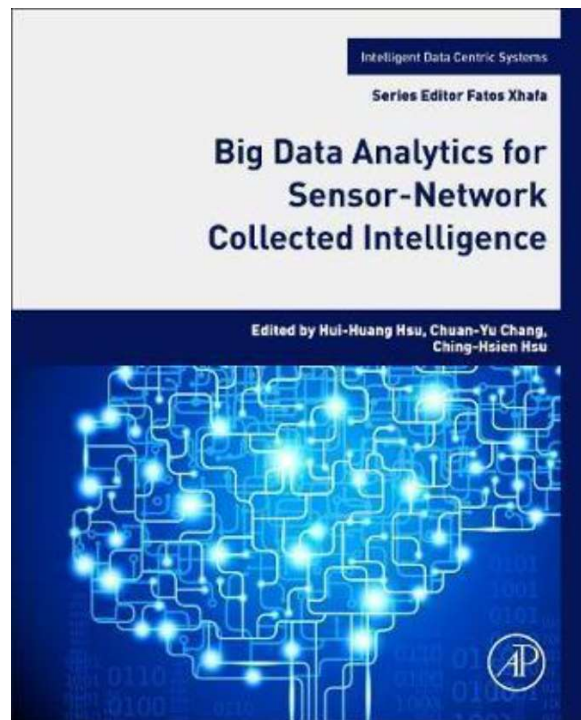




Academic Press is an imprint of
125 London Wall, London EC2Y 5DB
525 B Street, Suite 1800,
50 Hampshire Street, 5th
The Boulevard, Langford

© 2017 Elsevier Inc. All

No part of this publication
mechanical, including photocopying,
permission in writing from
the Publisher's permission
Center and the Copyright

This book and the individ

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1800, San Diego, CA 92101-4495, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

© 2017 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

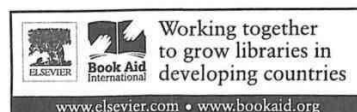
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-12-809393-1

For information on all Academic Press publications
visit our website at <https://www.elsevier.com/books-and-journals>



Contents

List of Contributors	xii
Preface	xvii

PART I BIG DATA ARCHITECTURE AND PLATFORMS

CHAPTER 1 Big Data: A Classification of Acquisition and Generation Methods	3
1 Big Data: A Classification	4
1.1 Characteristics of Big Data	4
2 Big Data Generation Methods	5
2.1 Data Sources	5
2.2 Data Types	6
3 Big Data: Data Acquisition Methods	8
3.1 Interface Methods	8
3.2 Interface Devices	11
4 Big Data: Data Management	12
4.1 Data Representation and Organization	12
4.2 Databases	16
4.3 Data Fusion and Data Integration	18
5 Summary	18
References	19
Glossary	20
 CHAPTER 2 Cloud Computing Infrastructure for Data Intensive Applications.....	 21
1 Introduction	22
2 Big Data Nature and Definition	23
2.1 Big Data in Science and Industry	23
2.2 Big Data and Social Network/Data	24
2.3 Big Data Technology Definition: From 6V to 5 Parts	24
3 Big Data and Paradigm Change	25
3.1 Big Data Ecosystem	25
3.2 New Features of the BDI	25
3.3 Moving to Data-Centric Models and Technologies	26
4 Big Data Architecture Framework and Components	27
4.1 Defining the Big Data Architecture Framework	27

vi Contents

4.4 NIST Big Data Reference Architecture	30
4.5 General Big Data System Requirements	33
5 Big Data Infrastructure	34
5.1 BDI Components	34
5.2 Big Data Stack Components and Technologies	36
5.3 Example of Cloud-Based Infrastructure for Distributed Data Processing	37
5.4 Benefits of Cloud Platforms for Big Data Applications	39
6 Case Study: Bioinformatics Applications Deployment on Cloud	39
6.1 Overall Description	39
6.2 UC1—Securing Human Biomedical Data	40
6.3 UC2—Cloud Virtual Pipeline for Microbial Genomes Analysis	41
6.4 Implementation of Use Cases and CYCLONE Infrastructure Components	42
7 CYCLONE Platform for Cloud Applications Deployment and Management	44
7.1 General Architecture for Intercloud and Multicloud	44
7.2 Ensuring Consistent Security Services in Cloud-Based Applications	45
7.3 Dynamic Access Control Infrastructure	46
8 Cloud Powered Big Data Applications Development and Deployment Automation	47
8.1 Demand for Automated Big Data Applications Provisioning	47
8.2 Cloud Automation Tools for Intercloud Application and Network Infrastructure Provisioning	49
8.3 SlipStream: Cloud Application Management Platform	49
9 Big Data Service and Platform Providers	51
9.1 Amazon Web Services and Elastic MapReduce	52
9.2 Microsoft Azure Analytics Platform System and HDInsight	52
9.3 IBM Big Data Analytics and Information Management	54
9.4 Cloudera	54
9.5 Pentaho	54
9.6 LexisNexis HPCC Systems as an Integrated Open Source Platform for Big Data Analytics	55
10 Conclusion	57
Acknowledgments	57
References	58
Glossary	60

