Preface xiii

PART I: CLOUD, BIG DATA, AND COGNITIVE COMPUTING  1

1 Principles of Cloud Computing Systems  3
   1.1 Elastic Cloud Systems for Scalable Computing  3
      1.1.1 Enabling Technologies for Cloud Computing  3
      1.1.2 Evolution of Scalable Distributed/Parallel Computing  6
      1.1.3 Virtualized Resources in Cloud Systems  9
      1.1.4 Cloud Computing versus On-Premise Computing  10
   1.2 Cloud Architectures Compared with Distributed Systems  13
      1.2.1 Basic Cloud Platform Architectures  13
      1.2.2 Public, Private, Community, and Hybrid Clouds  16
      1.2.3 Physical Clusters versus Virtual Clusters  19
      1.2.4 Comparison with Other Parallel/Distributed Systems  22
   1.3 Service Models, Ecosystems, and Scalability Analysis  25
      1.3.1 Cloud Service Models: IaaS, PaaS, and SaaS  25
      1.3.2 Scalability Laws in Evaluating Cloud Performance  29
      1.3.3 Cloud Ecosystem and User Environments  32
      1.3.4 Gartner Hype Cycle for Cloud Computing  35
      1.3.5 Interaction among SMACT Technologies  37
   1.4 Availability, Mobility, and Cluster Optimization  40
      1.4.1 Availability Analysis of Cloud Server Clusters  40
      1.4.2 Fault Tolerance in Virtual Cluster Operations  43
      1.4.3 Queueing Model of Multiserver Clusters in Clouds  44
      1.4.4 Multiserver Cluster Optimization for Cloud Computing  46
   1.5 Conclusions  50
Homework Problems  50
2 Data Analytics, Internet of Things and Cognitive Computing 57
  2.1 Big Data Science and Application Challenges 57
    2.1.1 Data Science and Big Data Characteristics 57
    2.1.2 Gartner Hype Cycle for the Internet of Things 59
    2.1.3 Towards a Big Data Industry 61
    2.1.4 Big Data Applications: An Overview 64
  2.2 The Internet of Things and Cloud Interactions 68
    2.2.1 IoT Sensing and Platform Architecture 70
    2.2.2 IoT Value Chains and Development Road Map 72
    2.2.3 Stand-alone and Cloud-centric IoT Applications 75
    2.2.4 Smart City and Smart Community Development 80
  2.3 Data Collection, Mining, and Analytics on Clouds 82
    2.3.1 Data Quality Control and Representations 82
    2.3.2 Data Mining and Data Analytics 88
    2.3.3 Upgrading Data Analytics on Clouds 89
    2.3.4 Cloud Resources for Supporting Big Data Analytics 93
  2.4 Neuromorphic Hardware and Cognitive Computing 97
    2.4.1 Cognitive Computing and Neuromorphic Processors 97
    2.4.2 SyNAPSE and Related Neurocomputer Projects at IBM 99
    2.4.3 Cambricom NPU at the Chinese Academy of Sciences 103
    2.4.4 Google’s TPU and Related AI Programs 104
  2.5 Conclusions 106
Homework Problems 107

PART II: CLOUD ARCHITECTURE AND SERVICE PLATFORM DESIGN 111

3 Virtual Machines, Docker Containers, and Server Clusters 113
  3.1 Virtualization in Cloud Computing Systems 113
    3.1.1 Basic Concept of Machine Virtualization 113
    3.1.2 Implementation Levels of Virtualization 116
    3.1.3 Resources Virtualization in Cluster or Cloud Systems 119
  3.2 Hypervisors for Creating Native Virtual Machines 121
    3.2.1 Virtual Machine Architecture Types 121
    3.2.2 Full Virtualization and Hosted Virtualization 125
    3.2.3 Paravirtualization with Guest OS Modification 127
    3.2.4 Comparison of Platform Virtualization Software Products and Toolkits 130
  3.3 Docker Engine and Application Containers 132
    3.3.1 Virtualization at Linux Kernel Level 132
3.4 Docker Containers and Deployment Requirements 136
   3.4.1 Docker Containers Created with Linux Kernel Functions 137
   3.4.2 Docker Containers versus Virtual Machines 139
   3.4.3 Architectural Evolution from VMs to Containers and Unikernel 141
3.5 Virtual Machine Management and Container Orchestration 144
   3.5.1 VM Management Solutions 144
   3.5.2 VM Migration for Disaster Recovery 147
   3.5.3 Docker Container Scheduling and Orchestration 149
3.6 Eucalyptus, OpenStack, and VMware for Cloud Construction 153
   3.6.1 Eucalyptus for Virtual Clustering in Private Clouds 153
   3.6.2 OpenStack Software for Building Private or Public Clouds 156
   3.6.3 VMware Virtualization Support for Building Hybrid Clouds 158
3.7 Conclusions 160
Homework Problems 161

