Practical Service Level Management: Delivering High-Quality Web-Based Services
By John McConnell, Eric Siegel
Publisher: Cisco Press
Pub Date: January 14, 2004
Pages: 304

Measure, manage, and improve the speed and reliability of web services

- Complete reference for creating relevant, effective Service Level Agreements
- Detailed discussions of both technical and business performance metrics and their statistical treatment
- Performance and management implications of various web services delivery infrastructures, including caching and load distribution
- Discussion of the transport infrastructure, including quality of service (QoS) technology and traffic shaping
- Instrumentation system design
- Measurement collection, aggregation, correlation, and use for real-time service level control and reporting
- Quick problem detection, "triage" problem diagnosis, and root-cause analysis
- Automated, policy-based system management
- Load testing, modeling, and capacity planning for web systems
- Calculation of return on investment for web infrastructure improvements
- Structured plan for implementation of SLM techniques

The web has become a major vehicle for transforming business processes, but ineffective management of web-based services can result in high costs and user dissatisfaction. Service Level Management (SLM) is therefore a competitive weapon in the web marketplace, providing the tools needed to improve performance and reliability of web services while simultaneously controlling costs.

Practical Service Level Management: Delivering High-Quality Web-Based Services shows you how you can measure, manage, and improve network performance and quality of experience (QoE) for critical web services. Starting with an explanation of SLM and common performance metrics, the book provides detailed discussions of methods to measure and improve performance. Service Level Agreements, instrumentation, performance-improvement technologies, load testing, and long-term planning are all covered in detail. This book provides both technical engineers and non-technical managers with an organized, cohesive plan for measuring, improving, and evaluating the performance of web-based services.

Whether you are delivering services to other businesses or directly to customers, Practical Service Level Management: Delivering High-Quality
Web-Based Services walks you through the complete process of designing a balanced solution for your situation. Use it to help design a system with the speed, reliability, and flexibility that are critical success factors for your business.

This book is part of the Networking Technology Series from Cisco Press, which offers networking professionals valuable information for constructing efficient networks, understanding new technologies, and building successful careers.
Summary

Chapter 5. Event Management

Event Management Overview
Basic Event Management Functions: Reducing the Noise and Boosting the Signal
A Market-Leading Event Manager: Micromuse
Summary

Chapter 6. Real-Time Operations

Reactive Management Proactive Management Automated Responses Examples of Commercial Operations Managers
Summary

Chapter 7. Policy-Based Management

Policy-Based Management The Need for Policies A Policy Architecture Policy Design Policy Product Examples
Summary

Chapter 8. Managing the Application Infrastructure

Interaction of Operations and Application Development Teams
Application-Level Metrics
Transaction Response Time: An Example of Dependence on Lower-Level Services Instrumenting Applications Summary

Chapter 9.
Managing the Server Infrastructure
Architecture of the Server Infrastructure Instrumentation of the Server Infrastructure Summary

Chapter 10.
Managing the Transport Infrastructure
Technical Quality Metrics for Transport Services QoS Technologies Managing Data Flows Among Organizations Summary

Part III. Long-term Service Level Management Functions
Chapter 11. Load Testing
The Performance Envelope Load Testing Benchmarks Load Test Beds and Load Generators Building Transaction Load-Test Scripts and Profiles Using the Test Results Summary
Chapter 12. Modeling and Capacity Planning
   Advantages of Simulation Modeling
   Complexity of Simulation Modeling
   Simulation Model Examples
   Capacity Planning
   Summary
Part IV. Planning and Implementation of Service Level Management
   Chapter 13. ROI: Making the Business Case
   Impact of ROI on the Organization
   A Basic ROI Model
   Soft Benefits
   ROI Case Study
   Summary
   Chapter 14. Implementing Service Level Management
   Phased Implementation of SLM
   An SLM Project Implementation Plan
   Summary
   Chapter 15. Future Developments
   The Demands of Speed and Dynamism
   Evolution of Management Systems
   Integration Architectural Trends for Web Management Systems
   Business Goals for Service
Copyright

Copyright © 2004 Cisco Systems, Inc.

Published by:
Cisco Press
800 East 96th Street
Indianapolis, IN 46240 USA

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the publisher, except for the inclusion of brief quotations in a review.

