

ITIL[®] Capacity Management -More than charts over coffee



A White Paper Metron Technology Ltd Metron-Athene Inc © 2006

Many organizations are embracing ITIL[®] as a model for best practices. This paper provides a brief background of ITIL and the ITSM processes, with particular attention paid to Capacity Management. The author also considers two different organizations and discusses Capacity Management as practiced in those environments, comparing and contrasting those practices with ITIL Capacity Management. Finally, recommendations are made for implementing ITIL Capacity Management in any environment, with additional focus given to the interfaces between ITIL Capacity Management and the other ITSM processes.

> By Rich Fronheiser Metron-Athene, Inc. rich.fronheiser@metron-athene.com

1. Introduction

This paper will, in general, focus on the IT Infrastructure Library (ITIL[®]) and how it has evolved as a model for IT best practices. Specifically, it will focus on the Capacity Management process – and how the role of the capacity manager has evolved (and must continue to evolve) in organizations aligning their processes with ITIL.

Section 2 of the paper will provide the reader unfamiliar with ITIL a reasonable background of ITIL and, specifically, IT Service Management (ITSM). Why has ITIL emerged as a model for best practices?

The Capacity Management processes as described within ITIL will be covered in Section 3. What are the goals and the levels (or sub-processes) of ITIL Capacity Management?

Section 4 will look specifically at two disparate organizations in the past decade and how Capacity Management has evolved in that decade under the alignment to ITIL and the other IT Service Management processes. Part of Section 4 will discuss the day-to-day activities of Capacity Management and how those were carried out by one of the organizations.

The remaining sections of the paper will discuss a detailed methodology for implementing ITIL Capacity Management and will also look at the interfaces between ITIL Capacity Management and the other ITSM processes.

1





2. ITIL Service Management

The IT Infrastructure Library (ITIL) was developed in the late 1980s and has become the worldwide standard of best practices in Service Management. Originally developed as a guide for the UK government, the framework has been found to be useful for all organizations, regardless of the line of business or the size of the organization.

Starting in the late 1980s, the UK's Office of Government Commerce (OGC) collected information about how different organizations tackled Service Management, analyzed this data, filtered out what would be most useful to the OGC and to its customers in the UK's government. Very soon after, other businesses outside of government found that this information and guidance was applicable to their businesses as well.

The IT Infrastructure Library is a series of seven books, three of which directly cover Service Management. Service Management, itself, is broken into two books – Service Support and Service Delivery. There is also a book on Planning to Implement Service Management. The other four books cover the business perspective, infrastructure management, applications management, and security management. For the purposes of this paper, we will focus on two books – Service Support and Service Delivery, which form the foundation for IT Service Management.





Since ITIL is design to be a descriptive, not proscriptive, framework, it very loosely (by design) describes models of organizing Service Management. The models show a best practices view of goals, activities, inputs, outputs, costs, and benefits of the processes. Most important to remember, ITIL does not (and does not attempt) to proscribe specific day-to-day activities, because these will vary between organizations. Instead, ITIL focuses on general best practices which can be used in very different ways depending on the business's need.

Therefore, companies will not abandon their existing activities and methodologies when implementing (rather, aligning towards) ITIL. Rather, ITIL provides a framework in which an organization can place its existing day-to-day activities and provide structure around those activities. More importantly, ITIL will help organizations identify areas for improvement and processes that require more attention and will help those organizations improve their IT practices.

IT Service Management is broken into two books within the IT Infrastructure Library. These two books detail the ten processes and one function (and the interfaces between them) that comprise IT Service Management.



Figure 2. ITIL Overview

Service Support (also known as the "Blue Book" in ITIL circles) consists of five processes (and one function) dedicated to operational, or day-to-day, tasks that ensure the user has access to appropriate services to support business requirements. These include:

