"As a manager who worked under Michael Fisher and Marty Abbott during my time at PayPal/eBay, the opportunity to directly absorb the lessons and experiences presented in this book is invaluable to me now working at Facebook."

—**Yishan Wong,** former CEO, Reddit, and former director of engineering, Facebook

THE ART OF SCALABILITY

Scalable Web Architecture, Processes, and Organizations for the Modern Enterprise

SECOND EDITION

MARTIN L. ABBOTT MICHAEL T. FISHER

FOREWORD BY **MARTY CAGAN,** FOUNDER, SILICON VALLEY PRODUCT GROUP

Praise for *The Art of Scalability, Second Edition*

"A how-to manual for building a world-class engineering organization with step-by-step instructions on everything including leadership, architecture, operations, and processes. A driver's manual for going from 0 to 60, scaling your business. With this book published, there's no excuse for mistakes—in other words, RTFM."

—Lon F. Binder, vice president, technology, Warby Parker

"I've worked with AKF for years on tough technical challenges. Many books address how to correct failing product architectures or problematic processes, both of which are symptoms of an unspoken problem. This book not only covers those symptoms, but also addresses their underlying cause—the way in which we manage, lead, organize, and staff our teams."

—Jeremy King, chief technology officer and senior vice president,
global ecommerce, Walmart.com

"I love this book because it teaches an important lesson most technology-focused books don't: how to build highly scalable and successful technology organizations that build highly scalable technology solutions. There's plenty of great technology coaching in this book, but there are also excellent examples of how to build scalable culture, principles, processes, and decision trees. This book remains one of my few constant go-to reference guides."

—Chris Schremser, chief technology officer, ZirMed

Praise for the First Edition

"This book is much more than you may think it is. Scale is not just about designing Web sites that don't crash when lots of users show up. It is about designing your company so that it doesn't crash when your business needs to grow. These guys have been there on the front lines of some of the most successful Internet companies of our time, and they share the good, the bad, and the ugly about how to not just survive, but thrive."

—Marty Cagan, founder, Silicon Valley Product Group

"A must read for anyone building a Web service for the mass market."

—Dana Stalder, general partner, Matrix Partners

"Abbott and Fisher have deep experiences with scale in both large and small enterprises. What's unique about their approach to scalability is they start by focusing on the true foundation: people and process, without which true scalability cannot be built. Abbott and Fisher leverage their years of experience in a very accessible and practical approach to scalability that has been proven over time with their significant success."

—Geoffrey Weber, vice president of internet operations/IT, Shutterfly

"If I wanted the best diagnoses for my health I would go to the Mayo Clinic. If I wanted the best diagnoses for my portfolio companies' performance and scalability I would call Martin and Michael. They have recommended solutions to performance and scalability issues that have saved some of my companies from a total rewrite of the system."

—Warren M. Weiss, general partner, Foundation Capital

"As a manager who worked under Michael Fisher and Marty Abbott during my time at PayPal/eBay, the opportunity to directly absorb the lessons and experiences presented in this book are invaluable to me now working at Facebook."

—Yishan Wong, former CEO, Reddit, and former director of engineering, Facebook

"*The Art of Scalability* is by far the best book on scalability on the market today. The authors tackle the issues of scalability from processes, to people, to performance, to the highly technical. Whether your organization is just starting out and is defining processes as you go, or you are a mature organization, this is the ideal book

to help you deal with scalability issues before, during, or after an incident. Having built several projects, programs, and companies from small to significant scale, I can honestly say I wish I had this book one, five, and ten years ago."

—Jeremy Wright, chief executive officer, b5media, Inc.

"Only a handful of people in the world have experienced the kind of growth-related challenges that Fisher and Abbott have seen at eBay, PayPal, and the other companies they've helped to build. Fewer still have successfully overcome such challenges. *The Art of Scalability* provides a great summary of lessons learned while scaling two of the largest internet companies in the history of the space, and it's a must-read for any executive at a hyper-growth company. What's more, it's well-written and highly entertaining. I couldn't put it down."

—Kevin Fortuna, partner, AKF Consulting

"Marty and Mike's book covers all the bases, from understanding how to build a scalable organization to the processes and technology necessary to run a highly scalable architecture. They have packed in a ton of great practical solutions from real world experiences. This book is a must-read for anyone having difficulty managing the scale of a hyper-growth company or a startup hoping to achieve hyper growth."

