



Capacity Planning for Web Services: Metrics, Models, and Methods

By Daniel A. Menasce, Virgilio A.F. Almeida

Download now

Read Online ➔

Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida

In *Capacity Planning for Web Services: Metrics, Models, and Methods*, two leading Web capacity planning experts introduce quantitative performance predictive models for every major Web scenario -- especially the latest Web services, e-commerce, and mobile applications. Discover how to identify and address potential performance problems, accommodate usage peaks, define service provider SLAs, benchmark effectively -- and optimize every trade-off between cost and performance.

↓ [Download Capacity Planning for Web Services: Metrics, Model ...pdf](#)

📖 [Read Online Capacity Planning for Web Services: Metrics, Mod ...pdf](#)

Capacity Planning for Web Services: Metrics, Models, and Methods

By Daniel A. Menasce, Virgilio A.F. Almeida

Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida

In *Capacity Planning for Web Services: Metrics, Models, and Methods*, two leading Web capacity planning experts introduce quantitative performance predictive models for every major Web scenario -- especially the latest Web services, e-commerce, and mobile applications. Discover how to identify and address potential performance problems, accommodate usage peaks, define service provider SLAs, benchmark effectively -- and optimize every trade-off between cost and performance.

Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida **Bibliography**

- Rank: #1651323 in Books
- Published on: 2001-09-21
- Original language: English
- Number of items: 1
- Dimensions: 9.40" h x 1.30" w x 7.00" l, 2.09 pounds
- Binding: Paperback
- 608 pages



[Download Capacity Planning for Web Services: Metrics, Model ...pdf](#)



[Read Online Capacity Planning for Web Services: Metrics, Mod ...pdf](#)

Download and Read Free Online Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida

Editorial Review

From the Back Cover

The #1 guide to Web capacity planning -- now completely updated!

- A quantitative analysis of Web service availability
- An integrated coverage of benchmarking, load testing, workload forecasting, and performance modeling of Web services
- Example and case studies show how to use each technique in the latest Web services, portals, search engines, mobile and streaming-media applications
- A quantitative framework for planning the capacity of Web services and understanding their behavior

The world's #1 book on Web capacity planning now covers the latest Web services, e-business, and mobile applications! *Capacity Planning for Web Services: Metrics, Models, and Methods* introduces quantitative performance predictive models for every major Web scenario, showing precisely how to identify and address both potential and actual performance problems. Coverage includes:

- Web services: protocols, interaction models, and unique performance, reliability, and availability challenges
- State-of-the-art capacity planning methodologies
- Spreadsheets implement the solutions of the models presented in the book
- Specific issues and workloads associated with HTTP and TCP/IP protocols
- Benchmarking current performance at system and component levels

From accommodating current usage peaks to defining service provider SLAs, Daniel A. Menasce and Virgilio Almeida cover every aspect of capacity planning — helping you optimize every tradeoff between cost and performance.

"This book is the best guide available to understanding the unique performance challenges involved in delivering today's Web services."

— Mark Crovella, Associate Professor, Boston University;
Technical Director, Network Appliance

"...a superb starting point for anyone wishing to explore the world of Web performance."

— Jeffrey P. Buzen, President of CMG;
Co-Founder, BGS Systems

"There is no other book like this. It is a first."

— Peter J. Denning, Professor of Computer Science, George Mason University and former President of the ACM

"Web servers have bursty and highly-skewed load characteristics. This book presents a new way to model, analyze, and plan for these new performance problems."

— Jim Gray, Senior Researcher, Microsoft Research;
1998 ACM Turing Award Recipient

"...a welcome approach to the performance analysis of today's Web-based Internet. ... no simple and practical treatment has been offered before, and theirs is a timely contribution."

— Leonard Kleinrock, Professor of Computer Science, UCLA;
Chairman and Founder, Nomadix, Inc.

About the Author

DANIEL A. MENASCÉ is Professor of Computer Science at George Mason University and a Fellow of the Association for Computing Machinery. He has consulted extensively on Web performance worldwide. Menascé holds a Ph.D. in Computer Science from UCLA.

