

**A New Book from Wiley Publisher to appear in late 2016 or early 2017**

# **Big-Data Computing with Smart Clouds and IoT Sensing**

**Kai Hwang**, University of Southern California, USA

**Min Chen**, Huazhong University of Science and Technology, China

## **Motivations and Objectives:**

The information industry is changing rapidly in recent years. We are crossing into the era of big data and machine intelligence. This trend is triggered by the widespread use of storage and computing clouds and by pervasive deployment of Internet of Things (IoT) platforms. To face the changes, we must be able to handle the discovery, storage, processing and application of big data in social networks, mobile systems, business, health-care, and scientific domains.

These new challenges demand extensive virtualization support and machine intelligence. For example, major companies like Amazon, Apple, Google, IBM, and Microsoft are all expanding their datacenters into cloud facilities and services. New applications are triggered by big data and IoT sensing in smart cities, autonomous car driving, critical decision making, and cognitive services in all works of life.

This book blends big-data theories with emerging technologies on smart clouds over the IoT. The data analysts and computer scientists must learn how to use clouds and IoT effectively to discover new knowledge or making critical decision intelligently. This book is aimed to close up the gaps between these learning groups. We encourage machine learning and collaborative work between data scientists and cloud programmers.

## **Audience and Our Unique Approach :**

To achieve the above goals, we write this book to meet the growing demand in Computer Science and Electrical Engineering curriculum. Four courses on *Big Data Analytics, Cloud Computing, Internet of Things, and Machine Learning* can use this book either as a required textbook or as a major reference. These courses are now offered at major universities in the US, Asia and Europe. In addition, this book should benefit computer professionals transforming their skills to meet the new challenges.

We take a technological fusion approach to integrating big-data theories, cloud design principles, IoT sensing, machine learning, data analytics, Hadoop and Spark programming in a single volume. Both authors have taught these courses at USC and HUST for years. The topics covered in this book are also in great demand at many universities globally.

The main theme of this book is to promote effective big-data computing on smart clouds that are fully supported by IoT sensing, machine learning, and analytics software systems. The book material is an outgrowth of authors' established research work and by their teaching experiences accumulated over the years. This book will benefit a wide scope of audience across interactive or collaborative applications.

## **A Quick Glance of All Chapters:**

The book has three Parts. Part 1 has two introductory chapters on data science and the roles of clouds and IoT for big-data computing. These two chapters lay the necessary background of

enabling technologies for exploring big-data cloud computing and expanding innovative IoT sensing applications.

Part 2 has three chapters related to cloud system architecture, IoT sensing platforms, software support, security precautions, and performance tuning techniques. We present the principles and construction of smart clouds such as the Amazon AWS, Google AppEngine, Microsoft Azure, and many clouds built by Apple, Facebook, etc. We review the evolution from MapReduce to Hadoop and Spark libraries for cloud applications. The IoT sensing is integrated with mobile and social networking technologies. We will also cover cloud security and performance issues.

Part 3 has three chapters covering the theories and applications of big data processing, IoT sensing, machine learning, deep learning, and data analytics. We present an in-depth treatment of health-care, social-media and deep learning systems supported by smart clouds and the IoT. Case studies report existing clouds and IoT applications with benchmarks suites over some real-life big datasets.

## Part 1: Big Data Science and Enabling Technologies

Chapter 1: Big Data Science for Knowledge Discovery

Chapter 2: Smart Clouds and IoT Support of Big Data

## Part 2: Smart Clouds and IoT for Massive Data Processing

Chapter 3: Cloud Systems and Infrastructure Design

Chapter 4: IoT Sensing, Mobile and Networking Technologies

Chapter 5: MapReduce, Hadoop and Spark Programming

Chapter 6: Cloud Security, Mobility and Performance

## Part 3: Machine Learning, IoT Sensing and Data Analytics

Chapter 7: Machine Learning and Data Analytics

Chapter 8: Deep Learning Algorithms and Applications

Chapter 9: Health-Care Applications with Cloud/IoT Support

Appendix A: Literature Guide and Book Website for Courseware

Appendix B: Introduction to the OPNET-IoT Simulator

Subject Index

## Contact of the Authors:

Kai Hwang : [kaihwang@usc.edu](mailto:kaihwang@usc.edu) and Min Chen: [minchen2012@hust.edu.cn](mailto:minchen2012@hust.edu.cn)