—Tom Keeven, partner, AKF Consulting

"*The Art of Scalability* is remarkable in its wealth of information and clarity; the authors provide novel, practical, and demystifying approaches to identify, predict, and resolve scalability problems before they surface. Marty Abbott and Michael Fisher use their rich experience and vision, providing unique and groundbreaking tools to assist small and hyper-growth organizations as they maneuver in today's demanding technological environments."

—Joseph M. Potenza, attorney, Banner & Witcoff, Ltd.

This page intentionally left blank

The Art of Scalability

Second Edition

This page intentionally left blank

The Art of Scalability

Scalable Web Architecture, Processes, and Organizations for the Modern Enterprise

Second Edition

Martin L. Abbott
Michael T. Fisher

Addison-Wesley
New York • Boston • Indianapolis • San Francisco
Toronto • Montreal • London • Munich • Paris • Madrid
Capetown • Sydney • Tokyo • Singapore • Mexico City

Library of Congress Cataloging-in-Publication Data

Abbott, Martin L.

The art of scalability : scalable web architecture, processes, and organizations for the modern enterprise / Martin L. Abbott, Michael T. Fisher.

pages cm

Includes index.

ISBN 978-0-13-403280-1 (pbk. : alk. paper)

1. Web site development. 2. Computer networks—Scalability. 3. Business enterprises—Computer networks. I. Fisher, Michael T. II. Title.

TK5105.888.A2178 2015

658.4'06—dc23

2015009317

ISBN-13: 978-0-13-403280-1
ISBN-10: 0-13-403280-2
Text printed in the United States on recycled paper at RR Donnelley in Crawfordsville, Indiana.
First printing, June 2015

Editor-in-Chief
Mark L. Taub

Executive Editor
Laura Lewin

Development Editor
Songlin Qiu

Managing Editor
John Fuller

Senior Project Editor
Mary Kesel Wilson

Copy Editor
Jill Hobbs

Indexer
Jack Lewis

Proofreader
Andrea Fox

Technical Reviewers
Roger Andelin
Chris Schremser
Geoffrey Weber

Editorial Assistant
Olivia Basegio

Cover Designer
Chuti Prasertsith

Compositor
The CIP Group

"To my father, for teaching me how to succeed, and to my wife Heather, for teaching me how to have fun."
—Marty Abbott

"To my parents, for their guidance, and to my wife and son, for their unflagging support."
—Michael Fisher

This page intentionally left blank

Contents

Foreword

Perhaps your company began as a brick-and-mortar retailer, or an airline, or a financial services company.

A retailer creates (or buys) technology to coordinate and manage inventory, distribution, billing, and point of sale systems. An airline creates technology to manage the logistics involved in flights, crews, reservations, payment, and fleet maintenance. A financial services company creates technology to manage its customers' assets and investments.

But over the past several years, almost all of these companies, as well as their counterparts in nearly every other industry, have realized that to remain competitive, they need to take their use of technology to an entirely different level—they now need to engage directly with their customers.

Every industry is being reshaped by technology. If they hope to maintain their place as competitive, viable enterprises, companies have no choice but to embrace technology, often in ways that go well beyond their comfort zone.

For example, most retailers now find they need to sell their goods directly to consumers online. Most airlines are trying very hard to entice their customers to purchase their air travel online directly through the airline's site. And nearly all financial services companies work to enable their customers to manage assets and trade directly via their real-time financial sites.

Unfortunately, many of these companies are trying to manage this new customer-facing and customer-enabling technology in the same way they manage their internal technology. The result is that many of these companies have very broken technology and provide terrible customer experiences. Even worse, they don't have the organization, people, or processes in place to improve them.

What companies worldwide are discovering is that there is a very profound difference between utilizing technology to help *run* your company, and leveraging technology to provide your actual products and services directly *for* your customers. It also explains why "technology transformation" initiatives are popping up at so many companies.

This book is all about this necessary transformation. Such a transformation represents a shift in organization, people, process, and especially culture, and scalability is at the center of this transformation.