Printed in the United States of America 1 2 3 4 5 6 7 8 9 0

First Printing January 2004

Library of Congress Cataloging-in-Publication Number: 2001097399

Warning and Disclaimer

This book is designed to provide information about service level management. Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an "as is" basis. The author, Cisco Press, and Cisco Systems, Inc. shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the author and are not necessarily those of Cisco Systems, Inc.

Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members from the professional technical community.

Readers' feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through e-mail at feedback@ciscopress.com. Please make sure to include the book title and ISBN in your message.
We greatly appreciate your assistance.

**Trademark Acknowledgments**

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc. cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

**Corporate and Government Sales**

Cisco Press offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales. For more information, please contact:

U.S. Corporate and Government Sales 1-800-382-3419 corpsales@pearsontechgroup.com

For sales outside of the U.S., please contact:

International Sales 1-317-581-3793 international@pearsontechgroup.com

**Credits**

<table>
<thead>
<tr>
<th>Role</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publisher</td>
<td>John Wait</td>
</tr>
<tr>
<td>Editor-in-Chief</td>
<td>John Kane</td>
</tr>
<tr>
<td>Executive Editor</td>
<td>Brett Bartow</td>
</tr>
<tr>
<td>Cisco Representative</td>
<td>Anthony Wolfenden</td>
</tr>
<tr>
<td>Cisco Press Program Manager</td>
<td>Sonia Torres Chavez</td>
</tr>
<tr>
<td>Manager, Marketing Communications, Cisco Systems</td>
<td>Scott Miller</td>
</tr>
<tr>
<td>Cisco Marketing Program Manager</td>
<td>Edie Quiroz</td>
</tr>
<tr>
<td>Production Manager</td>
<td>Patrick Kanouse</td>
</tr>
<tr>
<td>Acquisitions Editor</td>
<td>Michelle Grandin</td>
</tr>
<tr>
<td>Development Editor</td>
<td>Jill Batistick</td>
</tr>
<tr>
<td>Project Editor</td>
<td>Marc Fowler</td>
</tr>
<tr>
<td>Copy Editor</td>
<td>Jill Batistick</td>
</tr>
<tr>
<td>Technical Editors</td>
<td>David M. Fishman</td>
</tr>
<tr>
<td></td>
<td>John P. Morency</td>
</tr>
<tr>
<td></td>
<td>Richard L. Ptak</td>
</tr>
<tr>
<td>Team Coordinator</td>
<td>Tammi Barnett</td>
</tr>
<tr>
<td>Book Designer</td>
<td>Gina Rexrode</td>
</tr>
<tr>
<td>Cover Designer</td>
<td>Louisa Adair</td>
</tr>
</tbody>
</table>
Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the Cisco.com Web site at www.cisco.com/go/offices.
In Loving Memory

This book was finished as a final tribute to my late husband, John McConnell. I hope these words keep his ideas alive in the industry a little longer.

Grace Morlock McConnell

My perception of my very special son, John W. McConnell

All that we can be, we must be.
Find a star and never settle for less.
John was born to be one of a kind.
Making his way with a mind of his own,
and making a difference and making it known.
He had his dreams and hopes to pursue.

by his mother, Jeanette McConnell
Dedication

John W. McConnell

December 9, 1943  November 3, 2002

This book is dedicated to my wife, Grace, whose support has been so helpful in carving out the time and quiet needed for this project. My friends and Grace have also provided a supportive environment and tolerated my frequent absences to work with clients. Returning home to a warm community has been really important to me.

Acknowledgments

Many people have been part of this process of turning some ideas and experience into a book. First, my thanks to the Cisco Press team, especially Michelle Grandin. The steady enthusiasm and willingness of all to help are deeply appreciated.

In the same vein, the technical reviewers have been so helpful. I've had the pleasure of spending good time exchanging views with John Morency and Rich Ptak at many analyst conferences and other events; their suggestions for this manuscript were specific and helpful, and in some cases spurred some spirited discussions. Although I've never met David Fishman face to face, I'd be pleased to buy him a good meal someday as thanks for so many good suggestions and his attention to detail and integrity on getting it right.

