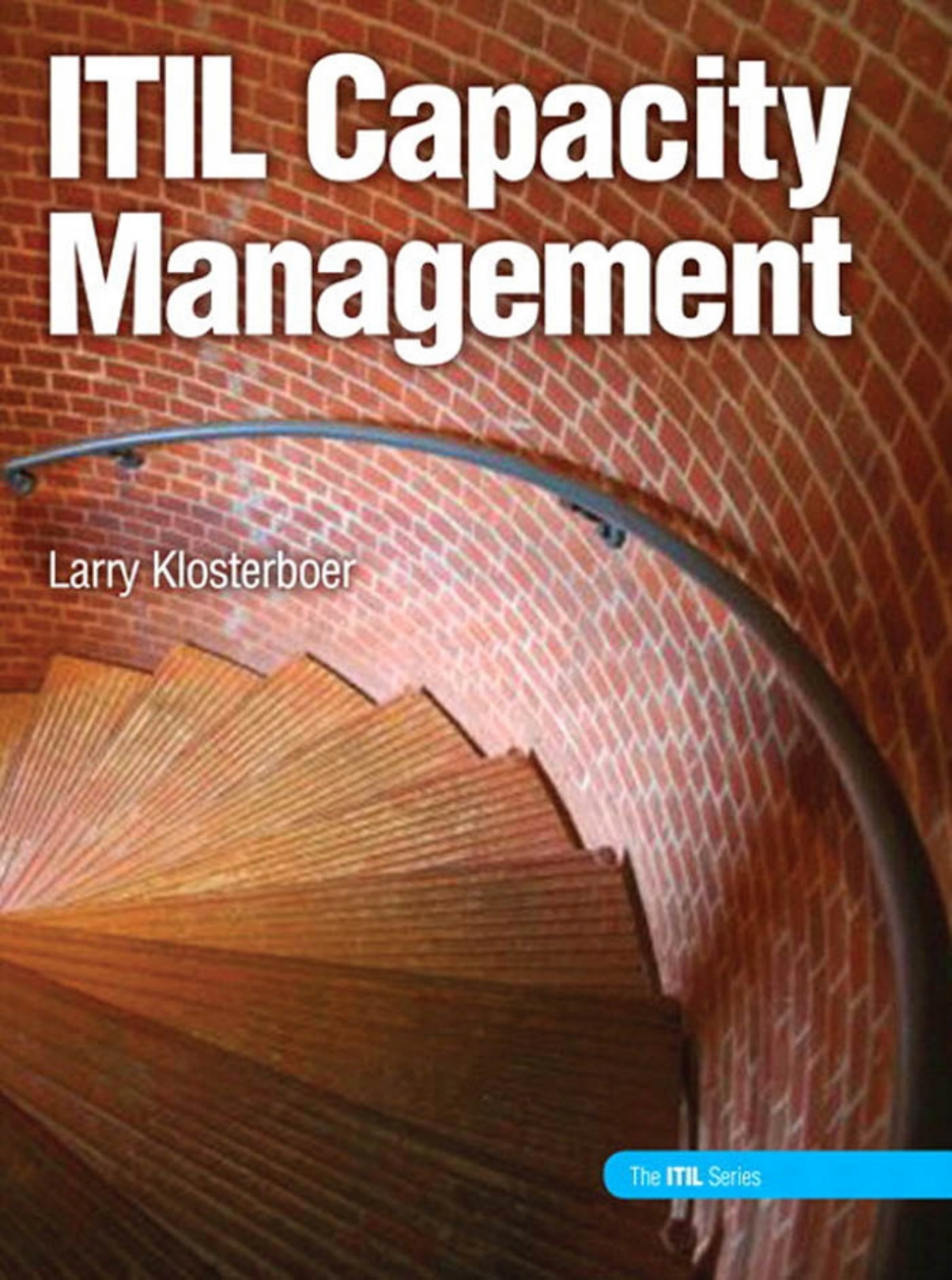
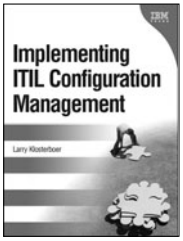
ITIL Capacity Management