VIRGILIO ALMEIDA, Professor of Computer Science at the Federal University of Minas Gerais, Brazil, holds a Ph.D. in Computer Science from Vanderbilt University and was formerly a visiting researcher at Xerox PARC and at HP Labs. Menascé and Almeida co-authored *Scaling for E-Business* (Prentice Hall PTR).

Excerpt. © Reprinted by permission. All rights reserved.

Preface

This is what makes this book unique and worth reading: Businesses, government agencies, and individuals are increasingly dependent on Web services for their day-to-day operation. As a consequence, quality of service, in particular performance and availability considerations, are becoming extremely important. Performance analysis of Web services is unique in many ways. First, Web services rely on large-scale systems that consist of thousands of computers, networks, software components, and users. Large-scale systems are inherently complex. The randomness associated with the way users request Web services compounds the problem of managing and planning the capacity of those services.

This book uses a quantitative approach to analyzing Web services. It provides a framework to understand the Web's complex relationships and how these relationships impact performance and availability of Web services. This approach lends itself to the development of performance and availability predictive models for managing and planning the capacity of Web services. Instead of relying on intuition, ad hoc procedures, and rules of thumb, we provide a uniform and sound way for dealing with performance problems. The models discussed here are based on probability fundamentals and on the theory of queuing networks. The techniques and models presented in this book provide a simple and practical approach to dealing with Web-based systems. Models are essential tools to simplifying the representation of complex systems and, at the same time, to capturing the most relevant characteristics of those systems.

There are several books on computer system performance evaluation, but they do not focus on Web-based systems. The Web has special features that make its performance problems unique and demand novel

approaches to dealing with them. This book presents a sound and practical approach to addressing these challenges. Also, this book is not tied to any specific technology or product; it is general enough to be valuable in any Web-based environment.

This is an example-driven book. Many real-world examples with solutions are used to illustrate the use of models and analytic techniques. The models are available as Microsoft Excel worksheets, which makes it easy for readers to understand detailed situations of capacity planning of real-world Web services. The case studies presented are inspired in problems that come up very often in the course of planning, designing, implementing, operating, and managing Web services. The examples, case studies, and spreadsheets allow the reader to immediately put into practice the methods and models discussed here.

Who Should Read This Book

Information technology professionals must ensure that the Web services under their management provide an acceptable quality of service to their users. Managers must avoid the pitfalls of inadequate capacity and meet users' performance expectations in a cost-effective manner. System administrators, Web masters, network administrators, capacity planners and analysts, managers, consultants, and other IT professionals will benefit from reading parts or the entire book. Its practical, yet sound and formal, approach provides the basis for understanding modern and complex networked environments.

This book can be used as a textbook for senior undergraduate and graduate courses in Computer Science and Computer Engineering. At the undergraduate level, the book is a good starting point to motivate students to learn the important implications and solutions to performance problems. In particular, this book is very useful to in providing students with the quantitative skills required to analyze the behavior of Web systems. At the graduate level, it can be used in System Performance Evaluation courses. This book offers a theoretical and practical foundation in performance modeling. The book can also be used as a supplement for systems courses, including Operating Systems, Distributed Systems, and Networking, both at the undergraduate and graduate levels. A series of exercises and short projects is available at the Web site associated with this book (www.cs.gmu.edu/~menasce/webservices/).

Book Organization

Chapter 1 introduces, through a series of examples, the importance of performance and availability considerations for Web services. The concepts of capacity planning, service levels, and workload evolution are defined and illustrated.

Chapter 2 presents a brief discussion on local and wide area networks and their protocols including TCP/IP, Ethernet, IEEE 802.11, Token Ring, and FDDI. The chapter revisits the client server (C/S) computing paradigm and discusses several kinds of servers, such as file servers, database servers, application servers, groupware servers, object servers, and Web servers. Various architectural C/S issues, including two-tier versus three-tier C/S architectures, are presented. The HTTP protocol is described in this chapter. The Peer-to-Peer model is also discussed as an alternative to the C/S paradigm used to support Web services. Standards-based technologies for Web services, such as Simple Object Access Protocol (SOAP) and Web Service Description Language (WSDL) are also presented. The concepts here are presented in light of performance considerations and tradeoffs.

