value and hard return on investment (ROI) organizations may achieve from any investment in Lifecycle Capacity Management solutions and processes.

- The closer Capacity Management processes and deliverables align to the business, the higher the value and ROI.
- The more automated, accurate, timely, and intelligent you can make the technology behind your Capacity Management processes and deliverables, the higher the value and ROI.

The Performance Analysis and Capacity Management Maturity Value Spectrum

Based on over 20 years of experience working with organizations worldwide in IT Performance Analysis and Capacity Planning, we have gained a deep understanding of the typical "states" of IT maturity in the Capacity Management value spectrum.

The following spectrum is ordered from lowest ROI and business value to highest, and describes some of the general characteristics of performance analysis and capacity management processes and deliverables at each level.

Level 1. Simple Metrics, Manual Reporting - Low ROI on investment in Capacity Management

Description

Processes and technologies are in place to provide relatively simplistic tracking and the manual reporting of "raw performance metrics" at a server (or virtual partition/guest) level. E.g. CPU is X% over time, I/O is Y% over time. Manual generation of weekly or monthly reports based on such "raw" metrics. Typically, for any environment larger than approximately 100 Servers (or Virtual Servers), not all servers and applications are encapsulated in the process.

Advantages

- No capital investment in capacity management tools
- Technologically and organizationally simple
- Manual generation of reports is easiest with this approach (e.g. create reports in Word, charts in Excel, using "cut and paste" from performance toolsets)

Many accomplish this by writing scripts or code to extract whatever historical metrics might be available from any real-time event management or monitoring solutions may already be in place.)

Disadvantages

- Time consuming and labor intensive
- Inefficient and error prone
- Minimal sets of metrics
- High ongoing costs of "ownership" due to custom nature of implementations (if any have been done programmatically)
- Difficult (usually staff cost prohibitive) to cover ALL the infrastructure, especially in larger environments
- Highest levels of Service Risk lead to very low resource capacity efficiency

Because metrics are usually not sufficiently detailed, or are server or partition level only, it is dangerous to run average utilization rates much above around 15% due to lack of understanding of real requirements of business services in terms of resources. (Long-term customer experience has shown that organizations at this level of maturity seldom run above 10-15% average utilization rates. Those that exceed this rate, have found a strong correlation with unacceptable Service Capacity Risk). Organizations at this level are realizing minimal or no benefit or value from this type of Capacity Management process.

Level 2. Simple Metrics, Automated Reporting - Low/Mid ROI on investment in Capacity Management

Description

Processes and technologies are in place to provide comprehensive tracking and fully automated reporting of "raw performance metrics" at server or virtual partition level. Automated generation of daily, weekly, and monthly reports based on metrics is being done across most, or better all, of the infrastructure.

Advantages

- Simple
- Eliminates human error
- Automated reporting frees staff time for more advanced (higher ROI) capacity management activities
- Automation allows for ALL resources to be covered (reducing Service Capacity Risks)

More frequent, automated reports that span the entire infrastructure offer potential to spot trends and problems sooner, meaning that underlying resources can safely be more heavily utilized (20-25%) without undue Capacity Service Risk. Efficiencies are increased and value and ROI from Capacity Management is increased significantly (40-50%) over Level 1, with the same, or slightly better, level of Service Capacity Risk.

Disadvantages

- Capital investment cost of Capacity Management tools
- Minimal sets of metrics

Because metrics are still server or partition level as well as being insufficiently detailed, it remains dangerous to run utilization rates much above around 25% due to lack of understanding of real requirements of business services in terms of resources.

Level 3. Application/Service Oriented Metrics, Manual Reporting and Analysis - Mid ROI on investment in Capacity Management

Description

Processes and technologies are in place for tracking, analyzing and reporting on metrics which align to key, server oriented aspects of business services (e.g. so called "application workloads"). Manual generation of reports on weekly, or monthly basis.

Advantages

- Better understanding of Business Service requirements for resources over time (for that portion of the infrastructure under management)
- 30% Higher average levels of average resource utilization than Level 2

Closer alignment of resource utilization to business service "consumers" has the potential to more safely allow higher levels of underlying average resource utilization (typically 30-40%) without increased Service Capacity Risks. This is primarily because metrics are more detailed, and more accurately reflect actual application and business service demands over time, increasing visibility and reducing risks.