- Scaling from hundreds of your employees using your technology, to millions of your customers depending on your technology

- Scaling from a small IT cost-center team serving their colleagues in finance and marketing, to a substantial profit-center technology team serving your customers
- More generally, scaling your people, processes, and technology to meet the demands of a modern technology-powered business

But why is technology for your customers so different and so much more difficult to manage than technology for your employees? Several reasons:

- You pay your employees to work at your company and use the technology you tell them they need to use. In contrast, every customer makes his or her own purchase decision—and if she doesn't want it, she won't use it. Your customers must *choose* to use your technology.
- With your own employees, you can get away with requiring training courses, reading manuals, or holding their hands if necessary. In contrast, if your customers can't figure out how to use your technology, they are just a click away from your competitor.
- For internal technology, we measure scale and simultaneous usage in the hundreds of users. For our customers, that scope increases to hundreds of thousands or very often millions of users.
- With internal technology, if a problem arises with the technology, the users are your employees and they are forced to deal with it. For your customers, an issue such as an outage immediately disrupts revenue streams, usually gets the attention of the CEO, and sometimes even draws the notice of the press.
- The harsh truth is that most customer technology simply has a dramatically higher bar set in terms of the definition, design, implementation, testing, deployment, and support than is necessary with most internal technology.

For most companies, establishing a true customer technology competency is *the single most important thing for them to be doing to ensure their survival*, yet remarkably some of them don't even realize they have a problem. They assume that "technology is technology" and the same people who managed their enterprise resource planning implementation shouldn't have too much trouble getting something going online.

If your company is in need of this transformation, then this book is essential reading. It provides a proven blueprint for the necessary change.

Marty and Michael have been there and done that with most of the technology industry's leading companies. I have known and worked with both of these guys for many years. They are not management consultants who could barely launch a

brochure site. They are hands-on leaders who have spent decades in the trenches with their teams creating technology-powered businesses serving hundreds of millions of users and customers. They are the best in the world at what they do, and this new edition is a goldmine of information for any technology organization working to raise its game.

—Marty Cagan
Founder, Silicon Valley Product Group

This page intentionally left blank

Acknowledgments

The authors would like to recognize, first and foremost, the experience and advice of our partner and cofounder Tom Keeven. The process and technology portions of this book were built over time with the help of Tom and his many years of experience. Tom started the business that became AKF Partners. We often joke that Tom has forgotten more about architecting highly available and scalable sites than most of us will ever learn.

We would also like to thank several AKF team members—Geoff Kershner, Dave Berardi, Mike Paylor, Kirk Sanford, Steve Mason, and Alex Hooper—who contributed their combined decades of experience and knowledge not only to this second edition, but also to AKF Partners' consulting practice. Without their help putting the concepts from the first edition into practice and helping to mature them over time, this second edition would not be possible.

Additionally, the authors owe a great debt of gratitude to this edition's technical reviewers—Geoffrey Weber, Chris Schremser, and Roger Andelin. All three of these individuals are experienced technology executives who have decades of hands-on experience designing, developing, implementing, and supporting large-scale systems in industries ranging from ecommerce to health care. They willingly agreed to accept our poorly written drafts and help turn them into easily consumable prose for the benefit of our readers.

This edition would not be possible without the support provided by the team at Addison-Wesley, including executive editor Laura Lewin, development editor Songlin Qiu, and editorial assistant Olivia Basegio. Laura quickly became the champion for a second edition after discussing the significant changes with regard to scaling systems and organizations that have occurred over the five years since the first edition was published. Songlin has been an invaluable partner in ensuring both the first and second editions of *The Art of Scalability* were consistent, clear, and correct. Olivia has saved us multiple times when technical challenges threatened to delay or derail us.

We further would like to recognize our colleagues and teams at Quigo, eBay, and PayPal. These are the companies at which we really started to build and test many of the approaches mentioned in the technology and process sections of this book. The list of names within these teams is quite large, but the individuals know who they are.

Finally, we'd like to acknowledge the U.S. Army and United States Military Academy. Together they created a leadership lab unlike any other we can imagine.

Multiple reviewers have reviewed this book as we have attempted to provide the best possible work for the reader. However, in a work this large, errors will inevitably occur. All errors in the text are completely the authors' fault.