Larry Klosterboer



The ITIL Series

Related Books of Interest



Implementing ITIL Configuration Management, 2nd Edition

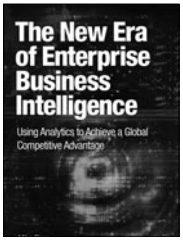
By Larry Klosterboer

ISBN: 0-13-138565-8

Practical, Real-World ITIL Configuration Management—From Start to Finish

Now, there's a practical, start-to-finish guide to ITIL configuration management for every IT leader, manager, and practitioner. ITIL-certified architect and solutions provider Larry Klosterboer helps you establish a clear roadmap for success, customize standard processes to your unique needs, and avoid the pitfalls that stand in your way.

You'll learn how to plan your implementation, deploy tools and processes, administer ongoing configuration management tasks, refine ITIL information, and leverage it for competitive advantage. Throughout, Klosterboer demystifies ITIL's jargon, illuminates each technique with real-world advice and examples, and helps you focus on the specific techniques that offer maximum business value in your environment.



The New Era of Enterprise Business Intelligence, 1st Edition

Using Analytics to Achieve a Global Competitive Advantage

By Mike Biere

ISBN: 0-13-707542-1

A Complete Blueprint for Maximizing the Value of Business Intelligence in the Enterprise

The typical enterprise recognizes the immense potential of business intelligence (BI) and its impact upon many facets within the organization—but it's not easy to transform BI's potential into real business value. In *The New Era of Enterprise Business Intelligence*, top BI expert Mike Biere presents a complete blueprint for creating winning BI strategies and infrastructure, and systematically maximizing the value of information throughout the enterprise.

This product-independent guide brings together start-to-finish guidance and practical checklists for every senior IT executive, planner, strategist, implementer, and the actual business users themselves. Drawing on thousands of hours working with enterprise customers, Biere helps decision-makers choose from today's unprecedented spectrum of options, including the latest BI platform suites and appliances. He offers practical, "in-the-trenches" insights on a wide spectrum of planning and implementation issues, from segmenting and supporting users to working with unstructured data.

Listen to the author's podcast at: ibmpressbooks.com/podcasts

Of course, capacity is also affected by hardware refresh. We make the assumption that obsolete hardware is always replaced with hardware that is faster, bigger, and better. What this really means is that the new component normally has a larger capacity than the one being replaced. From the perspective of a capacity manager, this greater capacity needs to be tracked because all the assumptions you might have been making about the capacity of an IT service using this component could now be false.

While a hardware refresh project might introduce more capacity in one part of an IT service, it does not necessarily increase the overall capacity of the entire service. Consider an IT service to provide an enterprise service bus that all applications can use to interoperate with one another. Refreshing the server that transmits messages would seem like a logical way to allow more messages to flow, thus increasing the capacity of the services. This would only be true, however, if the network bandwidth already supported the increase in message traffic through the server. Many very costly mistakes can be avoided if you remember to analyze the complete service before ordering hardware or software to add capacity where you think it might do the most good.

Grow the Business

Capacity management would be much easier if growth happened only in your existing base set of components and services. But we know that IT is a dynamic environment where new services are added, old services are retired, and existing services often add or lose components. We call this dimension "business growth" because it is largely determined by the business needs and priorities that your IT team is working with. This section describes how to track and manage IT capacity as you grow the business.

Project-Based Growth

Most IT organizations think growth happens as a result of projects. As the business identifies new needs, those needs get translated to IT project requests. When the projects are approved and implemented, they inevitably result in more IT demand. Some projects add capacity to meet their own demand, while others use existing capacity. Either way, the capacity manager must track the new capacity and the new demand.