Another group I want to acknowledge are the clients I've worked with around the world. I've gotten to learn a lot about how technology is actually used and to work with people who want to push the envelope.

Finally, my thanks to my friends and colleagues in the industry who constantly stimulate and challenge me. It's been a tremendous blessing to be among so many creative and independent thinkers and doers that have shaped the networking industry.

John McConnell

It's impossible to begin these acknowledgements without wishing that John were still alive. This is his book, not mine. He conceived it; he drafted it; he should have been writing this page. We all used to joke about how John "towered over the industry," and it wasn't just because of his height. In working from John's drafts to complete the book, in talking to colleagues about his work, and in remembering the easy, jovial way he talked about examples of industry practices, I was constantly reminded of his stature and of the friendly way he had. I think I can say, with confidence, that everyone in the industry truly misses him; I certainly do.

John's wife wanted to see this book come to publication, and Cisco Press went far out of their way to make that happen. Jill Batistick and Michelle Grandin, the editors, were wonderfully friendly and helpful; they made the process of working through the chapters almost enjoyable. The technical reviewers, Rich Ptak, John Morency, and David Fishman put a tremendous amount of work into the book. They didn't just point out my errors; they suggested corrections and entire new paragraphs that could improve the text. They were truly partners in bringing the book to publication.
I'd also like to thank Astrid Wasserman, of MediaLive International, Inc., (the organizers of Networld+Interop), who gave me a copy of John's proposed two-day seminar on Service Level Management. Although it was never presented, the seminar slides gave me a lot of insight into his ideas.

I have tried to stay close to John's original thoughts and text, although I have occasionally succumbed to temptation and added additional information. Minor additions occur in all chapters; major additions are in Chapter 2 (measurement statistics), Chapter 6 (triage for quick assignment of problems to appropriate diagnostic teams), Chapter 8 (transaction response time), and Chapter 11 (flash loads and abandonment). Most of the additions are topics that I had discussed with John at various conferences we attended together; I hope, and believe, that he would agree with them. In all cases when the author speaks directly to the reader, that author is John.

Eric Siegel
October 14, 2003

About the Authors

John McConnell was involved in networking for over 30 years. A member of the ARPANET working group, John contributed to early Internet architecture and protocol development. John has consulted with clients in the U.S., Europe, Asia, and the Middle East, and he has designed some of the first TCP/IP networks deployed in Europe and the Middle East.

John served as a consultant in the areas of systems and network management with a focus on Service Level Management (SLM), policy-based management solutions, and the emerging issues of management solutions for e-business.

John received a master's in electrical engineering and computer science from the University of California, Berkeley.

Eric Siegel, Principal Internet Consultant with Keynote Systems, Inc., "the Internet performance authority," first worked on the Internet in 1978. He wrote Designing Quality of Service Solutions for the Enterprise (John Wiley & Sons) and has taught Internet performance tuning, SLM, and quality of service (QoS) at major industry conferences, such as Networld+Interop.

Before joining Keynote Systems, Eric was a Senior Network Analyst at NetReference, Inc., where he specialized in network architectural design for Fortune 100 companies, and he was a Senior Network Architect with Tandem Computers, where he was the technical leader and coordinator for all of Tandem's data communications specialists worldwide. Eric also worked for Network Strategies, Inc. and for the MITRE Corporation, where he specialized in computer network design and performance evaluation. Eric received his B.S. and M.Eng. degrees in electrical engineering from Cornell University, where he was elected to the Electrical Engineering honor society.

About the Technical Reviewers

David M. Fishman is at Sun Microsystems, where he is responsible for availability measurement strategies in the office of Sun's Chief Customer Advocate. Prior to that, he managed Sun's strategic...
technology relationship with Oracle, driving technology alignment on High Availability (HA), Java
technology, and performance. Before joining Sun in 1996, Fishman held a variety of technical and
business development positions at Mercury Interactive Corporation in a variety of product management
and business development roles. Previous work experience includes high-tech marketing and
management in defense electronics, embedded systems, and office automation. David holds an MBA
from the School of Management at Yale University. He lives in Sunnyvale, California, with his wife
and two children.

John P. Morency is a 29-year veteran of the networking and telecommunications industries and
president of Momenta Research, Inc., a company that he founded in 2002. His industry experience
includes network software development, technical support, IT operations, industry consulting, product
marketing, and business development.