- (1) The Service Desk function. More than a help desk, the Service Desk is a onestop-shop for all service requests as well as all Incidents. The Service Desk is of utmost importance, because the Service Desk is the face of IT to the customer.
- (2) Incident Management, which is concerned with getting the user/customer back to business as soon as possible in accordance with the Service Level Agreements. Note that Incident Management will not necessarily accomplish this by "solving" an issue – the goal of Incident Management is to return service to the customer.
- (3) Problem Management, which is concerned primarily with finding the root cause of problems. A problem, in ITIL terms, can be repeated (or trended) incidents or a major incident and the goal of Problem Management is to find the root cause of the problem and determine an appropriate cause of action. Note the difference between Incident Management, which is mainly concerned with getting the users back up and running, and Problem Management, which focuses on the problem and its resolution.
- (4) Change Management, which is concerned with the change process – assessing, approving, and interfacing with other Service Management processes to ensure the impact of changes to the business are minimized.
- (5) Release Management, which is concerned with maintaining hardware and software release levels, maintaining a Definitive Software Library, and making sure rollouts are secure and managed properly, and
- (6) Configuration Management, which is an integral part of all Service Management processes, since current and accurate information about all configuration items and their relationships to one another is crucial for making decisions and assess impact of Incidents, Problems, and Changes.

Service Delivery (also known as the "Red Book") consists of five processes dedicated to tactical, or strategic, tasks that look at what the business requires of the IT provider to allow for adequate support to the customers. Service Delivery is concerned with providing this support both now and in the future. These processes are:

(1) Service Level Management, which is responsible for making sure Service Level Agreements (SLAs) and underpinning Operational Level Agreements (OLAs) or contracts are met. Those that perform this process will work closely with customers to ensure well-defined SLAs are negotiated.



3



- (2) Financial Management for IT Services. This process is responsible for budgeting, accounting, and (optionally) charging for IT services.
- (3) IT Service Continuity Management, which includes the traditional "disaster recovery" function, yet is also concerned with keeping the business running through resiliency, redundancy, and effective planning.
- (4) Availability Management, which deals with the design, implementation, measurement, and management of IT services to ensure the business requirements for availability are met.
- (5) Capacity Management, which will be covered in depth in the next section of the paper.

3. ITIL Capacity Management

According to the IT Infrastructure Library, Capacity Management is "responsible for ensuring that the Capacity of the IT infrastructure matches the evolving demands of the business in the most cost-effective and timely manner."

In general, the process includes:

- (1) Monitoring of performance of IT services and the infrastructure components that support those services.
- (2) Recommending tuning activities to make the best use of resources.
- (3) Understanding current demands on IT resources and making forecasts for future requirements
- (4) Influencing the demand for resources, usually in conjunction with Financial Management
- (5) Producing a Capacity Plan to enable the IT Service Provider to meet Service Level Agreements, both now and the in future.

Capacity Management has always been a balancing act in a number of different ways:

- (1) Cost against capacity. Purchases should be cost-justifiable against business need, and once those purchases are made, used in the most efficient manner.
- (2) Supply against demand. Capacity Management ensures that there is adequate (but not excessive) supply of processing power to meet current and future demands of the business.

The best way to balance those requirements is through an effective interface with Service Level Management and effective negotiating, monitoring, and policing of Service Level Agreements. This will be discussed more in depth in section 6 of the paper.

3.1 Objectives of ITIL Capacity Management

High-level objectives of ITIL Capacity Management sound pretty familiar to those who have worked in a formal Capacity Planning or Performance Management group in the past. Here is a list of typical objectives for an ITIL Capacity Management process:

- (1) Ensure the right level of IT investment
- (2) Identify and resolve bottlenecks
- (3) Evaluate tuning strategies
- (4) Improve and report/publish performance
- (5) "Right-size" or "Consolidate" where appropriate
- (6) Ensure accurate and timely procurements
- (7) Ensure effective service level management
- (8) Plan for workload growth, new apps, new sites
- (9) Avoid performance disasters

From these objectives, a goal statement of the Capacity Management process might be:

The Capacity Management process's goal is to make sure that adequate, costeffective capacity is available to meet the business needs, both now and in the future.



Sounds like something every company is doing or should be doing, right? So why is a best practice view like ITIL needed to best meet these goals?

As we'll discuss, the answer to this question is two-fold: First, the complexity of today's environment requires that we look at Capacity Management on levels different than just the hardware resource level. Secondly, effective Capacity Management requires effective interfaces to other Service Management processes, which we'll discuss in depth in section 6.

3.2 ITIL Capacity Management levels (or sub-processes)

The complexity of today's environment requires that Capacity Management be done on varying levels. Traditionally, Capacity Management (or Capacity Planning or Performance Management, depending on the company) dealt with the performance and capacity of individual hardware resources both now and in the future. Today's applications (or services) are so complex and the IT Services so important to the success of the business, that organizations cannot afford to look at hardware resources and technology in a vacuum, but must look at these resources in the context of an IT service or a business requirement.