**Production Status:** The book manuscript is presently under copy-editing and production. After proofreading, it will enter the final printing by late 2016. The hard copy is expected to have 400+ pages, to appear in late 2016 or early 2017.

# **Detailed Table of Contents** : (Revised as of April 8, 2016)

## ***Front Matter: Foreword, Preface and User Guide***

## ***Part 1: Big Data Science and Enabling Technologies***

### **Chapter 1: Big Data Science for Knowledge Discovery**

- 1.1 Enabling Technologies for Big Data Computing**
  - 1.1.1 Data Science and Related Disciplines
  - 1.1.2 Emerging Technologies for Big Data Computing
  - 1.1.3 Interactive SMACT Technologies
- 1.2 Social-Media Networking and Mobile Cloud Computing**
  - 1.2.1 Social Networks and Web Service Sites
  - 1.2.2 Mobile Cellular Core Networks
  - 1.2.3 Mobile Devices and Internet Edge Networks
  - 1.2.4 Mobile Cloud Computing Infrastructure
- 1.3 Big Data Mining and Analytics Strategy**
  - 1.3.1 Big Data Value Chain
  - 1.3.2 Strategy for Big Data Analytics
  - 1.3.3 Data Analytics Process Model
- 1.4 Machine Intelligence for Big Data Applications**
  - 1.4.1 Knowledge Discovery and Machine Learning
  - 1.4.2 Supervised and Unsupervised Learning Techniques
  - 1.4.3 An Overview of Big Data Applications
- 1.5 Conclusions, References and Exercises**

### **Chapter 2: Smart Clouds and IoT Support for Big Data**

- 2.1 Cloud Computing Models and Services**
  - 2.1.1 Enabling Technologies for Clouds
  - 2.1.2 A Generic Cloud Architecture
  - 2.1.3 Layered Development Cloud Services
  - 2.1.4 Clouds for Big Data Storage and Processing Engine
- 2.2 Virtualization in Cloud Computing Systems**
  - 2.2.1 Basic Concepts of Virtualization
  - 2.2.2 Hypervisors and Virtual Machines
  - 2.2.3 Docker Engine and Application Containers
  - 2.2.4 Application Software Libraries for Big Data
- 2.3 Sensing Technologies for Internet of Things**
  - 2.3.1 Enabling Technologies and Evolution of IoT
  - 2.3.2 Roadmap of Developing The Internet of Things
  - 2.3.3 IoT Sensing Technologies (RFID and Sensors)
- 2.4 IoT Architectures and Interaction Frameworks**
  - 2.4.1 IoT Architecture and Wireless Support
  - 2.4.2 Local vs, Global Positioning Technologies

- 2.4.3 Standalone vs. Cloud-Centric IoT applications
- 2.4.4 IoT Interaction Frameworks with Environments
- 2.5 Conclusions, References and Home Work**

## ***Part 2: Smart Clouds and IoT for Big Data Processing***

### **Chapter 3: Cloud Systems and Infrastructure Design**

#### **3.1 Introduction**

- 3.1.1 Public, Private, Community and Hybrid Clouds
- 3.1.2 Cloud Ecosystems and Mashup Services
- 3.1.3 Scalability and Fault Tolerance in Cloud Clusters
- 3.1.4 Availability of Server Clusters in Clouds