Because of his wide range of experience, John has a very unique ability to effectively assess the
business, technological, and operational impacts of new products and technologies. This is evidenced by
the significant business case and Total Cost of Ownership (TCO) work that John has done on behalf of
hundreds of Fortune 1000 clients over the past ten years, resulting in hundreds of millions of dollars in
both top- and bottom-line benefits.

John's current research is focused on the business benefits attributable to the implementation of wireless
LANs (Wi-Fi), network telephony, content networking, system and network security, Web services,
disaster recovery, and IT process automation.

He is the author of over 400 publications on the operations and business impact of new IT technology.
His speakership and publication credentials include Networld+Interop, Network World, Billing World,
Broadband Year, LightWave, Telecommunications, and Telecom-Plus International, among many
others.

Richard L. Ptak, founder of Ptak & Associates, Inc., has more than 25 years experience providing
consulting services on the use of IT resources to achieve competitive advantage. Ptak earned his B.S.
and M.S. at Kansas State University. His MBA was earned at the University of Chicago.

Preface

Some years ago I received a true pearl of wisdom from an industry colleague. "In order to truly
understand your profession," he advised, "you must make the effort to learn other disciplines that are
completely different from the one that you espouse."

That colleague was John McConnell, a man who truly understood this advice by walking the talk over
the course of his life. Born into a military family, John developed a keen understanding of the
importance of the global ecosystem at a very young age through his childhood experiences in both
Europe and the Far East. Despite being a shy, scholarly individual throughout primary and secondary
school, John also demonstrated the value of hard work and dedication by making the varsity rowing
team at U.C. Berkeley.

The strong work ethic that John nurtured at Berkeley served him well after he received his master's in
computer science in 1968. What differentiated John from many of his fellow graduates, however, was
the application of his craft to non-IT disciplines after graduation. Some of his first initiatives included
the application of computer technology to measure the rate of solar intensity upon the earth and the
development of a programming language that was designed to test the content and substance of moon
samples brought back to earth by the Apollo astronauts. In addition, John developed a number of network control programs for the ARPANET (the predecessor to today's Internet) in the mid-1970s when the state of the commercial data networking industry was in its true infancy.

John also spent a number of years in professional capacities that had very little to do with information technology. After graduate school, John became an accomplished massage therapist, hypnotist, and practitioner in the art of Rolfing, a technique for the detection, treatment, and removal of bodily stress and pain. In 1983, using his Rolfing technique, John was selected to work with the members of the U.S. bicycling Olympic team, and he applied this technique to aid the team in preparing for the 1984 Olympic games. Recently, when not consulting, John was training to become an instructor in the Ridhwan Foundation, an institution whose focus is the rediscovery and integration of the true self into one's own professional and personal life. Over the years, he had a myriad of personal interests including soaring, mountain climbing, bird watching, backpacking, rowing, and blues festivals. One of his most recent and satisfying accomplishments was the design, building, and completion of a second home in southern Costa Rica that effectively enabled both he and his wife Grace to really get away from it all.

First and foremost, John's professional focus in the IT industry was the advancement of technologies and products that improved the efficiency and the effectiveness of IT management.

Given his whole life background, John was especially dedicated to reducing the operational and business "pain points" associated with IT implementation and management. This focus is reflected in John's prior work, Internetworking Computer Systems and Managing Client/Server Environments, as well as in Practical Service Level Management: Delivering High-Quality Web-Based Services. John's numerous publications, conferences, and televised briefings reflect a focused dedication to the removal of technological barriers to the optimal effectiveness of IT organizations worldwide. His life experiences of a true Renaissance man uniquely enabled him to both understand and drive the level of change needed to not only improve state of the art, but also quality of life. John was indeed the "gold standard" of knowledge, professionalism, and personal integrity that made the pursuit of these goals not only a logical possibility, but, for many of us, a practical reality. The loss of John will be keenly felt for some time, but the goals and values that he aspired to and embraced will inspire and guide many of us for years to come.