Chapter 3 investigates, in detail, the nature of the delay incurred by a typical Web request. A complete analysis of the time to download a Web page is presented. This chapter shows how service times can be computed at single disks, disk arrays, networks, and routers. Queues and contention are defined more formally and some very basic and important performance results from Operational Analysis are introduced.

Chapter 4 discusses issues that affect performance of Web services. The chapter starts by looking at the sources of delay in Web environments. After discussing the end-user perspective to Web service performance, the chapter provides an assessment. It examines the components and protocols involved in the execution of Web services and analyzes their capacity and performance issues. New concepts such as heavy-tailed characteristics, traffic spikes, and network caching are analyzed through the use of simple performance models. The Wireless Application Protocol (WAP) is also presented in this chapter.

Chapter 5 introduces a step-by-step methodology to determining the most cost-effective Web service infrastructure. The main steps of the methodology are: understanding the environment, workload characterization, work-load model validation and calibration, performance and availability model development, model validation and calibration, workload forecasting, performance and availability prediction, cost model development, cost prediction, and cost/performance-availability analysis.

Chapter 6 describes and illustrates with examples the major steps required for the construction of workload models. The common steps to be followed by any workload characterization project include specification of the standpoint from which the workload will be analyzed, choice of the set of parameters that capture the most relevant characteristics of the workload for the purpose of the study, monitoring the system to obtain the raw performance data, analysis and reduction of performance data, and construction of a workload model. Some characteristics of Web workloads are discussed and power-laws models are introduced.

Chapter 7 presents several different standard industry benchmarks-such as TPC-C and TPC-W, SPEC CPU, SPECWEB, and WebStone-and shows how to use benchmark results as a complementary source of information to support the capacity planning methodology. Mathematical models that represent characteristics of actual Web workloads are studied in this chapter. This chapter also examines performance testing, which is commonly used to determine how users will experience the performance of a Web service.

Chapter 8 starts by introducing very simple models of Web-based systems. Complexity is progressively introduced and the solution to each model is presented using first principles and intuitive concepts. After a few models are presented, the approach is generalized. The models presented in this chapter are called system-level performance models since they view the system being modeled as a "black box."

Chapter 9 introduces powerful techniques to analyzing the performance of intranets and Web sites. The chapter considers the components that make up a Web-based system. Component-level models account for the different resources of the system and the way they are used by different requests. The solution techniques for component-level models are based on Queuing Networks (QN). Solution methods for both open and closed QNs with multiple classes of customers are presented in this chapter. The chapter also shows how system- and component-level models can be combined to address more complex modeling situations.

Chapter 10 shows how the performance models discussed in previous chapters can be specialized to handle specific Web performance issues. The chapter also shows how certain special features common in Web workloads, such as burstiness and heavy-tail distributions of file sizes, can be accounted for in performance models for the Web.

Chapter 11 introduces availability, reliability and performability as important metrics that define the quality of service provided by Web-based applications. A quantitative framework, based on the state transition diagram, is developed to model availability, reliability, performability, software aging and rejuvenation.

Chapter 12 shows how to apply existing forecasting methods and techniques to predict the future workload for Web services. Forecasting techniques includes linear and non linear regression models, moving averages, and exponential smoothing. A strategy and a methodology for workload forecasting is discussed here. It encompasses selection of the workload to be forecast, analysis of historical data and estimation of workload

growth, selection of a forecasting technique, application of the forecasting technique to the historical data, and analysis and validation of forecast results.

Chapter 13 presents a framework for collecting performance data in Web environments. Next, it discusses the main issues in the process of measuring the performance of Web-based systems. The chapter does not focus on any specific product or manufacturer. Instead, it presents general procedures for transforming typical measurement data into input parameters. The procedures can be thought of as a set of major guidelines for obtaining input parameters for performance models.

