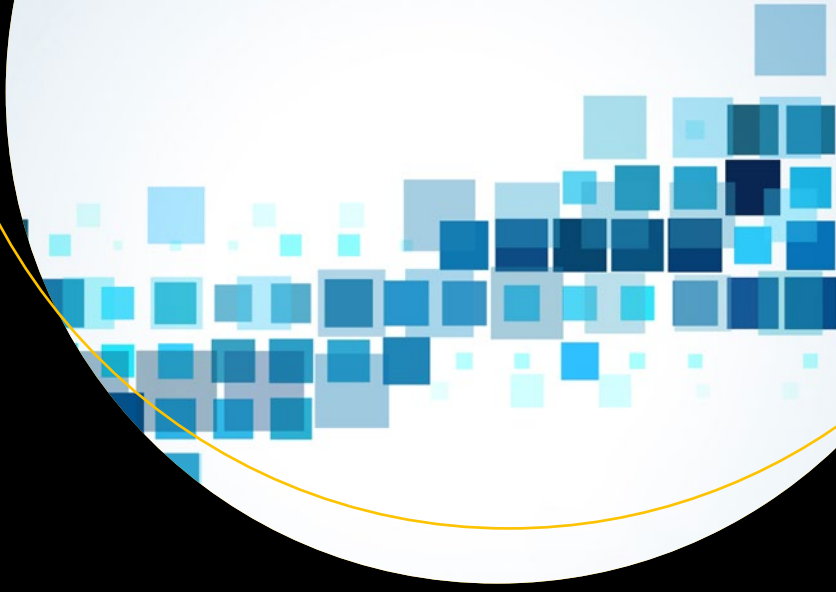

Beginning Apache Pig

Big Data Processing Made Easy

Balaswamy Vaddeman

Apress®

Beginning Apache Pig

Big Data Processing Made Easy

Balaswamy Vaddeman

Apress®

Beginning Apache Pig: Big Data Processing Made Easy

Balaswamy Vaddeman
Hyderabad, Andhra Pradesh, India

ISBN-13 (pbk): 978-1-4842-2336-9 ISBN-13 (electronic): 978-1-4842-2337-6

DOI 10.1007/978-1-4842-2337-6

Library of Congress Control Number: 2016961514

Managing Director: Welmoed Spahr
Lead Editor: Celestin Suresh John
Technical Reviewer: Manoj R. Patil
Editorial Board: Steve Anglin, Pramila Balan, Laura Berendson, Aaron Black, Louise Corrigan, Jonathan Gennick, Robert Hutchinson, Celestin Suresh John, Nikhil Karkal, James Markham, Susan McDermott, Matthew Moodie, Natalie Pao, Gwenan Spearing
Coordinating Editor: Prachi Mehta
Copy Editor: Kim Wimpsett
Compositor: SPi Global
Indexer: SPi Global
Artist: SPi Global

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at www.apress.com/bulk-sales.

Any source code or other supplementary materials referenced by the author in this text are available to readers at www.apress.com. For detailed information about how to locate your book's source code, go to www.apress.com/source-code/. Readers can also access source code at SpringerLink in the Supplementary Material section for each chapter.

Printed on acid-free paper

The six most important people in my life:

The late Kammari Rangaswamy (Teacher)

The late Niranjanamma (Mother)

Devaiah (Father)

Radha (Wife)

Sai Nirupam (Son)

Nitya Maithreyi (Daughter)

Contents at a Glance

Contents

About the Author

Balaswamy Vaddeman is a thinker, blogger, and serious and self-motivated big data evangelist with 10 years of experience in IT and 5 years of experience in the big data space. His big data experience covers multiple areas such as analytical applications, product development, consulting, training, book reviews, hackathons, and mentoring. He has proven himself while delivering analytical applications in the retail, banking, and finance domains in three aspects (development, administration, and architecture) of Hadoop-related technologies. At a startup company, he developed a Hadoop-based product that was used for delivering analytical applications without writing code.

In 2013 Balaswamy won the Hadoop Hackathon event for Hyderabad conducted by Cloudwick Technologies. Being the top contributor at Stackoverflow.com, he helped countless people on big data topics at multiple web sites such as Stackoverflow.com and Quora.com. With so much passion on big data, he became an independent trainer and consultant so he could train hundreds of people and set up big data teams in several companies.

About the Technical Reviewer

Manoj R. Patil is a big data architect at TatvaSoft, an IT services and consulting firm. He has a bachelor's of engineering degree from COEP in Pune, India. He is a proven and highly skilled business intelligence professional with 17 years of information technology experience. He is a seasoned BI and big data consultant with exposure to all the leading platforms such as Java EE, .NET, LAMP, and so on. In addition to authoring a book on Pentaho and big data, he believes in knowledge sharing, keeps himself busy in corporate training, and is a passionate teacher. He can be reached at on Twitter `@manojrpatil` and at `https://in.linkedin.com/in/manojrpatil` on LinkedIn.

Manoj would like to thank his family, especially his two beautiful daughters, Ayushee and Ananyaa, for their patience during the review process.