



Table of Contents

| | |
|--|--------------|
| Introduction | xxv |
| Book Preview..... | xxxii |
| Chapter 1: Getting an Overview of Big Data | 1 |
| What is Big Data?..... | 2 |
| History of Data Management - Evolution of Big Data..... | 5 |
| Structuring Big Data..... | 7 |
| Types of Data..... | 8 |
| Elements of Big Data | 12 |
| Volume..... | 13 |
| Velocity..... | 14 |
| Variety..... | 14 |
| Veracity | 14 |
| Big Data Analytics | 15 |
| Advantages of Big Data Analytics..... | 17 |
| Careers in Big Data..... | 20 |
| Skills Required | 22 |
| Future of Big Data..... | 22 |
| Summary..... | 25 |
| Quick Revise..... | 25 |
| Multiple-Choice Questions..... | 25 |
| Subjective Questions..... | 26 |
| Chapter 2: Exploring the Use of Big Data in Business Context..... | 29 |
| Use of Big Data in Social Networking..... | 30 |
| Business Intelligence..... | 33 |
| Marketing..... | 34 |
| Product Design and Development | 35 |
| Use of Big Data in Preventing Fraudulent Activities..... | 36 |
| Preventing Fraud Using Big Data Analytics | 37 |

Table of Contents

| | |
|--|-----------|
| Use of Big Data in Detecting Fraudulent Activities in Insurance Sector | 39 |
| Fraud Detection Methods | 39 |
| Use of Big Data in Retail Industry | 43 |
| Use of RFID Data in Retail..... | 44 |
| Summary..... | 46 |
| Quick Revise..... | 47 |
| Multiple-Choice Questions..... | 47 |
| Subjective Questions..... | 48 |
| Chapter 3: Introducing Technologies for Handling Big Data..... | 53 |
| Distributed and Parallel Computing for Big Data..... | 54 |
| How Data models and Computing models are different? | 58 |
| Introducing Hadoop..... | 60 |
| HDFS and MapReduce..... | 61 |
| How does Hadoop Function? | 63 |
| Cloud Computing and Big Data | 65 |
| Features of Cloud Computing..... | 66 |
| Cloud Deployment Models | 67 |
| Cloud Delivery Models..... | 70 |
| Cloud Services for Big Data..... | 72 |
| Cloud Providers in Big Data Market..... | 73 |
| In-Memory Computing Technology for Big Data | 74 |
| Summary | 75 |
| Quick Revise..... | 76 |
| Multiple-Choice Questions..... | 76 |
| Subjective Questions..... | 77 |
| Chapter 4: Understanding Hadoop Ecosystem | 83 |
| Hadoop Ecosystem | 84 |
| Hadoop Distributed File System..... | 86 |
| HDFS Architecture | 87 |
| Concept of Blocks in HDFS Architecture..... | 88 |
| NameNodes and DataNodes..... | 89 |
| The Command-Line Interface | 90 |
| Using HDFS Files..... | 90 |
| Hadoop-Specific File System Types | 91 |
| HDFS Commands..... | 93 |
| The org.apache.hadoop.io package | 94 |
| HDFS High Availability..... | 96 |