Disadvantages

- Capital investment cost of Capacity Management tools
- Complex, time-consuming and labor intensive
- Error-prone
- Impossible to achieve overall, or even much coverage of the infrastructure
- Limited foresight "distance" due to time required to generate manual reports and analysis

This level, while being more closely aligned on the Business dimension, suffers a similar lack of automation disadvantages as with level 1, making it impossible to cover all infrastructure and all resources without unrealistically high staffing levels.

Level 4. Application/Service Oriented Metrics, Automated Reporting and Analysis - Mid/High ROI on investment in Capacity Management

Description

Processes and technologies are in place for fully automated tracking, analysis and reporting on metrics that closely align to business services. In addition to application workload data, additional performance and capacity metrics are factored which more closely relate to business or service metrics (e.g. configuration, costing, response times, transaction rates, etc.) Automated generation of exception oriented analyses and reports is done on a daily, weekly, monthly basis.

Advantages

- 300-400% better total ROI and value than Level 1
- Elimination of human error
- Better understanding of Business Service requirements for resources over time
- Automation allows extending across most/all infrastructure
- Reduced staffing requirements
- Time saved via automation may be applied to increased Predictive Modeling efforts

This is the "sweet spot" or point of optimum balance for most organizations worldwide.

Closer alignment and better analysis of resource utilization to business service "consumers" can allow for much safer, and significantly higher levels of underlying resource utilization (40-50% or more depending on degree to which automated resource deployment solutions are leveraged) without increased Service Capacity Risks. Automation allows extending higher

resource utilizations across ALL infrastructure, without increased staffing requirements, further increasing ROI and value.

Disadvantages

- Requires commitment to Capacity Management as a key discipline within broader organizational change-management processes and controls.

Level 5. Application/Service Oriented Metrics, Automated Reporting and Analysis, Predictive Modeling - Highest ROI on investment in Capacity Management

Description

Processes and technology are in place for fully automated tracking, analysis and reporting on metrics that align closely to business services. In addition to application workload data, additional performance and capacity metrics are factored which more closely relate to business or service metrics (e.g. configuration, costing, response times, and/or transaction rates, etc.) Automated generation of exception oriented analysis and reports is done on a daily, weekly, monthly basis. Predictive analytics are used on appropriate and key elements of the infrastructure to completely understand relationship between the response time and throughput of business services and their underlying requirements for physical resource.

Advantages

- Elimination of human error
- Best possible understanding of Business Service requirements for resources over time (for that infrastructure being modeled on a regular basis)
- Lowest possible Service Capacity Risk for targeted Services

Closer alignment and best possible analysis of resource utilization to business service "consumers" over time and scenarios allows for safer, and highest possible levels of underlying resource utilization (60%+) without increased Service Capacity Risks. Automation allows extending higher resource utilizations across ALL infrastructure, without increased staffing requirements, further increasing ROI.

Disadvantages

- Potential to be staff intensive and/or expensive
- Cannot cover all the infrastructure with modeling
- Requires commitment to Capacity Management as a key discipline within broader organizational change-management processes and controls.

It is not possible to do predictive analytics against most or all infrastructure resources and services due to need for scarce staff time/resources, coupled with the rapid and dynamic rates of configuration change. But with appropriate and complete coverage with Level 4 processes, it is feasible to stay sufficiently ahead of potential issues such that modeling can be applied precisely where and when it is needed.

Accelerating Value and ROI from Capacity Management

As has been shown, significant incremental value and ROI from Capacity Management begins accruing as organizations increase their level within the Capacity Management Maturity Value Spectrum. Therefore the over-arching goal should be how to increase that maturity as rapidly and cost-effectively as possible.

Tool Considerations

When considering dimensions such as speed and cost for implementation of any set of processes and deliverables, the only practical approach is to implement management tool(s) that can deliver significant capabilities across the discipline lifecycle, as well as across the Technology - Business dimension.