This page intentionally left blank

About the Authors

Martin L. Abbott is a founding partner at the growth and scalability advisory firm AKF Partners. He was formerly chief operations officer at Quigo, an advertising technology startup sold to AOL, where he was responsible for product strategy, product management, technology development, and client services. Marty spent nearly six years at eBay, most recently as senior vice president of technology, chief technology officer, and member of the executive staff. Prior to his time at eBay, Marty held domestic and international engineering, management, and executive positions at Gateway and Motorola. He has served on the boards of directors of several private and public companies. Marty has a B.S. in computer science from the United States Military Academy, has an M.S. in computer engineering from the University of Florida, is a graduate of the Harvard Business School Executive Education Program, and has a Doctor of Management from Case Western Reserve University.

Michael T. Fisher is a founding partner at the growth and scalability advisory firm AKF Partners. Prior to cofounding AKF Partners, Michael was the chief technology officer at Quigo, a startup Internet advertising company that was acquired by AOL in 2007. Before his time at Quigo, Michael served as vice president, engineering and architecture, for PayPal, Inc., an eBay company. Prior to joining PayPal, he spent seven years at General Electric helping to develop the company's technology strategy and was a Six Sigma Master Black Belt. Michael served six years as a Captain and pilot in the U.S. Army. He received a Ph.D. and an MBA from Case Western Reserve University's Weatherhead School of Management, an M.S. in information systems from Hawaii-Pacific University, and a B.S. in computer science from the United States Military Academy (West Point). Michael is an adjunct professor in the design and innovation department at Case Western Reserve University's Weatherhead School of Management.

This page intentionally left blank

Introduction

Thanks for picking up the second edition of *The Art of Scalability: Scalable Web Architecture, Processes, and Organizations for the Modern Enterprise*. This book has been recognized by academics and professionals as one of the best resources available to learn the art of scaling systems and organizations. This second edition includes new content, revisions, and updates. As consultants and advisors to hundreds of hyper-growth companies, we have been fortunate enough to be on the forefront of many industry changes, including new technologies and new approaches to implementing products. While we hope our clients see value in our knowledge and experience, we are not ignorant of the fact that a large part of the value we bring to bear on a subject comes from our interactions with so many other technology companies. In this edition, we share even more of these lessons learned from our consulting practice.

In this second edition, we have added several key topics that we believe are critical to address in a book on scalability. One of the most important new topics focuses on a new organizational structure that we refer to as the Agile Organization. Other notable topics include the changing rationale for moving from data centers to clouds (IaaS/PaaS), why NoSQL solutions aren't in and of themselves a panacea for scaling, and the importance of business metrics to the health of the overall system.

In the first edition of *The Art of Scalability*, we used a fictional company called AllScale to demonstrate many of the concepts. This fictional company was an aggregation of many of our clients and the challenges they faced in the real world. While AllScale provided value in highlighting the key points in the first edition, we believe that real stories make more of an impact with readers. As such, we've replaced AllScale with real-world stories of successes and failures in the current edition.

The information contained in this book has been carefully designed to be appropriate for any employee, manager, or executive of an organization or company that provides technology solutions. For the nontechnical executive or product manager, this book can help you prevent scalability disasters by arming you with the tools needed to ask the right questions and focus on the right areas. For technologists and engineers, this book provides models and approaches that, once employed, will help you scale your products, processes, and organizations.

Our experience with scalability goes beyond academic study and research. Although we are both formally trained as engineers, we don't believe academic

programs teach scalability very well. Rather, we have learned about scalability by suffering through the challenges of scaling systems for a combined 30-plus years. We have been engineers, managers, executives, and advisors for startups as well as *Fortune 500* companies. The list of companies that our firm or we as individuals have worked with includes such familiar names as General Electric, Motorola, Gateway, eBay, Intuit, Salesforce, Apple, Dell, Walmart, Visa, ServiceNow, DreamWorks Animation, LinkedIn, Carbonite, Shutterfly, and PayPal. The list also includes hundreds of less famous startups that need to be able to scale as they grow. Having learned the scalability lessons through thousands of hours spent diagnosing problems and thousands more hours spent designing preventions for those problems, we want to share our combined knowledge. This motivation was the driving force behind our decisions to start our consulting practice, AKF Partners, in 2007, and to write the first edition of this book, and it remains our preeminent goal in this second edition.