| | |
|---|------------|
| Features of HDFS..... | 99 |
| MapReduce..... | 101 |
| Hadoop YARN..... | 102 |
| Introducing HBase..... | 102 |
| HBase Architecture..... | 103 |
| Regions..... | 104 |
| Storing Big Data with HBase..... | 104 |
| Interacting with the Hadoop Ecosystem..... | 105 |
| HBase in Operation - Programming with HBase..... | 105 |
| Combining HBase and HDFS..... | 106 |
| REST and Thrift..... | 108 |
| Data Integrity in HDFS..... | 110 |
| Features of HBase..... | 111 |
| Hive..... | 114 |
| Pig and Pig Latin..... | 114 |
| Sqoop..... | 114 |
| ZooKeeper..... | 114 |
| Flume..... | 115 |
| Oozie..... | 115 |
| Summary..... | 115 |
| Quick Revise..... | 115 |
| Multiple-Choice Questions..... | 115 |
| Subjective Questions..... | 116 |
| Chapter 5: Understanding MapReduce Fundamentals and HBase..... | 121 |
| The MapReduce Framework..... | 122 |
| Exploring the Features of MapReduce..... | 123 |
| Working of MapReduce..... | 124 |
| Exploring Map and Reduce Functions..... | 128 |
| Techniques to Optimize MapReduce Jobs..... | 131 |
| Hardware/Network Topology..... | 131 |
| Synchronization..... | 132 |
| File System..... | 132 |
| Uses of MapReduce..... | 133 |
| Role of HBase in Big Data Processing..... | 134 |
| Characteristics of HBase..... | 134 |
| Installation of HBase..... | 135 |
| Summary..... | 142 |
| Quick Revise..... | 143 |
| Multiple-Choice Questions..... | 143 |
| Subjective Questions..... | 145 |

| | |
|---|------------|
| Chapter 6: Understanding Big Data Technology Foundations..... | 149 |
| Exploring the Big Data Stack..... | 150 |
| Data Sources Layer | 151 |
| Ingestion Layer..... | 152 |
| Storage Layer..... | 153 |
| Physical Infrastructure Layer..... | 154 |
| Platform Management Layer..... | 156 |
| Security Layer..... | 157 |
| Monitoring Layer..... | 158 |
| Analytics Engine | 158 |
| Visualization Layer..... | 159 |
| Big Data Applications..... | 160 |
| Virtualization and Big Data..... | 162 |
| Virtualization Approaches | 166 |
| Server Virtualization | 166 |
| Application Virtualization..... | 167 |
| Network Virtualization..... | 167 |
| Processor and Memory Virtualization..... | 167 |
| Data and Storage Virtualization..... | 167 |
| Managing Virtualization with Hypervisor..... | 168 |
| Implementing Virtualization to Work with Big Data | 169 |
| Summary..... | 170 |
| Quick Revise..... | 170 |
| Multiple-Choice Questions..... | 170 |
| Subjective Questions..... | 172 |
| | |
| Chapter 7: Storing Data in Databases and Data Warehouses | 177 |
| RDBMS and Big Data | 178 |
| CAP Theorem..... | 183 |
| Issues with the Relational Model..... | 184 |
| Non-Relational Database | 184 |
| Issues with the Non-Relational Model..... | 185 |
| Polyglot Persistence..... | 186 |
| Integrating Big Data with Traditional Data Warehouses | 187 |
| Big Data Analysis and Data Warehouse..... | 190 |
| Changing Deployment Models in Big Data Era | 194 |
| Summary..... | 196 |
| Quick Revise..... | 196 |
| Multiple-Choice Questions..... | 196 |
| Subjective Questions..... | 196 |

| | |
|--|------------|
| Chapter 8: Processing Your Data with MapReduce | 199 |
| Recollecting the Concept of MapReduce Framework..... | 200 |
| Developing Simple MapReduce Application..... | 201 |
| Building the Application..... | 206 |
| Executing the Application | 208 |
| Points to Consider while Designing MapReduce..... | 213 |
| Summary..... | 214 |
| Quick Revise..... | 214 |
| Multiple-Choice Questions..... | 214 |
| Subjective Questions..... | 215 |
| | |
| Chapter 9: Customizing MapReduce Execution and Implementing MapReduce Program | 217 |
| Controlling MapReduce Execution with InputFormat..... | 218 |
| InputSplit | 218 |
| RecordReader | 219 |
| FileInputFormat | 220 |
| Implementing InputFormat for Compute-Intensive Applications | 221 |
| Implementing InputFormat to Control the Number of Maps | 224 |
| Implementing InputFormat for Multiple HBase Tables..... | 225 |
| Reading Data with Custom RecordReader | 228 |
| Organizing Output Data with OutputFormats..... | 231 |
| Customizing Data with RecordWriter | 232 |
| Optimizing MapReduce Execution with Combiner..... | 233 |
| Controlling Reducer Execution with Partitioners | 234 |
| Customizing the MapReduce Execution in Terms of YARN..... | 234 |
| Reduce Phase..... | 236 |
| Concept of Containers..... | 236 |
| Implementing a MapReduce Program for Sorting Text Data..... | 237 |
| Summary..... | 254 |
| Quick Revise..... | 254 |
| Multiple-Choice Questions..... | 254 |
| Subjective Questions..... | 255 |
| | |
| Chapter 10: Testing and Debugging MapReduce Applications..... | 257 |
| Debugging Hadoop MapReduce Locally | 258 |
| Performing Unit Testing for MapReduce Applications..... | 259 |
| Unit Testing the Mapper Component | 261 |
| Unit Testing the Reducer Component | 262 |
| Integration Testing of the Mapper-Reducer Combination..... | 263 |