#### **3.2 Cloud Infrastructure Design Principles**

- 3.2.1 Physical vs. Virtual Server Clusters
- 3.2.2 Market-Oriented Cloud Architecture
- 3.2.3 Converting Datacenters into Clouds
- 3.2.4 Resources Provisioning Methods
- 3.2.5 Cloud Services and Business Models

#### **3.3 IaaS, PaaS and SaaS Cloud Systems**

- 3.3.1 Infrastructure as a Service (IaaS) Clouds
- 3.3.2 Amazon AWS Architecture
- 3.3.3 AWS Service Offerings
- 3.3.4 Platform PaaS Clouds – Google App Engine
- 3.3.5 Application SaaS Clouds – The Salesforce Clouds

#### **3.4 Virtual Machine and Container Management and Cloud OS**

- 3.4.1 Virtual Machine Creation and Management
- 3.4.2 Live VM Migration and Disaster Recovery
- 3.4.3 OpenStack for Constructing Cloud Infrastructure
- 3.4.4 VM/Container Scheduling and Orchestration
- 3.4.5 Cloud Operating Systems: Eucalyptus and vSphere/4

#### **3.5 Conclusions, References, and Exercises**

### **Chapter 4: IoT Sensing, Mobile and Networking Technologies**

#### **4.1 Introduction**

- 4.1.1 Sensing, Networking and Embedded Computing
- 4.1.2 Layered Development of IoT Platforms

#### **4.2 Radio Frequency Identification (RFID)**

- 4.2.1 RFID Technology and Tagging Devices
- 4.2.2 RFID System Architecture
- 4.2.3 RFID for Manufacturing Retailing and Logistics

#### **4.3 Sensors and Related Sensing Technologies**

- 4.3.1 Sensors Design and Platform Architecture
- 4.3.2 Hardware Components and Operating Systems
- 4.3.3 Sensing Through Smart Phones

- 4.3.4 Software-Defined IoT Sensing
- 4.4 IoT Wireless Networking and Positioning Technologies**
  - 4.6.1 Spectrum of IoT Wireless Technologies
  - 4.6.2 Wide-Range Wireless Sensing Networks
  - 4.6.3 Wireless Sensor Networks and Body Area Networks
  - 4.6.4 Global Positioning Systems
  - 4.6.4 Location-Sensitive IoT Applications
- 4.5 Context-Aware and Cloud-assisted IoT Processing**
  - 4.5.1 Context-Aware Computing for IoT Development
  - 4.5.2 Cloud-Assisted IoT for Health-Care Services
  - 4.5.3 OPNET-based IoT Simulation Mobile Cloud based IoT Sensing
  - 4.6.4 Cloud-Assisted IoT for Healthcare Services
- 4.7 Conclusions, References and Exercises**

## **Chapter 5: MapReduce, Hadoop and Spark Programming**

- 5.1 Evolution of Scalable Parallel Computing**
  - 5.1.1 Characteristic of Scalable Parallel Computing
  - 5.1.2 From MapReduce to Hadoop and Spark
  - 5.1.3 Software Libraries for Big-Data Cloud Applications
- 5.2 MapReduce Computing in Batch Mode**
  - 5.2.1 Basic Concept of MapReduce Paradigm
  - 5.2.2 MapReduce for Matrix Multiplication
  - 4.2.3 Variants of MapReduce Computing
- 5.3 Hadoop Programming and Recent Extensions**
  - 5.3.1 Hadoop YARN for Resource Management
  - 5.3.2 Hadoop Distributed File System (HDFS)
  - 5.3.3 Recent Hadoop Feature Extensions
- 5.4 Core Concepts of Spark Libraries**
  - 5.4.1 Spark Core for General-Purpose Applications
  - 5.4.2 In-Memory Computation and Language Support
  - 5.4.3 Spark Resilient Distributed Datasets (RDDs)
- 5.5 Spark for Real-Time Cloud Computing**
  - 5.5.1 Spark SQL with Structured Data
  - 5.5.2 Spark Streaming for Live Stream of Data
  - 5.5.3 Spark MLlib for Machine Learning
  - 5.5.4 Spark GraphX for Graph Processing
- 5.6 Conclusions, References and Exercises**