John Morency, President, Momenta Research
May 2003
Chapter 1. Introduction

The World Wide Web is the catalyst for the changes in our communications, work styles, business processes, and ways of seeking entertainment and information. The Internet is just the transport infrastructure for the web-based services that drive so much innovation. Note, however, that the Internet generally gets all the credit. As Thomas Friedman writes in The Lexus and the Olive Tree:

The Internet is going to be like a huge vise that takes the globalization system that I have described and keeps tightening and tightening that system around everyone, in ways that will only make the world smaller and smaller and faster and faster with each passing day.

This is an accurate description of the environment that most of us deal with directly on a daily basis. The Internet is a tremendous business engine, and, as it transforms the ways we do business, it is being transformed in turn by the ways we use it. We must learn how to manage the growing array of online business services or risk being marginalized by a faster moving and more dynamic business environment.

In this introductory chapter, I discuss the following:

- The types of e-business services
- A definition of webbed services and the webbed ecosystem
- Service Level Management (SLM)
- The structure of this book

E-business Services

E-business is a generic term defining business activities that are carried out totally, or in part, through electronic communications between distributed organizations and people. These activities are characterized by speed, flexibility, and constant change.

The Internet has become the vehicle for transforming business processes. The reasons for its ascendancy include the following:

- The Internet protocols are the only workable set of technologies that really provide a high degree of interoperability among different systems.
- The wide geographic reach of the Internet increases the size of any potential market.
- Internet economies make it feasible to distribute information and transact business globally.
- The introduction of the browser and its supporting technologies make the Internet much easier to use, thereby increasing the potential market.

There are many ways of segmenting and describing the large variety of services available through the Internet and the Web. A simple classification that covers most services is based on the relationship of the business to customers, business partners, and employees. For example, the process shown in Figure 1-1 describes a simple situation involving all three types of relationships: business to business (B2B), business to consumer (B2C), and business to employee (B2E). These segments are an easy way of organizing our thinking about services, although it's important to remember that business processes in the real world will have many variations and overlaps.
The following sections discuss each relationship type in turn.

**B2B**

B2B services are a broad category that incorporates transactions among different businesses and government agencies. Many current B2B services, such as supply chain management and credit authorization, use the Internet to drive down the costs and delays associated with current processes and to boost their productivity.

B2B is rapidly broadening to include more than supply chain management and credit authorization. Functions such as shipping, billing, and Customer Relationship Management (CRM) are now often external to the business; other businesses provide and host these specialized services as a utility. For example, entry of a customer's order can result in more than the functions of pricing, authorizing, assembling, and shipping; a modern system might use B2B links to provide the customer with a shipment tracking number from the shipping company, and it might interact with an external CRM service to reflect the current purchases and other factors of the customer's profile. Meanwhile, the sales person might be indirectly using B2B links to handle her commissions and personnel data through outsourced employee management services, and engineering staff might use B2B links for collaborative design.

Thanks to the Web, B2B is rapidly transforming into an even more dynamic set of services from which an enterprise can select in real time. No one wants to be dependent on a single supplier or customer; everyone must deal with competitive pressures exerted from both sides. Services such as credit authorization and shipping are examples of those that can be selected in real time based on their
performance or costs. Other services and supplies may be selected from web-based exchanges or e-markets.

B2B processes can be complex. They must follow the business requirements for tracking orders, negotiating contracts, arranging payments, and reporting outcomes that govern these processes when they take place without the automation of electronic communications.

Note that new benefits become available, although at the cost of additional complexity, when B2B replaces older systems. For example, organizations can change their business processes to increase their business effectiveness by obtaining real-time information on order volumes, revenue rates, cancelled orders, and other factors. This additional information, while adding to complexity, provides value in addition to the acceleration of the processes themselves by identifying further efficiencies.

Continuous monitoring of B2B suppliers, partners, and web infrastructure (communications, hosting, and exchanges) is necessary to determine whether they are meeting their service quality commitments.

As in conventional commerce, managing across organizations adds complexity. All the links in the B2B services chains are known, but these links are controlled by many different organizations, are complex, and may change rapidly as services are selected in real time. Managing B2B services therefore requires cooperation with the management teams of the other participants and, possibly, with third-party measurement organizations to assure true end-to-end service quality.