Table of Contents

| | |
|---|------------|
| Performing Local Application Testing with Eclipse | 265 |
| Logging for Hadoop Testing..... | 266 |
| Application Log Processing..... | 268 |
| Defensive Programming in MapReduce..... | 270 |
| Summary..... | 272 |
| Quick Revise..... | 272 |
| Multiple-Choice Questions..... | 272 |
| Subjective Questions..... | 274 |
| Chapter 11: Understanding Hadoop YARN Architecture | 277 |
| Background of YARN..... | 278 |
| Revisiting MapReduce | 279 |
| Limitations of MapReduce | 280 |
| Advantages of YARN..... | 285 |
| YARN Architecture | 286 |
| ResourceManager | 286 |
| ApplicationManager..... | 287 |
| Integration of ResourceManager and ApplicationManager | 287 |
| Working of YARN | 288 |
| YARN Schedulers | 290 |
| CapacityScheduler | 290 |
| Fair Scheduler..... | 293 |
| Backward Compatibility with YARN | 295 |
| Script Compatibility | 296 |
| Binary Compatibility..... | 296 |
| Source Compatibility..... | 296 |
| YARN Configurations..... | 296 |
| YARN Commands..... | 299 |
| Administration Commands..... | 300 |
| User Commands..... | 301 |
| YARN Containers | 301 |
| Registry | 303 |
| Log Management in Hadoop 1 | 304 |
| Log Management in YARN..... | 304 |
| Summary..... | 306 |
| Quick Revise..... | 306 |
| Multiple-Choice Questions..... | 306 |
| Subjective Questions..... | 307 |

| | |
|---|------------|
| Chapter 12: Exploring Hive..... | 311 |
| Introducing Hive..... | 312 |
| Getting Started with Hive..... | 314 |
| Hive Services | 318 |
| Hive Variables..... | 319 |
| Hive Properties..... | 319 |
| Hive Queries..... | 319 |
| Data Types in Hive..... | 320 |
| Built-In Functions in Hive | 321 |
| Hive DDL..... | 323 |
| Creating Databases..... | 323 |
| Viewing a Database..... | 324 |
| Dropping a Database..... | 324 |
| Altering Databases..... | 324 |
| Creating Tables..... | 325 |
| Creating a Table Using the Existing Schema..... | 325 |
| Dropping Tables..... | 326 |
| Altering Tables..... | 326 |
| Using Hive DDL Statements | 327 |
| Data Manipulation in Hive..... | 329 |
| Loading Files into Tables..... | 329 |
| Inserting Data into Tables..... | 329 |
| Update in Hive..... | 331 |
| Delete in Hive..... | 332 |
| Using Hive DML Statements..... | 332 |
| Data Retrieval Queries | 333 |
| Using the SELECT Command..... | 334 |
| Using the WHERE Clause..... | 334 |
| Using the GROUP BY Clause..... | 334 |
| Using the HAVING Clause..... | 335 |
| Using the LIMIT Clause..... | 335 |
| Executing HiveQL Queries..... | 335 |
| Using JOINS in Hive..... | 337 |
| Inner Joins..... | 337 |
| Outer Joins..... | 338 |
| Cartesian Product Joins..... | 340 |
| Map-Side Joins | 340 |
| Joining Tables..... | 341 |
| Bucketing | 345 |