Chapter 14 wraps up the book by emphasizing the approach and methods used to develop and solve analytical performance models of Web services. It also discusses properties that should be pursued for the design and implementation of large-scale information systems. For any service provider, whether financial services, merchants, or government services, it is necessary to guarantee the quality of service, which is represented by key issues such as trustworthiness and security, performance, and availability. The chapter discusses the importance of modeling to address these key issues. Appendix A contains a glossary of the important terms introduced in the book.

Appendix B provides a description of the Microsoft Excel workbooks that accompany this book and describes the C program provided to derive burstiness parameters for HTTP logs.

Book's Web Site and Authors' Addresses

The Web site at www.cs.gmu.edu/~menasce/webservices/ will be used to keep the readers informed about new developments related to the book and to store the various Excel workbooks described in the book.

The authors' e-mail, postal addresses, and Web sites are:

Professor Daniel A. Menasce
Department of Computer Science, MS 4A5
George Mason University
Fairfax, VA 22030-4444
United States
(703) 993-1537
menasce@cs.gmu.edu
www.cs.gmu.edu/faculty/menasce.html

Professor Virgilio A. F. Almeida
Department of Computer Science
Universidade Federal de Minas Gerais
P.O. Box 920
31270-010 Belo Horizonte, MG
Brazil
+55 31 3499-5860
virgilio@dcc.ufmg.br
www.dcc.ufmg.br/~virgilio

Users Review

From reader reviews:

Leslie Marcellus:

Do you have favorite book? Should you have, what is your favorite's book? Reserve is very important thing for us to be aware of everything in the world. Each book has different aim or maybe goal; it means that e-book has different type. Some people really feel enjoy to spend their time for you to read a book. These are reading whatever they acquire because their hobby is actually reading a book. Consider the person who don't like looking at a book? Sometime, man feel need book whenever they found difficult problem or maybe exercise. Well, probably you should have this Capacity Planning for Web Services: Metrics, Models, and Methods.

Kimberly Williams:

Book will be written, printed, or created for everything. You can understand everything you want by a guide. Book has a different type. As it is known to us that book is important matter to bring us around the world. Beside that you can your reading proficiency was fluently. A publication Capacity Planning for Web Services: Metrics, Models, and Methods will make you to possibly be smarter. You can feel much more confidence if you can know about almost everything. But some of you think this open or reading any book make you bored. It's not make you fun. Why they may be thought like that? Have you trying to find best book or suited book with you?

Andre Botsford:

This Capacity Planning for Web Services: Metrics, Models, and Methods are usually reliable for you who want to be a successful person, why. The main reason of this Capacity Planning for Web Services: Metrics, Models, and Methods can be among the great books you must have is giving you more than just simple examining food but feed a person with information that possibly will shock your earlier knowledge. This book is usually handy, you can bring it just about everywhere and whenever your conditions in e-book and printed ones. Beside that this Capacity Planning for Web Services: Metrics, Models, and Methods giving you an enormous of experience like rich vocabulary, giving you tryout of critical thinking that we realize it useful in your day pastime. So , let's have it and enjoy reading.

Miranda Durkee:

This book untitled Capacity Planning for Web Services: Metrics, Models, and Methods to be one of several books that best seller in this year, this is because when you read this guide you can get a lot of benefit upon it. You will easily to buy this book in the book shop or you can order it by means of online. The publisher of the book sells the e-book too. It makes you easier to read this book, since you can read this book in your Cell phone. So there is no reason for you to past this publication from your list.

**Download and Read Online Capacity Planning for Web Services:
Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F.
Almeida #1YNFBW9SL7R**

Read Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida for online ebook

Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida Free PDF d0wnl0ad, audio books, books to read, good books to read, cheap books, good books, online books, books online, book reviews epub, read books online, books to read online, online library, greatbooks to read, PDF best books to read, top books to read Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida books to read online.

Online Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida ebook PDF download

Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida Doc

Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida Mobipocket

Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida EPub

1YNFBW9SL7R: Capacity Planning for Web Services: Metrics, Models, and Methods By Daniel A. Menasce, Virgilio A.F. Almeida