B2C

B2C garnered most of the early attention from the trade press and analysts as traditional businesses took advantage of the Internet's wide geographic reach and low costs for reaching customers. Some businesses (eBay and Amazon.com, for example) were founded to exploit this new market opportunity.

B2C sites continually add new services of their own while offering links to related businesses and services in an attempt to offer one-stop shopping and selling to their customers. This is a highly competitive segment with little customer loyalty. The wide selection of competing sites draws customers away whenever any one site has a service disruption.

B2C environments are characterized by a lack of visibility and management control of the customer-access infrastructure, which is the set of networks, caches, and other systems that consumers use to connect to the B2C site. Customers usually don't want measurement tools embedded in their systems, and the access infrastructure providers also resist making their internal performance readily visible. There is also limited visibility into the performance of partner sites (advertisers and other third parties), which are important parts of the customer's perception of total site performance. The span of control and management available to B2C sites is therefore usually limited to monitoring and managing their internal operations (inside the firewall) as well as measurement of Internet delays and performance as seen from various points on the edge of the Internet.

B2E

B2E services are also known as the intranet. These services help improve the internal effectiveness of an organization and help it keep pace with its customers and business partners. Many B2E services enable employees to query their benefits, schedule vacations, fill out expense reports, and conduct a set of activities that formerly required a large staff to coordinate.

B2C and B2E services use the web browser as the access device. Transactions are initiated from the browser to deliver information and activate a range of business processes. However, B2E environments
are the only ones that enable administrators to have control of both ends—the servers as well as the
desktops, cell phones, and personal communicators used to access them.

Webbed Services and the Webbed Ecosystem

In this book, I use the term webbed services to describe the set of business services that are based on a
component approach to systems design. This design is driven from the Web and its associated
technologies, regardless of the specific technologies used. Because webbed services are constructed
from a set of interconnected software components and services that can be reused in multiple places,
they can usually avoid some of the expense, time, and effort associated with building and modifying
monolithic applications.

Webbed services is a very inclusive term; it's increasingly difficult to find services that are not
somehow tied into the Web. As a case in point, I was recently speaking about webbed services at a large
retail organization, and someone in the group stated that their main application did not fit into the
webbed category because it was a stand-alone Oracle Financials application. However, further
discussion soon revealed that their international operations used real-time currency conversion
decisions. The real-time exchange rates in the Oracle Financials application were, in fact, accessed
through the Web.

Indeed, webbed services are now taking on many of the characteristics of an ecosystem, which is a
group of independent but interrelated elements comprising a unified whole. A smooth business process
depends on each element carrying out its tasks accurately and quickly, with consideration for
maintaining balances among all the elements. In a well-balanced webbed ecosystem, all elements bear
appropriate shares of the load. None is overwhelmed, none is underutilized. Balance is concurrently
maintained between service quality and service cost. The ecosystem metaphor is gaining momentum as
online processes evolve to dynamically select their elements (underlying services) based on their current
behavior and performance.

The webbed ecosystem perspective also holds within any subgroup of systems. For instance, hosting
facilities use a range of technologies, such as prioritizing devices, bandwidth managers, global load
balancers, and caches, to deliver online business services. These systems also need balanced
management; adding bandwidth when servers are congested is a wasteful investment.

Service Level Management

Service quality is extremely important, given the accelerating number of critical business processes
going online. Customers and business partners go elsewhere if the services they want are not available
or are performing sluggishly. Unfortunately, good service quality is a dynamic target and the demands
continue to tighten. Competitors will match or exceed service quality levels and create pressure toward
matching or bettering theirs.

Service Level Management (SLM) is the process of managing network and computing resources to
ensure the delivery of acceptable service quality at an acceptable price in an acceptable time frame. It
focuses on the behavior of the services rather than on tracking the status of every router, switch, and
server in the environment. Through SLM, service quality is guaranteed and priced for different levels of service.

SLM is a competitive weapon in the marketplace, offering the guarantees needed to transition critical business activities online. Poorly managed services have harmed many businesses when their web sites crashed, their applications slowed to a crawl, or their Web content was not attractively presented or was too difficult to navigate. Good service quality helps retain customers and differentiate your organization from those that have not yet mastered the art of managing service quality.