When considering the Capacity Management Lifecycle (Measure, Analyze, Plan and Report), we suggest the following categories as being particularly valuable at whatever Capacity Management maturity level is present or desired:

- **Automation** - the ability to fully automate all the analysis and reporting processes and deliverables associated with Capacity Management lifecycle in order to reduce errors, standardize processes, and scale to the entire business
- **Embed Expertise** - the ability to embed staff skills and expertise so existing staff skills can be cost-effectively leveraged against more infrastructure, applications and business services
- **Built-in intelligence and experience** - so less time is required from scarce, expensive existing staff experts, freeing their valuable time for other/more important activities
- **Flexibility and adaptability** - so any/all critical metrics required to align to the Services and Business may be accomplished; from multiple disparate sources
- **Scalability** - so the discipline of Capacity Management may be applied at the highest possible maturity levels against ALL the infrastructure, minimizing Service Capacity Risk
- **Rapid installation and non-invasive footprint** - to speed and maximize value and ROI and ensure minimal ongoing operating costs

The application of tool(s) encompassing one (or preferably more) of these capability areas, against whatever current level of Capacity Management Maturity is present in the current organization cannot help but provide immediate value and ROI. The more capabilities that are provided, the larger and faster the ROI and Value realized.

Conclusion

Today's IT environments are facing unprecedented pressures. Massive change is being imposed by consolidation and virtualization projects. Capital resource budgets have never before seen the scrutiny they are now facing. Customers and competitive pressures continue to mandate ever-increasing increases in business service capability, scope and number.

Above all, regulatory scrutiny of organizational preparedness continues to force ever more standardization of processes, deliverables and decision making in the name of compliance and risk management.

With ever more business services being supported by more powerful, and virtualized/consolidated resource environments, never before have so many factors been aligned in a perfect storm targeted squarely at Service Capacity Risk.

It should therefore be the goal of every IT organization to ascertain at what level of Capacity Management maturity their current processes align, and more importantly, how they can best and most quickly optimize them to increase the value and ROI from their discipline of Capacity Management.



About The Author

At Solution Labs, David Wagner is extending his career-long dream of making technology as efficient as possible by helping customers optimize their performance and capacity. Dave uses his over 27 years of technology experience and background in performance analysis and optimization to lead Solution Labs Sales and Marketing efforts, and is responsible for growing their worldwide base of successful and satisfied customers.

In his 11 years at BMC Software, Dave ran Solutions Marketing for areas including Data Center Automation (now Service Automation) and BMC Virtualization Solutions, delivering thought leadership via articles, interviews and a blog on capacity and virtualization. Dave ran Product Management and Marketing for the BMC Performance Assurance product line – responsible for growing the distributed solutions from \$2M/year in sales in 1996, to an installed base representing over \$500M in 2005, with over 1500 customers worldwide. With many years of industry participation focused on the Performance Analysis and Capacity Planning market, Dave has presented papers and participated in panels and forums such as Network World, Computer Measurement Group, as well as multiple and diverse vendor User Groups and communities.

Previous efforts included leading Data General Enterprise Management efforts, responsible for ideation and delivery of the first graphical PC-based management console for Unix Servers and RAID storage arrays (progenitor to the EMC NaviSphere solution). Earlier, at Stratus Computer, he was a founding member and Secretary of the Transaction Processing Performance Council (TPC) an industry standards organization dedicated to quantifying performance efficiency. In the early 1980's, Dave spent 7 years at Digital Equipment Corporation, where in addition to be an OS Engineer, he created and delivered a series of advanced performance analysis and capacity optimization courses to Digital's largest customers.

About Solution Labs Inc.

Solution Labs is uniquely positioned and qualified to deliver world-class Server Performance Analysis and Capacity Planning solutions that meet today's challenges. Founded by experts in this area, Solution Labs is managed and staffed by highly skilled Performance Analysts and Capacity Planners for the benefit of Performance Analysts and Capacity Planners. Solution Labs staff has over 50 years of combined expertise and experience using BMC Performance Assurance solutions to solve customer Capacity Management challenges.

The Solution Labs Inc. mission is to create and deliver the most comprehensive and easy to use performance analysis solution for today's complex Capacity Management challenges. Our singular focus is on delivering our customers a solution that enables them to

“Accelerate Resource Planning Intelligence”

Solution Labs continually strives to be a highly strategic partner towards our customer's success.

- Our products' comprehensive automation, embedded intelligence, and highly adaptable architecture are the *enablers* of that success
- Our expertise and continual customer interactions are the *guarantors* of that success

For more information, please contact us: sales@solution-labs.com, or 1 800 754 7149, or feel free to visit us at www.solution-labs.com.