# Table of Contents

Foreword ..................................................................................................................................................... xviii  
Preface ........................................................................................................................................................ xx  
Acknowledgment ........................................................................................................................................ xxviii  

## Section 1  
**Big Data Technologies, Methods, and Algorithms**

### Chapter 1  
Technologies for Big Data .......................................................................................................................... 1  
  *Kapil Bakshi, Cisco Systems Inc., USA*

### Chapter 2  
Applying the K-Means Algorithm in Big Raw Data Sets with Hadoop and MapReduce ......................... 23  
  *Ilias K. Savvas, TEI of Larissa, Greece*  
  *Georgia N. Sofianidou, TEI of Larissa, Greece*  
  *M-Tahar Kechadi, University College Dublin, Ireland*

### Chapter 3  
Synchronizing Execution of Big Data in Distributed and Parallelized Environments ............................. 47  
  *Gueyoung Jung, Xerox Research Center Webster, USA*  
  *Tridib Mukherjee, Xerox Research Center India, India*

### Chapter 4  
Parallel Data Reduction Techniques for Big Datasets ................................................................................ 72  
  *Ahmet Artu Yildirim, Utah State University, USA*  
  *Cem Özdogan, Çankaya University, Turkey*  
  *Dan Watson, Utah State University, USA*
Section 2
Big Data Storage, Management, and Sharing

Chapter 5
Techniques for Sampling Online Text-Based Data Sets .......................................................... 95
Lynne M. Webb, University of Arkansas, USA
Yuanxin Wang, Temple University, USA

Chapter 6
Big Data Warehouse Automatic Design Methodology ............................................................. 115
Francesco Di Tria, Università degli Studi di Bari Aldo Moro, Italy
Ezio Lefons, Università degli Studi di Bari Aldo Moro, Italy
Filippo Tangorra, Università degli Studi di Bari Aldo Moro, Italy

Chapter 7
Big Data Management in the Context of Real-Time Data Warehousing ................................ 150
M. Asif Naeem, Auckland University of Technology, New Zealand
Gillian Dobbie, University of Auckland, New Zealand
Gerald Weber, University of Auckland, New Zealand

Chapter 8
Big Data Sharing Among Academics ...................................................................................... 177
Jeonghyun Kim, University of North Texas, USA

Section 3
Specific Big Data

Chapter 9
Scalable Data Mining, Archiving, and Big Data Management for the Next Generation
Astronomical Telescopes ........................................................................................................ 196
Chris A. Mattmann, California Institute of Technology, USA
Andrew Hart, California Institute of Technology, USA
Luca Cinquini, California Institute of Technology, USA
Joseph Lazio, California Institute of Technology, USA
Shakeh Khudikyan, California Institute of Technology, USA
Dayton Jones, California Institute of Technology, USA
Robert Preston, California Institute of Technology, USA
Thomas Bennett, SKA South Africa Project, South Africa
Bryan Butler, National Radio Astronomy Observatory (NRAO), USA
David Harland, National Radio Astronomy Observatory (NRAO), USA
Brian Glendenning, National Radio Astronomy Observatory (NRAO), USA
Jeff Kern, National Radio Astronomy Observatory (NRAO), USA
James Robnett, National Radio Astronomy Observatory (NRAO), USA
Chapter 10
Efficient Metaheuristic Approaches for Exploration of Online Social Networks ................................. 222
  Zorica Stanimirović, University of Belgrade, Serbia
  Stefan Mišković, University of Belgrade, Serbia

Chapter 11
Big Data at Scale for Digital Humanities: An Architecture for the HathiTrust Research Center....... 270
  Stacy T. Kowalczyk, Dominican University, USA
  Yiming Sun, Indiana University, USA
  Zong Peng, Indiana University, USA
  Beth Plale, Indiana University, USA
  Aaron Todd, Indiana University, USA
  Loretta Auvil, University of Illinois, USA
  Craig Willis, University of Illinois, USA
  Jiaan Zeng, Indiana University, USA
  Milinda Pathirage, Indiana University, USA
  Samitha Liyanage, Indiana University, USA
  Guangchen Ruan, Indiana University, USA
  J. Stephen Downie, University of Illinois, USA

Chapter 12
GeoBase: Indexing NetCDF Files for Large-Scale Data Analysis ........................................................... 295
  Tanu Malik, University of Chicago, USA

Chapter 13
Large-Scale Sensor Network Analysis: Applications in Structural Health Monitoring .................... 314
  Joaquin Vanschoren, University of Leiden, The Netherlands
  Ugo Vesperi, University of Leiden, The Netherlands
  Shengfa Miao, University of Leiden, The Netherlands
  Marvin Meeng, University of Leiden, The Netherlands
  Ricardo Cachucho, University of Leiden, The Netherlands
  Arno Knobbe, University of Leiden, The Netherlands

Section 4
Big Data and Computer Systems and Big Data Benchmarks

Chapter 14
Accelerating Large-Scale Genome-Wide Association Studies with Graphics Processors .................... 349
  Mian Lu, Institute of High Performance Computing, A*STAR, Singapore
  Qiong Luo, Hong Kong University of Science and Technology, Hong Kong
Chapter 15
The Need to Consider Hardware Selection when Designing Big Data Applications Supported by Metadata
Nathan Regola, University of Notre Dame, USA
David A. Cieslak, Aunalytics, Inc., USA
Nitesh V. Chawla, University of Notre Dame, USA

Chapter 16
Excess Entropy in Computer Systems
Charles Loboz, Microsoft Corporation, USA

Chapter 17
A Review of System Benchmark Standards and a Look Ahead Towards an Industry Standard for Benchmarking Big Data Workloads
Raghunath Nambiar, Cisco Systems, Inc., USA
Meikel Poess, Oracle Corp., USA

Compilation of References

About the Contributors

Index