Table of Contents

| | |
|---|------------|
| Summary | 345 |
| Quick Revise..... | 346 |
| Multiple-Choice Questions..... | 346 |
| Subjective Questions..... | 347 |
| Chapter 13: Analyzing Data with Pig..... | 349 |
| Introducing Pig | 350 |
| The Pig Architecture..... | 350 |
| Benefits of Pig..... | 351 |
| Installing Pig..... | 351 |
| Properties of Pig..... | 352 |
| Running Pig | 354 |
| Running Pig Programs | 355 |
| Schema..... | 355 |
| Getting Started with Pig Latin | 356 |
| Pig Latin Application Flow..... | 357 |
| Pig Latin Structure..... | 357 |
| Working with Operators in Pig..... | 358 |
| FOREACH | 359 |
| ASSERT | 359 |
| FILTER..... | 360 |
| GROUP..... | 360 |
| ORDER BY | 361 |
| DISTINCT | 362 |
| JOIN | 363 |
| LIMIT..... | 364 |
| SAMPLE..... | 365 |
| SPLIT | 365 |
| FLATTEN..... | 366 |
| Debugging Pig..... | 367 |
| Multi-Query Execution | 368 |
| Working with Functions in Pig..... | 369 |
| Error Handling in Pig..... | 372 |
| Summary..... | 373 |
| Quick Revise..... | 373 |
| Multiple-Choice Questions..... | 373 |
| Subjective Questions..... | 374 |

| | |
|--|----------------|
| Chapter 14: Using Oozie | 379 |
| Introducing Oozie..... | 380 |
| Main Functional Components of Oozie..... | 381 |
| Benefits of Oozie | 382 |
| Installing and Configuring Oozie..... | 383 |
| Understanding the Oozie Workflow..... | 389 |
| Execution of Asynchronous Actions in Oozie | 392 |
| Implementing the Oozie Workflow..... | 393 |
| Oozie Recovery Capabilities..... | 396 |
| Oozie Workflow Life Cycle | 397 |
| Oozie Coordinator | 398 |
| Types of Oozie Coordinator | 398 |
| Oozie Coordinator Lifecycle Operations | 401 |
| Oozie Bundle | 402 |
| Oozie Parameterization with EL..... | 404 |
| Workflow Functions..... | 405 |
| Coordinator Functions | 405 |
| Bundle Functions | 405 |
| EL Functions | 405 |
| Oozie Job Execution Model | 406 |
| Accessing Oozie | 409 |
| Oozie SLA | 409 |
| Event Status | 410 |
| SLA Status..... | 410 |
| Oozie Activity..... | 410 |
| The Oozie SLA Subsystem..... | 410 |
| SLA Language Schema..... | 411 |
| Summary | 412 |
| Quick Revise..... | 412 |
| Multiple-Choice Questions..... | 412 |
| Subjective Questions..... | 414 |
| Chapter 15: NoSQL Data Management | 417 |
| Introduction to NoSQL | 418 |
| Why NoSQL?..... | 419 |
| Characteristics of NoSQL..... | 419 |
| History of NoSQL | 420 |
| Types of NoSQL Data Models..... | 420 |
| Key Value Data Model | 420 |
| Column-Oriented Data Model | 422 |

