

Web Server Technology

The Advanced Guide for World Wide Web
Information Providers

Nancy J. Yeager
Robert E. McGrath

An excellent book for anyone who wants to understand
all the issues surrounding Web server technology.

— Ed Krol, author of *The Whole Internet User's Guide & Catalog*

Web Server Technology

The Advanced Guide for
World Wide Web Information Providers

Nancy J. Yeager
Robert E. McGrath

National Center for Supercomputing Applications



Morgan Kaufmann Publishers, Inc.
SAN FRANCISCO, CALIFORNIA

Sponsoring Editor	Michael B. Morgan
Production Manager	Yonie Overton
Production Editor	Elisabeth Beller
Text Design	Mark Ong, Side by Side Studios
Cover Design	Martin Heirakuji Graphic Design
Cover Photograph	Photonica/Joshua Sheldon
Copyeditor	Ken DellaPenta
Proofreader	Judith Abrahms
Composition	Nancy Logan
Illustration	Cherie Plumlee
Indexer	Valerie Robbins
Printer	Courier Corporation

Morgan Kaufmann Publishers, Inc.
 Editorial and Sales Office
 340 Pine Street, Sixth Floor
 San Francisco, CA 94104-3205
 USA

Telephone	415/392-2665
Facsimile	415/982-2665
E-mail	mkp@mkp.com
WWW	http://www.mkp.com

Order toll free 800/745-7323

© 1996 by Morgan Kaufmann Publishers, Inc.
 All rights reserved
 Printed in the United States of America

00 99 98 97 96 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the publisher.

Library of Congress Cataloging-in-Publication Data is available for this book.

ISBN 1-55860-376-X

Cha

memory and disk storage systems, where access to the same information means accessing the same storage location. Popular data is seen as a concentration of accesses to particular locations, or a highly *localized* pattern.

The degree of locality in Web requests is rather low; browsing the Internet is a very unstructured activity. Cache hit rates seem to reflect this. While the hit rate for documents at one Web server might be as high as 85 percent (Kwan, McGrath, and Reed 1995a, 1995b), proxy servers show rates from 10 to 50 percent (Glassman 1994; O'Callaghan 1995; Wessels 1995). Simulations suggest that rates of 60 to 90 percent may be possible if a caching server is optimally configured (Pitkow and Recker 1994; Worrell 1994). Since the hit rate depends on what users request, it is very difficult to predict the future effectiveness of any cache, especially since the Web is an extremely fluid place, constantly changing and expanding. Even a cache that is working very well today may be ineffective in a few weeks if the users become interested in new parts of the Web.

Stale Documents

One way to maximize cache hits is to keep as many documents as possible in the cache as long as possible. This means people will sometimes get out-of-date (stale) information. Documents change at variable and unpredictable times. Studies have shown that 20 to 25 percent of documents change in 7 to 10 days, with 5 percent changing in one day and a few changing constantly (Muntz and Honeyman 1992). The implication of these studies is that if cached copies of a document are kept for one day, one in twenty of them will be out of date. Worst of all, some of the most popular documents, such as weather reports and current news, are interesting primarily *because* they change frequently.

5.3.5 Tuning a Caching Server

There are several parameters that can be adjusted when managing a caching proxy. The details of how to configure the cache vary for each type of server, but each caching server has much the same general parameters to tune.

First, the amount of storage available sets the limit on the cache. The more information that can be cached, the more likely that requests will be satisfied by a cache hit. How many documents can fit in a cache? Most Web documents (HTML text or small images) are fairly small, tens of kilobytes of storage. Many thousands of documents of this size can easily be stored in a practical disk cache. However, multimedia documents featuring many images and sound clips are much larger and are gaining in popularity. And movies (digital video) are even larger. Thus, as Web documents generally become larger, fewer can fit in the same-size cache. Fortunately, it is relatively easy to provide large amounts of storage for a proxy server cache.

The caching server also manages what documents are cached and how the cached copies are kept consistent with the originals. What documents should be cached? The results of scripts should never be cached, nor should documents

Table 5.4