Let's first consider a traditional project that implements a new business application. If this is a standard commercial application and not custom developed, the software publisher provides a description of the environment that should be deployed. Normally this includes a specific operating system type, the amount of CPU, memory and disk space required, and often the prerequisite versions of supporting software. The project team acquires exactly that configuration for production use and sometimes acquires a matching configuration for other uses such as test, training, and preproduction staging.

Assume that the application requires a Java™-based application server, a database, and a web server. The project team knows that this application will be heavily used, so they specify that each

of these functions should be on a separate server. In an environment without adequate capacity management, each function of the application is placed on its own physical server with just the middleware that supports that function—the Java environment with the application on one server, the web pages and configuration files with the web server on a second machine, and the database with a database management system on a third machine. When you move to a second similar application, you need six separate servers, as shown in Figure 4.2.

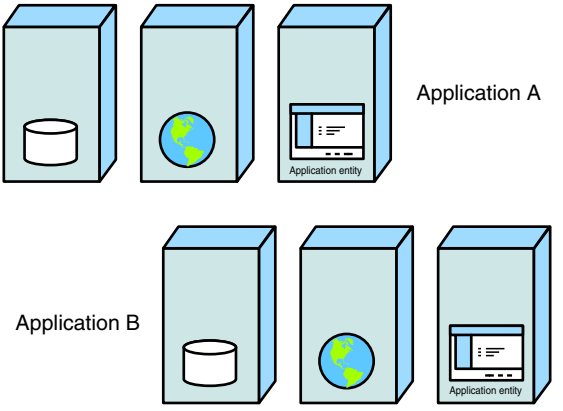


Figure 4.2 Traditional application deployment uses many servers.

If you see nothing wrong with this deployment method, capacity management will save your organization lots of money. With adequate capacity management you should quickly recognize that a single server can hold one licensed copy of the database management system and support dozens or even hundreds of individual databases that might be used by many different applications. Instead of using a separate database server for every application that requires one, you create a database server capacity pool that many applications can share. Of course, you'll want more memory and CPU power in your server than you would have assigned for a dedicated server, but even with more expensive hardware you will still save money over separate hardware and separate DBMS software for each application.

Of course, this same theory holds true for web servers and Java application servers. You can create and manage a single pool of capacity and then share that capacity with new applications as they are built. This is a much more cost-effective solution, but it depends on your ability to manage these capacity pools instead of relying on the vendors' advice for server sizes. If your organization hasn't already taken advantage of shared servers in this way, this alone can be a huge win for your capacity management program. Figure 4.3 shows how you can save three servers by sharing two different applications on the same set of database, web, and application servers.

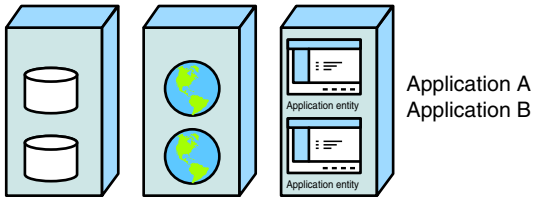


Figure 4.3 Shared deployment treats each server as a capacity pool.

An alternative way to provide capacity growth for new applications is by deploying virtual servers. In this scheme, you purchase very large servers and deploy a virtualization engine such as Microsoft Hyper-V or VMware ESX. These engines enable you to host dozens of separate operating system images, known as "virtual machines," on a single physical server. Using virtualization, you can still deploy a separate, dedicated database, web, and application server for each new application. The difference is that all these can be deployed on one piece of hardware, greatly reducing data center costs and consumption over either the dedicated server or the shared server approach and enabling you to build new servers dramatically faster than if you were following a physical server model. Figure 4.4 illustrates this approach for two applications.

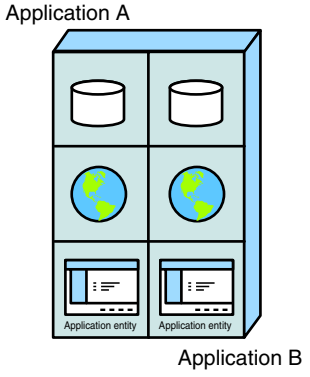


Figure 4.4 Virtualization greatly reduces the number of servers needed.