Table of Contents

| | |
|---|------------|
| Document Data Model | 423 |
| Graph Databases | 424 |
| Schema-Less Databases | 426 |
| Materialized Views | 426 |
| Distribution Models | 427 |
| CAP Theorem | 428 |
| Sharding | 429 |
| Summary | 431 |
| Quick Revise | 431 |
| Multiple-Choice Questions | 431 |
| Subjective Questions | 432 |
| Chapter 16: Data Movement with Flume and Sqoop | 435 |
| Flume Architecture | 436 |
| Recoverability and Multihop Flows | 437 |
| Complex Flow | 438 |
| Reliability and Failure Handling | 439 |
| Sqoop | 443 |
| How Sqoop Works | 443 |
| Sqoop Architecture | 446 |
| Importing Data | 447 |
| Importing Data into HDFS | 447 |
| Importing Data into Hive | 448 |
| Sqoop2 vs Sqoop | 448 |
| Summary | 451 |
| Quick Revise | 451 |
| Multiple-Choice Questions | 451 |
| Subjective Questions | 452 |
| Chapter 17: Introduction to Mahout | 457 |
| What is Mahout? | 458 |
| Machine Learning | 460 |
| Supervised Learning | 461 |
| Unsupervised Learning | 461 |
| Collaborative Filtering (Recommendation) | 464 |
| User-Based Recommendation | 465 |
| Item-Based Recommendation | 466 |
| Clustering | 466 |
| Classification | 467 |
| Mahout Algorithms | 467 |

| | |
|---|------------|
| Environment for Mahout..... | 469 |
| Java Installation on Linux..... | 469 |
| Hadoop Installation on Linux..... | 469 |
| Mahout Installation on Linux..... | 472 |
| Summary..... | 473 |
| Quick Revise..... | 473 |
| Multiple-Choice Questions..... | 473 |
| Subjective Questions..... | 473 |
| Chapter 18: Understanding Analytics and Big Data | 477 |
| Comparing Reporting and Analysis..... | 478 |
| Reporting..... | 479 |
| Analysis..... | 480 |
| The Analytic Process | 482 |
| Types of Analytics | 483 |
| Basic Analytics | 483 |
| Advanced Analytics | 484 |
| Operationalized Analytics..... | 485 |
| Monetized Analytics..... | 485 |
| Characteristics of Big Data Analysis | 486 |
| Points to Consider during Analysis | 486 |
| Frame the Problem Correctly | 486 |
| Statistical Significance or Business Importance? | 487 |
| Making Inferences versus Computing Statistics..... | 490 |
| Developing an Analytic Team..... | 490 |
| Skills Required for an Analyst..... | 491 |
| Convergence of IT and Analytics..... | 492 |
| Understanding Text Analytics | 493 |
| Summary..... | 494 |
| Quick Revise..... | 494 |
| Multiple-Choice Questions..... | 494 |
| Subjective Questions..... | 496 |
| Chapter 19: Analytical Approaches and Tools to Analyze Data..... | 501 |
| Analytical Approaches..... | 502 |
| Ensemble Methods..... | 502 |
| Text Data Analysis..... | 503 |
| History of Analytical Tools..... | 505 |
| Graphical User Interfaces..... | 505 |
| Point Solutions | 506 |

Table of Contents

| | |
|--|------------|
| Data Visualization Tools..... | 506 |
| Introducing Popular Analytical Tools..... | 508 |
| The R Project for Statistical Computing..... | 509 |
| IBM SPSS..... | 509 |
| SAS..... | 511 |
| Comparing Various Analytical Tools..... | 513 |
| Installing R..... | 515 |
| Installing R on a Windows Computer..... | 515 |
| Installing R on a Macintosh Computer..... | 521 |
| Installing R on a Linux Computer..... | 523 |
| Installing RStudio..... | 525 |
| Installing RStudio on Windows..... | 525 |
| Installing RStudio on Linux..... | 528 |
| Summary..... | 530 |
| Quick Revise..... | 530 |
| Multiple-Choice Questions..... | 530 |
| Subjective Questions..... | 532 |
| Chapter 20: Exploring R..... | 535 |
| Exploring Basic Features of R..... | 536 |
| Statistical Features..... | 537 |
| Programming Features..... | 538 |
| Packages..... | 538 |
| Graphical User Interfaces..... | 539 |
| Exploring RGUI..... | 539 |
| R Console..... | 540 |
| Developing a Program..... | 541 |
| Quitting R..... | 542 |
| Exploring RStudio..... | 542 |
| Handling Basic Expressions in R..... | 546 |
| Basic Arithmetic in R..... | 546 |
| Mathematical Operators..... | 548 |
| Variables in R..... | 549 |
| Calling Functions in R..... | 550 |
| Working with Vectors..... | 550 |
| Storing and Calculating Values in R..... | 552 |
| Creating and Using Objects..... | 555 |
| Interacting with Users..... | 556 |
| Handling Data in R Workspace..... | 557 |
| The ls() Function..... | 557 |