In view of these challenges, three levels of Capacity Management were defined within ITIL: Resource Capacity Management, Service Capacity Management, and Business Capacity Management.



Figure 3. ITIL Capacity Management Levels

Let's look at each one of these in depth:

(1) Resource Capacity Management is the traditional view of Capacity Planning and Performance Management where the focus is on the management of the individual components of the IT Infrastructure. In this sub-process, all resources are monitored; data recorded and reported, analyzed, trended, and modeled. Usually, at this level, Capacity Management is assisted by technology specialists who can help the Capacity Manager apply specialist knowledge to recommendations and decisions.

Some objectives at the Resource level may include:

- (a) Identify and understand the capacity and utilization of each component part of the IT infrastructure
- (b) Recommend optimization of hardware and software
- (c) Measure and store resource usage at a process level
- (d) Identify bottlenecks and potential future problems

- (e) Characterize workloads and business drivers
- (f) Evaluate alternative upgrades to meet workloads
- (g) Proactive rather than reactive
- (h) No surprises in performance or IT budgets
- (2) Service Capacity Management is the level where the focus is the management of IT Services (better known as "IT Applications" in many environments). This process is responsible for monitoring and measuring all services to ensure that SLAs are met and to ensure that end-to-end capacity is adequate for these services both now and in the future.

Some objectives are the Service level may include:

- (a) Identify and understand the IT services
- (b) Assess services' use of resources
- (c) Identify services' working patterns, peaks, troughs
- (d) Ensure SLA targets are viable
- (e) Monitor performance to identify SLA violations
- (f) Resource data aggregated by service (application)
- (g) Prevent difficulties wherever possible
- (h) Proactive rather than reactive
- (3) Business Capacity Management is responsible for considering future business requirements for IT services and ensuring that they are planned and implemented in a timely fashion. This can be done by using existing data and incorporating trends and models to predict future requirements as well as using business plans identifying new services, growth and improvements in existing services, etc.

Some objectives and methods for achieving those objectives at the Business level may include:

- (a) Published corporate performance objectives
- (b) Standard local metrics defining contribution
 - i. Unification of analytic information
 - ii. Improved managers' business insights
 - iii. Greater accountability via Key Performance Indicators (KPI)
 - iv. Resource data aggregated by service/application and then weighted
- (c) Enterprise framework for measurement
 - i. Published reports and exception reports
 - ii. Automated alarms and interpretations
 - iii. Interactive Dashboards for alert/drill down
 - iv. Predicted outcomes across framework
- (d) Business agility to adjust as necessary
 - i. Strategic modeling to view scenarios
 - ii. Ensured focus and drive to growth
 - iii. Effective liaison between IT and management





4. Utility to transportation to insurance (oh my)

My personal career experience has included working in three medium-to-large sized companies as a capacity manager in the last decade or so: The first was an electric utility, the second was an overnight shipping company, and the third was an insurance company. We'll look at two of these – the utility company and the insurance company.

4.1 Utilities - turning the lights on

In a 5 analyst, 50-server utility company environment, little was automated. Capacity Management was solely at the resource level and mainly consisted of looking at daily charts and making sure that the CPU utilization of the systems was below 90% at all times.

When customers started complaining about performance, management started holding daily "morning meetings" to talk about operations issues that were fueled by these internal and external customer complaints.

While the ongoing joke was that the meeting was simply a mechanism for the director to ensure everyone was at work by 8 AM, the analyst who couldn't provide at least a cursory analysis of any trouble ticket-related question was usually the subject of the director's scorn for a while. Staff typically referred to the meeting as the morning "get yelled at" session.

Fueled by the need for rapid information retrieval and by people like my director (and by the desire for the yelling to stop), even companies like mine that had only a few dozen servers started automating the data capture, collection, and retention process to meet the increasing reporting and analysis demands of the data center management.

Vendor tools were developed and purchased by many companies, analysts were trained to use those tools, data was stored in a systematic manner for easy retrieval, analysis, and reporting, and the nature of the performance analyst changed from a person who spent a lot of time generating reports that might be viewed by one or two people to a person who spent more time doing analysis of the data that used to be so hard to get.