Virtualization gives you, as a capacity manager, a completely different set of capacity pools to manage. Each physical server becomes a pool of CPU and a pool of memory, and you allocate resources from these pools each time a new virtual machine is needed. When all the physical CPU and memory on the server reaches a high utilization rate, another server is added and the capacity pools have more space in them.

Service-Based Growth

When your organization gets more mature in its journey toward IT service management, you will begin to realize that growth happens in the context of services. New services are introduced that require completely new infrastructure components or place additional demands on existing components. A service change such as increasing the availability may dictate that new components are added to the mix. Even a service retirement might generate a temporary need for more capacity as data is archived or transition activities take place. As a service-oriented IT organization, you must plan for the capacity changes that service transitions produce.

You will understand and document the growth of a service only if you have a complete understanding of the components that make up the service. As discussed in Chapter 2, "The Geography of Managing Capacity," this is the base for all service capacity management. Assuming that you do have this understanding, tracking service growth is a simple matter of following the same formula you use for calculating service capacity utilization in the first place.

Service growth is more difficult to project into the future as services become more complex. Additional components that make up the service become additional variables that might change the growth projections. This doesn't mean you shouldn't try, however. When you first begin to track service capacity and use trends to project service growth, you may not be very accurate, but you will be acquiring important skills. As your capacity management maturity and skill increase, your forecasts will include trends and known future changes to the service and will become more accurate.

Over time you will observe the ways that your services grow. When you record these observations, you will be able to develop better procedures for tracking service capacity.

Summary and Next Steps

This chapter takes the optimistic view that your organization is growing, and describes how to measure and manage the corresponding growth of your IT environment. In many cases your IT environment may be growing even if the overall organization is not. Tracking and managing IT growth is an important role for the capacity manager.

You first learned about base growth—the natural growth that happens as your IT systems are used and as they evolve to newer versions. You learned that the capacity manager is responsible for setting the upper limit for system growth because in many cases the computer systems and software don't do that for themselves. You also learned that there is a gap between the way that tools report this growth and the way it should be displayed for management consumption. It is up to the capacity manager to bridge that gap.

You next looked at how business growth can impact IT capacity growth. You saw that the business can be expanding in one of two ways. It can provide new goods or services to its customers, which will generally grow revenue and probably expenses as well. When this growth

occurs, there is almost always a resulting growth in IT utilization, which may involve growing the infrastructure. The other way that business can grow is by implementing internal projects to cut costs, enhance security, optimize compliance with industry regulations, or make the business more sound in other ways. These projects may or may not increase the demand on IT, and consolidation projects often reduce the IT demand. The savvy capacity manager needs to understand how the business is growing and be ready to respond with the best possible information to help the business make decisions.

The first four chapters have introduced you to the key concepts of capacity management. We've set the context of capacity management against the ITIL framework, and you've learned how capacity management must interact with other ITIL process disciplines to be successful. You looked at capacity pools and capacity streams as a metaphor for what ITIL calls "component" capacity management and "service" capacity management. You've learned how to make predictions about future IT demand, and in this chapter you've begun to think about the ways that growth impacts your capacity management program. Now that you have a firm foundation, you can explore the best practices in capacity management as defined by the ITIL process.

In the next chapter you look at the first of those best practices—the capacity management information system (CMIS). You'll learn why you need a CMIS and how to manage it. You will explore the CMIS in a way that doesn't depend on any single tool or implementation approach, and then we circle back around to the best tools in Chapter 10, "Choose Capacity Management Tools."

Part II

Best Practices in Capacity Management

In Part I, you learned the general concepts about managing IT capacity. In this second part of the book, you learn the specific details needed to begin actively managing capacity for your organization. The chapters in this part delve into the details that are mentioned as best practices in the ITIL documents. By pulling from these best practices, you should have a solid base to implement your very own capacity management program.