4 Cloud Architectures and Service Platform Design 167
4.1 Cloud Architecture and Infrastructure Design 167
   4.1.1 Public Clouds and Service Offerings 167
   4.1.2 Business Models of Cloud Services 170
   4.1.3 Converting Data Centers to Cloud Platforms 174
   4.1.4 Elastic Resources Provisioning Methods 178
4.2 Dynamic Deployment of Virtual Clusters 180
   4.2.1 Virtual Cluster Deployment Projects 181
   4.2.2 Virtual Cluster Configuration Adaptation 183
   4.2.3 Virtualization Support for Data Center Clusters 184
   4.2.4 VMware vSphere 6: A Commercial Cloud Operating System 185
4.3 Amazon AWS Cloud and Service Offerings 188
   4.3.1 Three Cloud Architectures and Services Convergence 188
   4.3.2 AWS EC2 Compute Engine and S3 Storage Cloud 192
   4.3.3 Other AWS Cloud Service Offerings 195
4.4 Google App Engine and Microsoft Azure 200
   4.4.1 Google App Engine and Compute Engine 200
   4.4.2 Google Hardware/Software Support for Machine Learning Services 205
   4.4.3 Microsoft Azure and Service Offerings 206
4.5 Salesforce, IBM SmartCloud, and Other Clouds 212
   4.5.1 Salesforce Clouds for SaaS Services 212
   4.5.2 IBM SmartCloud, IoT, and Cognitive Projects 215
   4.5.3 Clouds at SGI, NASA, and CERN 218
4.6 Conclusions 223
Homework Problems 223

5 Clouds for Mobile, IoT, Social Media, and Mashup Services 229
5.1 Wireless Internet and Mobile Cloud Computing 229
  5.1.1 Mobile Devices and Internet Edge Networks 229
  5.1.2 Wi-Fi, Bluetooth, and Wireless Sensor Networks 232
  5.1.3 Cloudlet Mesh for Mobile Cloud Computing 233
  5.1.4 Mobile Clouds and Colocation Clouds 236
5.2 IoT Sensing and Interaction with Clouds 240
  5.2.1 Local and Global Positioning Systems 241
  5.2.2 Cloud-Based RAN for Building Mobile Networks 242
  5.2.3 IoT Interaction Frameworks with Clouds and Devices 246
5.3 Cloud Computing in Social Media Applications 250
  5.3.1 Social Media Big-Data Industrial Applications 251
  5.3.2 Social Networks and API for Social Media Applications 255
  5.3.3 Social Graph Properties and Representations 258
  5.3.4 Social Graph Analysis on Smart Clouds 262
5.4 Multicloud Mashup Architecture and Service 264
  5.4.1 Cloud Mashup Architecture for Agility and Scalability 265
  5.4.2 Multicloud Mashup Service Architecture 268
  5.4.3 Skyline Discovery of Mashup Services 273
  5.4.4 Dynamic Composition of Mashup Services 275
5.5 Conclusions 277
Homework Problems 278

PART III: PRINCIPLES OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE MACHINES 283

6 Machine Learning Algorithms and Model Fitting 285
6.1 Taxonomy of Machine Learning Methods 285
  6.1.1 Categories of Machine Learning Algorithms 285
  6.1.2 Supervised Machine Learning Algorithms 289
  6.1.3 Unsupervised Machine Learning Algorithms 290
6.2 Supervised Regression and Classification Methods 291
  6.2.1 Linear Regression for Prediction or Forecasting 291
  6.2.2 Decision Trees for Machine Learning 299
  6.2.3 Bayesian Classifier with Training Samples 303
  6.2.4 Support Vector Machines (SVM) 307
6.3 Clustering and Dimensionality Reduction Methods 310
PART IV: CLOUD PROGRAMMING AND PERFORMANCE BOOSTERS  401

8  Cloud Programming with Hadoop and Spark  403
  8.1  Scalable Parallel Computing Over Large Clusters  403
    8.1.1 Characteristics of Scalable Computing  403
    8.1.2 From MapReduce to Hadoop and Spark  404
    8.1.3 Application Software Libraries for Big Data Processing  406
  8.2  Hadoop Programming with YARN and HDFS  407
    8.2.1 The MapReduce Compute Engine  408
    8.2.2 MapReduce for Parallel Matrix Multiplication  413
    8.2.3 Hadoop Architecture and Recent Extensions  416
    8.2.4 Hadoop Distributed File System  421
    8.2.5 Hadoop YARN for Resource Management  424
  8.3  Spark Core and Resilient Distributed Data Sets  426
    8.3.1 Spark Core for General-Purpose Applications  426
    8.3.2 Resilient Distributed Data Sets  429
    8.3.3 Spark Programming with RDDs for DAG Tasks  432
  8.4  Spark SQL and Streaming Programming  435
    8.4.1 Spark SQL with Structured Data  436
    8.4.2 Spark Streaming with Live Stream of Data  437
    8.4.3 Spark Streaming Application Examples  440
  8.5  Spark MLlib for Machine Learning and GraphX for Graph Processing  442
    8.5.1 Spark MLlib Library for Machine Learning  442
    8.5.2 Some MLlib Application Examples  444
    8.5.3 Spark GraphX for Graph Processing  445
    8.5.4 Some GraphX Programming Examples  448
  8.6  Conclusions  452

Homework Problems  453

9  TensorFlow, Keras, DeepMind, and Graph Analytics  463
  9.1  TensorFlow for Neural Network Computing  463
    9.1.1 Key Concepts of TensorFlow  463
    9.1.2 Tensors, Variables, Feed, and Fetch Operations  466
    9.1.3 Distributed TensorFlow Execution Environment  470
    9.1.4 Execution Sessions in TensorFlow Programs  473
  9.2  TensorFlow System for Deep Learning  476
    9.2.1 Layered TensorFlow System Architecture  477
    9.2.2 TensorFlow Installation on Various Host Machines  480
    9.2.3 TensorFlow Ecosystem for Distributed Resources Sharing  482
    9.2.4 TensorFlow for Handwritten Digit Recognition  484