Once the process was automated, the analysts could easily arrive at the office by 7: 30AM and have enough information by 8AM in order to survive the morning meeting and drill down into the problem afterwards. This time period became known as "charts over coffee" and was the inspiration for the title of the paper.

However, life was good – then again, with about 50 servers and 5 dedicated analysts, the tasks were hardly Herculean. With only about ten servers per person and the entire data capture and manipulation process automated, the analysts could spend a good portion of the working day becoming very familiar with the data patterns. Problems such as unusually high user activity or looping processes or long batch windows were quickly detected and corrected.

Unfortunately, even back in the salad days where the number of servers and analysts was in a more reasonable proportion than today, there was a high level of disconnect between the data center and the business units.

When business managers would ask for details on why performance was poor, analysts (such as the Author, unfortunately) would write highly technical documents that demonstrated servers in terms of resource consumption and using metrics and terms that no business manager could possibly be expected to understand.

For reasons more related to pride than understanding, the business managers asked few questions and generally accepted the explanations and recommendations of the technical staff, even though their real questions generally went unanswered. Notice also that the last few paragraphs reference servers, not applications, and certainly not end-to-end user experience or response times. A decade ago, many of the applications running on distributed systems ran on one, or at most a few, servers.

Rather than consider applications as a whole and considering the end-user experience (Service Capacity Management), analysts would look at reports for individual servers and take a very large step by theorizing that if all the servers have "reasonable" resource consumption numbers that the application likely had acceptable performance.

This Resource-oriented mentality only added to the disconnect that existed between the data center management and the business unit management.

So eventually, many of the Capacity Management activities described in the previous section were happening at the utility company – Monitoring, Analysis, Tuning, Implementation, and Storage of the (utilization) Capacity Management Data, yet only on a Resource level. And most of the data was used only in a reactive way.

4.2 "Insuring" proper ITIL Capacity Management

After a while, I moved on -- from moving electricity to moving packages to moving insurance.

I most recently worked in an environment with almost a thousand distributed servers, with about 75% of those in the production tier.

At least 300 of those servers were various flavors of UNIX, with the rest being various releases of Microsoft Windows.

Many of the applications in that environment were distributed across multiple tiers, including the mainframe, and many of the applications were complex, web-based applications that required a lot of specialized knowledge to administer.

So, considering the huge increase in the number of servers and the increasing complexity of the distributed applications, the number of analysts responsible for daily reporting, analysis, and Capacity Planning changed as well.

Indeed, the number decreased – from 5 to 4. Instead of each analyst being responsible for about ten servers, each analyst was responsible for hundreds of servers and dozens of services.

And yet, Capacity Management there was proactive, respected, and worked better than in a Utility Company where there was a much better analyst-to-server ratio. How?

Capacity Management, at the insurance company, was a vital part of the business and of every new and existing project. From the project kick-off meeting, Capacity Management was involved – learning business requirements, understanding technical requirements, and also learning about end-user expectations. As part of this process, Capacity Management would work closely with Service Level Management in negotiating SLAs with the client and then making sure that any SLAs could be properly monitored and policed once the service was in place, in production.

The following is a list of Capacity Management activities according to ITIL and how the insurance company tackled those activities:

- (1) Monitoring Capturing of utilization data of each resource and service was automated and target thresholds were set at a level that ensured SLAs were met. Exceeding of thresholds potentially raised an incident or problem for further analysis.
- (2) Analysis Tools were purchased to help analysts identify trends to establish utilization baselines and investigate threshold or SLA breaches.