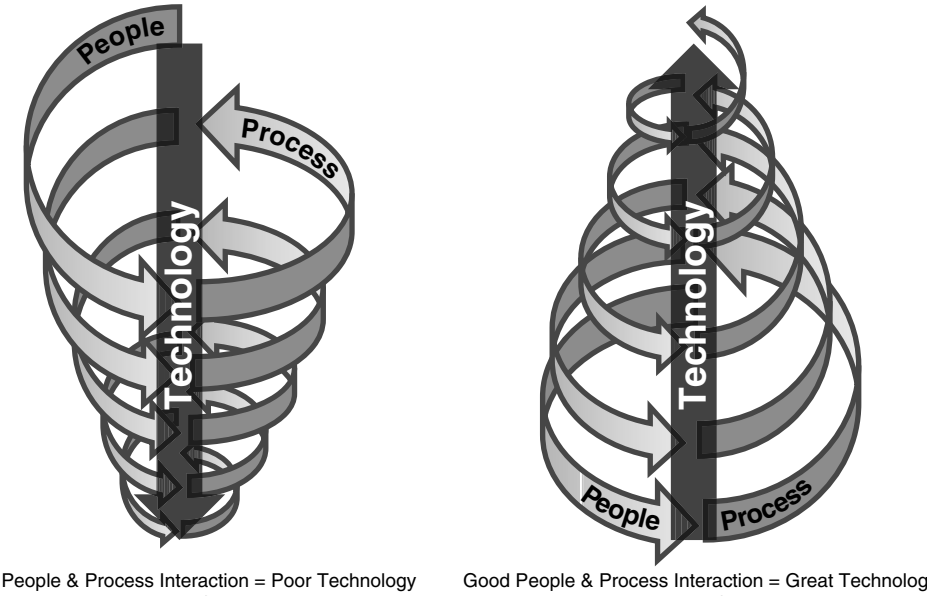
Scalability: So Much More Than Just Technology

Pilots are taught, and statistics show, that many aircraft incidents are the result of multiple failures that snowball into total system failure and catastrophe. In aviation, these multiple failures, which are called an error chain, often start with human rather than mechanical failure. In fact, Boeing identified that 55% of all aircraft incidents involving Boeing aircraft between 1995 and 2005 had human factors–related causes.[1]

Our experience with scalability-related issues follows a similar trend. The chief technology officer (CTO) or executive responsible for scale of a technology platform may see scalability as purely a technical endeavor. This perception is the first, and very human, failure in the error chain. Because the CTO is overly technology focused, she fails to define the processes necessary to identify scalability bottlenecks—failure number two. Because no one is identifying bottlenecks or chokepoints in the architecture, the user count or transaction volume exceeds a certain threshold and the entire product fails—failure number three. The team assembles to solve the problem, but because it has never invested in processes to troubleshoot incidents and their related problems, the team misdiagnoses the failure as "the database needs to be tuned"—failure number four. The vicious cycle goes on for days, with people focusing on different pieces of the technology stack and blaming everything from

firewalls, to applications, to the persistence tiers to which the apps speak. Team interactions devolve into shouting matches and finger-pointing sessions, while services remain slow and unresponsive. Customers walk away, team morale flat-lines, and shareholders are left holding the bag.

The key point here is that crises resulting from an inability to scale to end-user demands are almost never technology problems alone. In our experience as former executives and advisors to our clients, scalability issues start with organizations and people, and only then spread to process and technology. People, being human, make ill-informed or poor choices regarding technical implementations, which in turn sometimes manifest themselves as a failure of a technology platform to scale. People also ignore the development of processes that might help them learn from past mistakes and sometimes put overly burdensome processes in place, which in turn might force the organization to make poor decisions or make decisions too late to be effective. A lack of attention to the people and processes that create and support technical decision making can lead to a vicious cycle of bad technical decisions, as depicted in the left side of Figure I.1. This book is the first of its kind focused on creating a virtuous cycle of people and process scalability to support better, faster, and more scalable technology decisions, as depicted in the right side of Figure I.1.