## **Chapter 6: Cloud Security, Mobility and Performance**

- 6.1 Introduction**
  - 6.1.1 What Are Cloud Performance and The QoS?
  - 6.1.2 On Trusted Cloud Computing Environments
- 6.2 Cloud Security and Privacy Protection**

- 6.2.1 Cloud Security and Privacy Issues
- 6.2.2 Cloud Security Infrastructure
- 6.2.3 Data Privacy Protection Schemes
- 6.3 Mobile Computing with Remote Clouds**
  - 6.3.1 Mobile Access of Remote Clouds
  - 6.3.2 Cloudlets for Mobile Cloud Computing
  - 6.3.3 Cloudlet Mesh for Trusted Offloading in Mobile Clouds
  - 6.3.4 Defense against Security Threats in a Mobile World
- 6.4 Cloud Mashup Service Discovery and Composition**
  - 6.4.1 Inter-Cloud and Mashup Services
  - 6.4.2 Cloud Mashup Service Models
  - 6.4.3 Discovery and Composition of Mashup Services
  - 6.4.4 Composition of Mashup Services
  - 6.4.5 Performance of Cloud Mashup Services
- 6.5 Cloud Performance Metric and Benchmarks**
  - 6.5.1 Scale-Out and Scale-Up Strategies
  - 6.5.2 Cloud Performance Metrics and Benchmarks
  - 6.5.3 Cloud Benchmark Suites
- 6.6 Scaling Strategies and Reported Performance Results**
  - 6.6.1 Elasticity Analysis of Cloud Performance
  - 6.6.2 Scale-Out and Scale-Up Performance
  - 6.6.3 Relative Merits of Different Scaling Strategies
- 6.7 Conclusions, References and Exercises**

## ***Part 3: Machine Learning, IoT Sensing and Data Analytics***

### **Chapter 7: Machine Learning and Data Analytics**

- 7.1 Introduction**
- 7.2 Regression Techniques**
  - 7.2.1 Linear Regression
  - 7.2.2 Logistic Regression
- 7.3 Classification Techniques**
  - 7.3.1 Bayesian Classifiers
  - 7.3.2 Bayesian Belief Network
  - 7.3.3 Rule-based Classification
  - 7.3.4 Nearest Neighbor Classifier
- 7.4 Clustering Techniques**
  - 7.4.1 Cluster Analysis
  - 7.4.2 K- means clustering
  - 7.4.3 Agglomerative hierarchical clustering
  - 7.4.4 Density-based clustering
- 7.5 Decision Trees and Support Vector Machines**

- 7.5.1 Decision Tree
- 7.5.2 Random Forest
- 7.5.3 Support Vector Machines
- 7.5.4 Correlation Analysis

## **7.6 Conclusions, References and Exercises**

# **Chapter 8: Deep Learning Algorithms and Applications**

## **8.1 Introduction**

- 8.1.1 Deep Learning Mimics Human Senses
- 8.1.2 Representation Learning

## **8.2 Auto-encode and Deep Neural Networks**

- 8.2.1 Artificial Neural Networks
- 8.2.2 Auto Encoding
- 8.2.3 Deep Neural Networks

## **8.3 Restricted Boltzmann Machine and Deep Belief Networks**

- 8.3.1 Restricted Boltzmann Machine
- 8.3.2 Deep Belief Network Machine

## **8.4 Deep Convolutional Neural Networks**

- 8.4.1 Convolution
- 8.4.2 Pooling
- 8.4.3 Deep Convolutional Neural Networks

## **8.5 Image Understanding via Deep Learning**

- 8.5.1 Image Understanding
- 8.5.2 Convolutional Neural networks and Medical Image Understanding