9



- (3) Tuning Capacity Management, via monitoring and analysis activities, were able to identify potential tuning recommendations, e.g. workload balancing and disk traffic balancing.
- (4) Implementation Capacity Management worked closely with Change Management to best implement any recommendations that came from the monitoring, analysis, and tuning activities.
- (5) Storage of Capacity Management Data data was stored in a Capacity Database (CDB). As is most often the case, this data was not stored in a single relational database, but rather in multiple data repositories that were easily accessible by Capacity Management. Data stored included business level data, applicationspecific data, and utilization data. Data was also stored for financial and accounting purposes, although a chargeback system was not (yet) in place.
- (6) Demand Management demand management was mostly limited to the mainframe and mainly to encourage testing after the prime shift. Testing during prime shift would have created latent demand on the mainframe and would've likely required a significant upgrade of the existing hardware. However, by limiting testing to off-hours, the company was able to use the capacity more efficiently.
- (7) Modeling The insurance company heavily utilized analytic and simulation modelling for Resource and Service Capacity Management using predictions of additional and new business provided in the Business Capacity Management subprocess.
- (8) Application Sizing Data was captured during the application development and testing process and used, in combination with analytic and simulation modeling, to influence quality development and ensure services had adequate capacity on rollout and in the foreseeable future and would be able to meet the agreed-upon service level agreements.
- (9) Production of the Capacity Plan A capacity plan was written for each new service and on a recurring basis for critical services in the company. While a bit different from the ITIL recommendation of a global, annual capacity plan, the process of writing formal plans, even at this level, ensured that all new and all critical applications received the proper level of attention.

Most of the Capacity Management done in this environment was proactive in nature. However, when performance problems did occur (as they always do), the company was well positioned to analyze those issues because of the maturity of the Capacity Management process.

5. Implementing ITIL Capacity Management

So, how does a company get its Capacity Management process from the utility company described here to the insurance company described here?

Implementing Capacity Management is not a matter of buying a bunch of tools and hiring a handful of employees. Rather, implementing Capacity Management successfully depends on the same factors as successfully completing any other project. Therefore, a project management discipline should be followed, and some necessary steps are listed here:

- (1) Identify a sponsor. Without an executive sponsor, implementing ITIL Capacity Management will not succeed. Usually, implementing an ITIL process will involve purchasing tools, committing and training resources, and ensuring cooperation with other existing processes. Without an executive sponsor fully on board, the project will not succeed.
- (2) Identify the project team. Many will be involved and getting commitment to such a project up front is necessary for the project's success.
- (3) Identify the process owner. Within ITIL, there must be a person who, ultimately, owns a process. This may be a manager of a formal Capacity Management group or just a "functional manager" who pulls together needed resources to make the process work, but someone must be responsible and accountable for the process's success.
- (4) Ensure that proper funding is available. While this would seem to fit in with identifying an executive sponsor, it is worth mentioning again because that sponsor must be made aware of the costs involved with properly implementing ITIL Capacity Management.
- (5) Determine the scope of the project. While ITIL recommends that the Capacity Management process is responsible for all resources with finite capacity, a phased approach that covers a subset of those resources that will result in "quick wins" and visibility of the process (in a positive way) will help ensure success.
- (6) Develop the process mission and vision. It's important to communicate a raison d'être to all, especially the business units. If the project team and process owner cannot communicate the mission and vision effectively, it will be hard for Capacity Management to receive necessary buy-in from the business units and other parts of IT.
- (7) Determine SMART (Specific, Measurable, Achievable, Realistic, and Time) objectives.
- (8) Have a communication and awareness campaign. Ongoing communication will allow Capacity Management to communicate important information (mission, vision) upfront and those quick wins once the project goes live.
- (9) Conduct a gap analysis. A gap analysis is necessary when implementing any ITIL process as many of the tools and activities are (likely) already in place in the organization. The gap analysis will look at existing people, processes, and tools, as well as the current budget and the current interfaces to other ITSM processes and find what needs to be added or improved in order to put the desired level of ITIL Capacity Management in place.
- (10) Design the process. Will Capacity Management be centralized or distributed throughout a global organization? Which tools were identified in the gap analysis? Will the capacity database be centralized, distributed throughout the organization, or a hybrid that combines elements of both? Will adequate business, service, resource, and financial data be captured in order to meet the Capacity Management needs of the business?



11.



- (11) It is important to design the Capacity Management process keeping in mind the other Service Support and Service Delivery processes, NOT in isolation. Identify which processes will provide information to Capacity Management and which processes will receive information from Capacity Management and work closely with those process owners and teams.
- (12) Evaluate, select, and implement tools. Potential tools would include a capacity database (CDB), modeling tools, analysis tools, reporting tools, statistical packages, etc. Some of these tools would be identified in the gap analysis.
- (13) Implement the process. Establish monitoring at the desired scope, establish the CDB, train staff, link resource and service data to SLAs and SLRs, plan and produce capacity plans, etc.
- (14) Review. Is the utilization of resources and services being stored in the CDB? Is the right level of data being captured? Are the SLAs policed and Service Level Management notified of breaches? Are reports produced at the right level and on time? Is the Capacity Plan produced on time and is it accepted by management?

