SpringerBriefs in Computer Science

For further volumes:
http://www.springer.com/series/10028

Min Chen • Shiwen Mao • Yin Zhang
Victor C.M. Leung

Big Data

Related Technologies, Challenges and Future Prospects

Springer

Min Chen
School of Computer Science
and Technology
Huazhong University of Science
and Technology
Wuhan, China

Yin Zhang
School of Computer Science
and Technology
Huazhong University of Science
and Technology
Wuhan, China

Shiwen Mao
Auburn University
Auburn, AL, USA

Victor C.M. Leung
Electrical and Computer Engineering
The University of British Columbia
Vancouver, BC
Canada

ISSN 2191-5768
ISSN 2191-5776 (electronic)
ISBN 978-3-319-06244-0
ISBN 978-3-319-06245-7 (eBook)
DOI 10.1007/978-3-319-06245-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014937319

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

"How big is big?" Science writer Stephen Strauss asks in his fun book for kids titled *How Big is Big* and explains that "bigness is something no one can consume."

In this book, we aim to answer this interesting question, but in the context of computer data. In the *big data* era, we are dealing with the explosive increase of global data and enormous datasets. Unlike seemingly similar terms such as "massive data" or "very big data," *big data* refers to the datasets that could not be perceived, acquired, managed, and processed by traditional Information Technology (IT) and software/hardware tools within a tolerable time. It can be characterized by four Vs, i.e., Volume (great volume), Variety (various modalities), Velocity (rapid generation), and Value (huge value but very low density).

In this book, we provide a comprehensive overview of the background and related technologies, challenges and future prospects of big data. We first introduce the general background of big data and review related technologies, such as cloud computing, Internet of Things (IoT), data centers, and Hadoop. We then focus on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. For each phase, we introduce the general background, discuss the technical challenges, and review the latest advances. We next examine the several representative applications of big data, including enterprise management, IoT, online social networks, healthcare and medical applications, collective intelligence, and smart grid. This book is concluded with a discussion of open problems and future directions. We aim to provide the readers a comprehensive overview and big-picture of this exciting area. We hope this monograph could be a useful reference for graduate students and professionals in related fields, and general readers who will benefit from an understanding of the big data field.

We are grateful to Dr. Xuemin (Sherman) Shen, the SpringerBriefs Series Editor on Wireless Communications. This book would not be possible without his kind support during the process. Thanks also to the Springer Editors and Staff, all of whom did their usual excellent job in getting this monograph published.

This work was supported by China National Natural Science Foundation (No. 61300224), the Ministry of Science and Technology (MOST), China, the International Science and Technology Collaboration Program (Project No.:

2014DFT10070), and the Hubei Provincial Key Project (No. 2013CFA051). Shiwen Mao's research is supported in part by the US National Science Foundation (NSF) under Grants CNS-1320664, CNS-1247955, CNS-0953513, and DUE-1044021, and through the NSF Broadband Wireless Access & Applications Center (BWAC) Site at Auburn University (NSF Grant IIP-1266036). The research of Victor Leung is supported by the Canadian Natural Sciences and Engineering Research Council, BC Innovation Council, Qatar Research Foundation, TELUS, and other industrial partners. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the foundation.

Wuhan, China — Min Chen
Auburn, AL — Shiwen Mao
Wuhan, China — Yin Zhang
Vancouver, BC, Canada — Victor C.M. Leung
January 2014

Contents

Acronyms

AMI	Advanced Metering Infrastructure
APT	Advanced Persistent Threat
BI	Business Intelligence
BLOB	Binary Large Object or Basic Large Object
BPM	Business Process Management
BSON	Binary JSON
CEO	Chief Executive Officers
CIO	Chief Information Officer
DAS	Direct Attached Storage
DMA	Direct Memory Access
ETL	Extract, Transform and Load
ERCIM	European Research Consortium for Informatics and Mathematics
GUI	Graphic User Interface
HDFS	Hadoop Distributed File System
HGP	Human Genome Project
HQL	HyperTable Query Language
ICT	Information and Communications Technology
IDC	International Data Corporation
IoT	Internet of Things
IT	Information Technology
LHC	Large Hadron Collider
Libpcap	Packet Capture Library
MMF	Multi-Mode Fiber
MPI	Message Passing Interface
MR	MapReduce
MVCC	Muti-Version Concurrency Control
NAS	Network Attached Storage
NER	Named Entity Recognition
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NSF	National Science Foundation

OFDM	Orthogonal Frequency-Division Multiplexing
OLAP	On-Line Analytical Processing
OpenMP	Open Multi-Processing
PB	Petabyte
PMU	Phasor Measurement Unit
PNUTS	Platform for Nimble Universal Table Storage
RAID	Redundant Array of Independent Disks
RDBMS	Relational Database Management System
SAN	Storage Area Network
SDK	Software Development Kit
SDSS	Sloan Digital Sky Survey
SNS	Social Networking Services
SSD	Solid-State Drive
TB	Terabyte
TOMS	Topic-oriented Multimedia Summarization System
TOR	Top Rack Switches
URL	Uniform Resource Locator
WDM	Wavelength Division Multiplexing
ZC	Zero-copy