## **8.6 Natural Language Processing via Deep Learning**

- 8.6.1 Natural Language Processing
- 8.6.3 Convolutional Neural networks and Medical Text Analytics

## **8.7 Conclusions, References and Exercise**

# **Chapter 9: Health-Care Applications with Cloud/IoT Support**

## **9.1 Introduction**

- 9.1.1 IoT Sensing for Environment and Healthcare Monitoring
- 9.1.2 Cloud-assisted Environment and Healthcare Services

## **9.2 Healthcare and Medical Data**

- 9.2.1 Clinical Data
- 9.2.2 Behavioral and Emotional Data
- 9.2.3 Medical Image and Health Check Data

## **9.4 IoT-based Healthcare Applications**

- 9.4.1 IoT Sensing for Body Signal
- 9.4.2 Healthcare Monitoring System
- 9.4.3 Exercise Promotion
- 7.4.4 Medical Healthcare System

## **9.5 Applications of Healthcare Big Data Analytics**

- 9.5.1 Big Data Healthcare Problem
- 9.5.2 Healthcare Big Data Source
- 9.5.3 Medical Big Data Healthcare Platform
- 9.5.5 Decision-making via Healthcare Big Data Analytics

## 9.6 Emotion-aware Healthcare

- 9.6.1 Applications of Affective Computing
- 9.6.2 Mental Healthcare
- 9.6.3 Emotion-aware Computing and Services
- 9.6.4 Emotion-care via Robotics Cloud Technologies

## 9.7 Conclusions, References and Exercises

# Appendix A: Literature Guide and Book Website for Courseware

# Appendix B: Introduction To OPNET- IoT Simulator with Tested Applications

## Subject Index

## Biographical Sketches of Authors:

**Kai Hwang** is a Professor of Electrical Engineering and Computer Science at the University of Southern California (USC). With a Ph.D. from the University of California, Berkeley, he specializes in computer architecture, wireless Internet, cloud computing, and network security. An IEEE Life Fellow, he has served as the founding Editor-in-Chief of the *Journal of Parallel and Distributed Computing* (JPDC) for 28 years. He has published 8 books, including *Computer Architecture and Parallel Processing* (McGraw-Hill 1983), *Advanced Computer Architecture* (McGraw-Hill 2010). The American Library Association has rated his recent book, *Distributed and Cloud Computing* (with Fox and Dongarra) as an outstanding academic title published by Elsevier in 2012.

Over the years, Dr. Hwang has published 250 scientific papers. According to Google Scholars, his books/papers are cited over 15,000 times with an h-index of 53. His book on Computer Architecture was cited more than 2300 times and his best paper on PowerTrust for P2P computing was cited 660 times. He has received the Lifetime Achievement Award from *IEEE CloudCom* 2012. He has produced 21 Ph.D. students at USC and Purdue University, 4 of them recognized as IEEE Fellow and one an IBM Fellow. He has delivered 50+ keynote and distinguished lectures in IEEE/ACM conferences or at major universities and research centers worldwide.

**Min Chen** is a Professor of Computer Science and Technology at Huazhong University of Science and Technology (HUST) in Wuhan, China. He received the Ph.D. in 2004 in Communication and Information Systems from the South China University of Science and Technology in Guangzhou, China. He has performed post-doctoral research at the University of British Columbia in Canada and served on the faculty of Seoul National University in Korea for 6 years.

An IEEE Senior Member, he has published 180 papers in the areas of Internet of Things, Mobile Cloud, Body Area Networks, Healthcare Big Data, and Cyber Physical Systems, most in IEEE or ACM publications. He has received two Best Paper Awards from the IEEE QShine 2008 and ICC 2012. He has published two books: *OPNET IoT Simulation* (2015) with HUST Press and *Big Data Related Technologies* (2014) with the Springer Computer Science Series. He has served as editors for 7 IEEE/ACM Journals and Magazines. He delivered the keynote speech at *IEEE CloudCom* 2015.