As part of the post-implementation review and ongoing review, look at critical success factors, such as:

- Are forecasts accurate?
- Is there an understanding of current and future technology?
- · Can cost-effectiveness be demonstrated?
- Does Capacity Management have knowledge of the business plans and is that knowledge incorporated in the Capacity Plan?





6. Interfaces to other ITIL ITSM processes

Capacity Management cannot be successful without close interfaces with the other ITSM processes.

As a vital element of the planning process, Capacity Management interfaces with the Service Delivery processes in order to improve those processes. Those processes and examples of interfaces include:

- Financial Management -- budgeting, accounting, and charging
- Availability Management -- availability metrics are frequently captured and stored by Capacity Management and included in the Capacity Plan
- IT Service Continuity Management -- Capacity Management typically models ITSCM scenarios
- Service Level Management -- Capacity Management typically polices SLAs and helps to negotiate appropriate SLAs with the business units.

Further, ITIL Capacity Management provides support for all operational performance and capacity issues. The more the Service Support processes rely on Capacity Management, the better those processes will be. Those processes and examples of interfaces include:

- Configuration Management The Capacity Database (CDB) is a subset of the Configuration Management Database (CMDB). The CMDB provides Capacity Management with technical, service, utilization, financial, and business data. Without this data, Capacity Management cannot function effectively.
- Incident Management Frequently, Capacity Management detects performance and capacity related alerts automatically. Incident Management will also provide information to help Capacity Management analyze and recommend solutions to capacity related issues.
- Problem Management Capacity Management will provide specialist support to identify, diagnose, and resolve capacity-related problems and also support proactive Problem Management through analysis and identification of trends.
- Change Management Capacity Management is represented on the Change Advisory Board (CAB) to assess the impact of changes on capacity. Additional capacity requirements and recommendations are requests for change (RFCs) handled via Change Management.
- Release Management Capacity Management can help ensure needed capacity based on the release strategy (network bandwidth, for example, in the case of a network distribution). Capacity audits can be used to delay releases if there is insufficient bandwidth.

13_____ ITIL Capacity Management - More than charts over coffee



7. Conclusion

In fact, many of today's companies are doing a pretty good job of ITIL Capacity Management without realizing (or formalizing) it. Many of the activities mentioned earlier in the paper are being completed and Capacity Management is a vital part of many IT environments.

A thorough gap analysis will help best align a company's existing Capacity Management process to ITIL. However, ITIL Capacity Management cannot be done in isolation – Capacity Management relies on all other IT Service Management processes and is also relied upon by those same processes. It is vital that those interfaces exist and relationships between process owners and teams prosper.

In summary, ongoing improvement of ITIL Capacity Management and a strong alignment with the other IT Service Management processes not only makes ITIL Capacity Management better, but all of ITSM, IT, and the business itself better as well.

8. References (Acknowledgments)

- [1] "Best Practice for Service Delivery," Office of Government Commerce, ISBN 0-11-330017-4, 2001.
- [2] "Corporate Performance Management as a Pragmatic Process in an ITIL World," A. Grummitt, Proceedings of the Computer Measurement Group, December 2004.
- [3] "ITIL Capacity Management Deep Dive," C. Molloy, Proceedings of the Computer Measurement Group, December 2005.
- [4] "ITIL Capacity Management in Practice," Metron Technology, Ltd., 2005.

9. Trademarks

ITIL is a registered trademark and a registered community trademark of the Office of Government Commerce (OGC) and is registered in the United States Patent and Trademark Office.







©2006 Metron Technology Ltd / Metron-Athene Inc Metron, Metron-Athene and the Metron logo as well as Athene and other names of products referred to herein are trade marks or registered trade marks of Metron Technology Limited. Other products and company names mentioned herein may be trade marks of the respective owners. Any rights not expressly granted herein are reserved. ITIL is a Registered Trade Mark, and a Registered Community Trade Mark of the Office of Government

ITIL is a Registered Trade Mark, and a Registered Community Trade Mark of the Office of Government Commerce, and is Registered in the U.S. Patent and Trademark Office.