2	On-Premise Private Clouds for Big Data	66
2.1	Security of Cloud Computing Systems.....	66
2.2	Advantages of On-Premise Private Clouds	66
3	Introduction to Selected Open Source Cloud Environments	67
3.1	OpenNebula	67
3.2	Eucalyptus	69
3.3	Apache CloudStack	70
3.4	OpenStack.....	71
4	Heterogeneous Computing in the Cloud.....	75
4.1	Exclusive Mode.....	75
4.2	Sharing Mode	75
5	Case Study: The EMS, an On-Premise Private Cloud	76
6	Conclusion.....	78
	Disclaimer.....	78
	References	79

PART II BIG DATA PROCESSING AND MANAGEMENT

CHAPTER 4	Efficient Nonlinear Regression-Based Compression of Big Sensing Data on Cloud.....	83
1	Introduction	83
1.1	Motivation	84
1.2	Organization of the Chapter.....	85
2	Related Work and Problem Analysis.....	86
2.1	Related Work.....	86
2.2	Problem Analysis: Real-World Requirements for Nonlinear Regression.....	87
3	Temporal Compression Model Based on Nonlinear Regression	88
3.1	Nonlinear Regression Prediction Model.....	89
4	Algorithms	91
4.1	Algorithm for Nonlinear Regression.....	91
4.2	Nonlinear Regression Compression Algorithm Based on MapReduce	92
5	Experiments	94
5.1	Experiment Environment and Process	94
5.2	Experiment for the Compression With Nonlinear Regression.....	95
5.3	Experiment for Data Loss and Accuracy.....	96

viii Contents

CHAPTER 5	Big Data Management on Wireless Sensor Networks.....	99
1	Introduction	99
2	Data Management on WSNs.....	101
2.1	Storage	102
2.2	Query Processing.....	102
2.3	Data Collection.....	105
3	Big Data Tools	106
3.1	File System	106
3.2	Batch Processing	107
3.3	Streaming Data Processing	108
4	Put It Together: Big Data Management Architecture	109
4.1	Batch Layer	109
4.2	Serving Layer	110
4.3	Speed Layer	110
5	Big Data Management on WSNs.....	111
5.1	In-Network Aggregation Techniques and Data Integration Components	111
5.2	Exploiting Big Data Systems as Data Centers	112
6	Conclusion	114
	References	114
	Glossary	116
 CHAPTER 6	 Extreme Learning Machine and Its Applications in Big Data Processing.....	 117
1	Introduction	117
1.1	Background.....	117
1.2	Artificial Neural Networks.....	117
1.3	Era of Big Data	12
1.4	Organization	12
2	Extreme Learning Machine.....	12
2.1	Traditional Approaches to Train ANNs.....	12
2.2	Theories of the Extreme Learning Machine	12
2.3	Classical ELM	12
2.4	ELM for Classification and Regression.....	12
2.5	ELM for Unsupervised Learning	12
3	Improved Extreme Learning Machine With Big Data.....	12
3.1	Shortcomings of the Extreme Learning Machine for Processing Big Data	12

4 Applications.....	144
4.1 ELM in Predicting Protein Structure.....	144
4.2 ELM in Image Processing.....	145
4.3 ELM in Cancer Diagnosis.....	145
4.4 ELM in Big Data Security and Privacy.....	146
5 Conclusion.....	146
References.....	146
Glossary.....	150

PART III BIG DATA ANALYTICS AND SERVICES

CHAPTER 7 Spatial Big Data Analytics for Cellular Communication Systems.....	153
1 Introduction.....	153
2 Cellular Communications and Generated Data.....	154
3 Spatial Big Data Analytics.....	155
3.1 Statistical Foundation for Spatial Big Data Analytics.....	155
3.2 Spatial Pattern Mining From Spatial Big Data Analytics.....	156
4 Typical Applications.....	161
4.1 BS Behavior Understanding Through Spatial Big Data Analytics.....	161
4.2 User Behavior Understanding Through Spatial Big Data Analytics.....	163
5 Conclusion and Future Challenging Issues.....	164
Acknowledgments.....	164
References.....	164
Glossary.....	166
 CHAPTER 8 Cognitive Applications and Their Supporting Architecture for Smart Cities.....	 167
1 Introduction.....	167
2 CSE for Smart City Applications.....	168
2.1 Architecture Specification.....	169
2.2 Big Data Analysis and Management.....	171
3 Anomaly Detection in Smart City Management.....	171
3.1 Related Work to Anomaly Detection.....	172
3.2 Challenges and Benefits of Anomaly Detection in Smart Cities.....	173
4 Functional Region and Socio-Demographic Regional Patterns Detection in Cities.....	174

4.1 Discovering Functional Regions..... 174

4.2 Deep Learning and Regional Pattern Detections..... 176

5 Summary..... 181

References 181

Glossary 184