Effective SLM is also an economic weapon. Managing resources more effectively reduces costs, creates more revenue opportunities, and leverages technology investments.

Finally, SLM is a means to build the solid business relationships that make online business initiatives successful.

The basic terminology of SLM is as follows:

- **Service Level Agreement (SLA)** A formal, negotiated contract between a service provider and a service user that defines the services to be provided, the service quality goals (often called service level indicators and service level objectives), and the actions to be taken if the service provider does not comply with the SLA terms.

- **Quality of Service (QoS)** A technology-centered concept that focuses on the performance of the transport and service technologies underlying a webbed service. Examples are service availability, response time, and the technologies that measure and assure specific levels of transport infrastructure performance (packet loss, network transit time, and transit time variations).

- **Quality of Experience (QoE)** A customer-centered concept that focuses on monitoring and assessing service quality from the end-customer perspective. This includes someone using a browser to access information and order merchandise, or a business conducting a series of exchanges to order products, negotiate terms, and arrange payment.

QoE is the most important to customers, yet it is also the most difficult to evaluate. For example, I recently visited a large company that derives over half its revenues online. They were justly proud of a new initiative that reduced web page download times by two seconds. However, the content was so dense and difficult to navigate that users still needed a long time to understand the directions and identify the buttons or links they wanted to use next. Improved technical performance did not appreciably raise the QoE in this case; users wasted at least two seconds looking for what they wanted.

The book is divided into four parts:

- **Part I: Service Level Agreements and Introduction to Service Level Management** (Chapters 1–3): Chapter 1 defines the webbed ecosystem, and Chapter 2 defines and discusses SLAs along with typical technical and business process metrics, their statistical treatment, and recommendations for writing SLAs. Chapter 3 outlines both the overall architecture for service delivery on the Web and the overall architecture for managing that delivery. (Additional, detailed examples of technical metrics and management architectures are given in Part II.)

- **Part II: Components of the Service Level Management Infrastructure** (Chapters 4–10): The first group of chapters (4–7) in this part discusses the details of the service management
infrastructure. The group starts with measurement collection and aggregation technologies (Chapter 4) and then continues through the filtering and integration of real-time, measurement-detected events (Chapter 5). It concludes with the use of those filtered events by the operations staff (Chapter 6) and by automated, policy-based management systems (Chapter 7).

The second group of chapters (8–10) in this part steps through the major systems used for web service delivery. It looks at the ways they can be used to improve service delivery and also discusses their specific instrumentation needs, using the system management infrastructures described in the first part of this section. Chapter 8 investigates the instrumentation and management of applications and of end-user access devices, such as browsers. Chapter 9 looks at web server systems, including servers, load balancers, and content distribution networks. Finally, Chapter 10 discusses instrumentation and management of the transport infrastructure, including QoS technology and traffic shaping to achieve policy objectives.

- **Part III: Long-term Service Level Management Functions (Chapters 11–12)** This part covers load testing, modeling, and capacity planning. No management system can provide necessary quality if the web serving system, as a whole, has insufficient capacity.
- **Part IV: Planning and Implementation of Service Level Management (Chapters 13–15)** Calculation of Return on Investment (ROI) for SLM is critical to the justification and design of an implementation; it's covered in Chapter 13. Chapter 14 provides guidance for using the information in this book to design an SLM system for your particular situation, and the part ends with discussion in Chapter 15 of possible future developments in SLM.

### Summary

The Internet, and the Web, are transforming business processes for interaction among businesses, government, suppliers, customers, and employees. As more and more critical business processes go online, the service quality of those processes becomes more important to the success of business as a whole.

SLAs are the formal, negotiated contracts between service providers and service users that define the services to be provided, their quality goals, and the actions to be taken if the SLA terms are violated.

SLM is the process of managing network and computing resources to ensure the delivery of acceptable service quality, usually as defined in an SLA, at an acceptable price in an acceptable time frame. It is a competitive weapon in the marketplace because it can improve customer relationships, create more revenue opportunities, and reduce costs.

### Chapter 2. Service Level Management

Service Level Management (SLM) is a key for delivering the services that are necessary to remain competitive in the Internet environment. Service quality must remain stable and acceptable even when there are substantial changes in service volumes, customer activities, and the supporting infrastructures.