Bad People & Process Interaction = Poor Technology
Vicious Cycle

Good People & Process Interaction = Great Technology
Virtuous Cycle

Figure I.1 *Vicious and Virtuous Technology Cycles Utility*

Art Versus Science

Our choice of the word *art* in the title of this book is a deliberate one. *Art* conjures up images of a fluid nature, whereas *science* seems much more structured and static. It is this image that we heavily rely on, as our experience has taught us that there is no single approach or way to guarantee an appropriate level of scale within a platform, organization, or process. A successful approach to scaling must be crafted around the ecosystem created by the intersection of the current technology platform, the characteristics of the organization, and the maturity and appropriateness of the existing processes. This book focuses on providing skills and teaching approaches that, if employed properly, will help solve nearly any scalability or availability problem.

This is not to say that we don't advocate the application of the scientific method in nearly any approach, because we absolutely do. *Art* here is a nod to the notion that you simply cannot take a "one size fits all" approach to any potential system and expect to meet with success.

Who Needs Scalability?

Any company that continues to grow ultimately will need to figure out how to scale its systems, organizations, and processes. Although we focus on Web-centric products through much of this book, we do so only because the most unprecedented growth has been experienced by Internet companies such as Google, Yahoo, eBay, Amazon, Facebook, LinkedIn, and the like. Nevertheless, many other companies experienced problems resulting from an inability to scale to new demands (a lack of scalability) long before the Internet came of age. Scale issues have governed the growth of companies from airlines and defense contractors to banks and colocation facility (data center) providers. We guarantee that scalability was on the mind of every bank manager during the consolidation that occurred after the collapse of the banking industry.

The models and approaches that we present in our book are industry agnostic. They have been developed, tested, and proven successful in some of the fastest-growing companies of our time; they work not only in front-end customer-facing transaction-processing systems, but also in back-end business intelligence, enterprise resource planning, and customer relationship management systems. They don't discriminate by activity, but rather help to guide the thought process on how to separate systems, organizations, and processes to meet the objective of becoming highly scalable and reaching a level of scale that allows the business to operate without concerns about its ability to meet customer or end-user demands.

Book Organization and Structure

We've divided the book into five parts. Part I, "Staffing a Scalable Organization," focuses on organization, management, and leadership. Far too often, managers and leaders are promoted based on their talents within their area of expertise. Engineering leaders and managers, for example, are very often promoted based on their technical acumen and aren't given the time or resources needed to develop their business, management, and leadership acumen. Although they might perform well in the architectural and technical aspects of scale, their expertise in organizational scale needs is often shallow or nonexistent. Our intent is to provide these managers and leaders with a foundation from which they can grow and prosper as managers and leaders.

Part II, "Building Processes for Scale," focuses on the processes that help hypergrowth companies scale their technical platforms. We cover topics ranging from technical issue resolution to crisis management. We also discuss processes meant for governing architectural decisions and principles to help companies scale their platforms.

Part III, "Architecting Scalable Solutions," focuses on the technical and architectural aspects of scale. We introduce proprietary models developed within AKF Partners, our consulting and advisory practice. These models are intended to help organizations think through their scalability needs and alternatives.

Part IV, "Solving Other Issues and Challenges," discusses emerging technologies such as grid computing and cloud computing. We also address some unique problems within hyper-growth companies such as the immense growth and cost of data as well as issues to consider when planning data centers and evolving monitoring strategies to be closer to customers.

Part V, "Appendices," explains how to calculate some of the most common scalability numbers. Its coverage includes the calculation of availability, capacity planning, and load and performance.

The lessons in this book have not been designed in the laboratory, nor are they based on unapplied theory. Rather, these lessons have been designed and implemented by engineers, technology leaders, and organizations through years of struggling to keep their dreams, businesses, and systems afloat. The authors have had the great fortune to be a small part of many of these teams in many different roles—sometimes as active participants, at other times as observers. We have seen how putting these lessons into practice has yielded success—and how the unwillingness or inability to do so has led to failure. This book aims to teach you these lessons and put you and your team on the road to success. We believe the lessons here are valuable for everyone from engineering staffs to product staffs, including every level from the individual contributor to the CEO.

This page intentionally left blank

This page intentionally left blank

Index

C

D

E

F

G

H

I

M

N

O

P

Q

R

T

Y

Z