| | |
|--|------------|
| The rm() Function | 558 |
| The getwd() Function | 559 |
| The save() Function | 559 |
| The load () Function..... | 560 |
| Executing Scripts..... | 561 |
| Creating Plots | 564 |
| Accessing Help and Documentation in R..... | 565 |
| Summary | 569 |
| Quick Revise..... | 569 |
| Multiple-Choice Questions..... | 569 |
| Subjective Questions..... | 571 |
| | |
| Chapter 21: Reading Datasets and Exporting Data from R..... | 573 |
| Using the c() Command | 574 |
| Reading and Combining Numerical Data | 574 |
| Reading and Combining Text Data | 579 |
| Reading Both Numeric and Text Values in R | 580 |
| Using the scan() Command | 581 |
| Reading the Text Data Using the scan() Command | 583 |
| Using Clipboard to Create the Data | 584 |
| Reading the Data of a File from Disk | 586 |
| Reading Multiple Data Values from Large Files | 589 |
| Using the read.csv() Command | 590 |
| Using the read.table() Command..... | 591 |
| Reading Data from RStudio..... | 592 |
| Exporting Data from R..... | 595 |
| Using the write.table() Command | 595 |
| Using the write.csv() Command..... | 599 |
| Summary | 600 |
| Quick Revise..... | 600 |
| Multiple-Choice Questions..... | 600 |
| Subjective Questions..... | 602 |
| | |
| Chapter 22: Manipulating and Processing Data in R | 603 |
| Selecting the Most Appropriate Data Structure..... | 604 |
| Creating Data Subsets | 605 |
| Creating Subsets in Vectors | 605 |
| Creating Subsets in Data Frames | 607 |
| Merging Datasets in R..... | 609 |
| Using the merge() Function..... | 610 |

Table of Contents

| | |
|--|------------|
| Using the cbind() Function..... | 613 |
| Using the rbind() Function..... | 613 |
| Sorting Data..... | 614 |
| Sorting Data..... | 615 |
| Ordering Data..... | 615 |
| Reverse Sort..... | 616 |
| Putting Your Data into Shape..... | 618 |
| Transposing Data..... | 618 |
| Converting Data to Wide or Long Formats..... | 619 |
| Melting Data to Long Format..... | 619 |
| Casting Data to Wide Format..... | 621 |
| Managing Data in R Using Matrices..... | 622 |
| Reshaping a Vector into a Matrix..... | 622 |
| Accessing Matrix and Subsetting the Data..... | 623 |
| Managing Data in R Using Data Frames..... | 624 |
| Creating Data Frames..... | 624 |
| Accessing Data Frames..... | 625 |
| Merging Data Frames..... | 626 |
| Performing Operations on Data Frames..... | 628 |
| Summary..... | 629 |
| Quick Revise..... | 630 |
| Multiple-Choice Questions..... | 630 |
| Subjective Questions..... | 631 |
| Chapter 23: Working with Functions and Packages in R..... | 633 |
| Using Functions Instead of Scripts..... | 634 |
| Transforming the R Script into a Function..... | 634 |
| Returning Results in R..... | 637 |
| Reducing the Number of Lines in an R Function..... | 639 |
| Assigning the Function Objects..... | 640 |
| Writing Function Without Braces..... | 641 |
| Using Arguments in Functions..... | 642 |
| Using Dot Argument in Function..... | 644 |
| Passing Functions as Arguments..... | 645 |
| Anonymous Functions..... | 645 |
| Local and Global Environment of Functions..... | 646 |
| Built-in Functions in R..... | 647 |
| Numeric Functions..... | 648 |
| Character Functions..... | 648 |

| | |
|---|------------|
| Statistical Probability Functions..... | 649 |
| Miscellaneous Functions..... | 650 |
| Introducing Packages..... | 650 |
| Working with Packages..... | 651 |
| Summary..... | 656 |
| Quick Revise..... | 656 |
| Multiple-Choice Questions..... | 656 |
| Subjective Questions..... | 657 |
| Chapter 24: Performing Graphical Analysis in R..... | 659 |
| Using Plots..... | 660 |
| Using Plots for a Single Variable..... | 661 |
| Using Plots for Two Variables..... | 677 |
| Using Plots for Multiple Variables..... | 680 |
| Designing Special Plots..... | 681 |
| Saving Graphs to External Files..... | 683 |
| Advanced Features of R..... | 683 |
| Swapping X and Y Axes..... | 684 |
| Plotting Discrete Datasets..... | 685 |
| Plotting Continuous Datasets..... | 687 |
| Reversing the Direction of an Axis..... | 690 |
| Summary..... | 692 |
| Quick Revise..... | 692 |
| Multiple-Choice Questions..... | 692 |
| Subjective Questions..... | 693 |
| Chapter 25: Integrating R and Hadoop and Understanding Hive..... | 695 |
| RHadoop – An Integration of R and Hadoop..... | 696 |
| Installing RHadoop..... | 697 |
| Using RHadoop..... | 705 |
| Text Mining in RHadoop..... | 706 |
| Data Analysis Using the MapReduce Technique in RHadoop..... | 709 |
| Data Mining in Hive..... | 711 |
| Summary..... | 712 |
| Quick Revise..... | 712 |
| Multiple-Choice Questions..... | 712 |
| Subjective Questions..... | 713 |
| Chapter 26: Data Visualization-I..... | 715 |
| Ways of Representing Visual Data..... | 716 |

Table of Contents

| | |
|--|------------|
| Techniques Used for Visual Data Representation..... | 718 |
| Types of Data Visualization | 723 |
| Applications of Data Visualization..... | 725 |
| Visualizing Big Data..... | 725 |
| Deriving Business Solutions..... | 726 |
| Turning Data into Information | 726 |
| Tools Used in Data Visualization | 727 |
| Open-Source Data Visualization Tools..... | 730 |
| Analytical Techniques Used in Big Data Visualization | 731 |
| Tableau Products | 731 |
| Installation of Tableau Public..... | 732 |
| Summary..... | 737 |
| Quick Revise..... | 738 |
| Multiple-Choice Questions..... | 738 |
| Subjective Questions..... | 738 |
| Chapter 27: Data Visualization with Tableau (Data Visualization-II) | 741 |
| Introduction to Tableau Software..... | 742 |
| Tableau Desktop Workspace..... | 747 |
| Operations on Data..... | 757 |
| Data Analytics in Tableau Public..... | 757 |
| Using Visual Controls in Tableau Public..... | 768 |
| Overview of Tableau 9.0 | 774 |
| Summary..... | 779 |
| Quick Revise..... | 779 |
| Multiple-Choice Questions..... | 779 |
| Subjective Questions..... | 780 |
| Chapter 28: Social Media Analytics and Text Mining..... | 783 |
| Introducing Social Media..... | 784 |
| Introducing Key Elements of Social Media | 786 |
| Introducing Text Mining..... | 787 |
| Understanding Text Mining Process..... | 790 |
| Sentiment Analysis | 791 |
| Performing Social Media Analytics and Opinion Mining on Tweets..... | 793 |
| Online Social Media Analysis | 797 |
| Summary..... | 800 |
| Quick Revise..... | 800 |
| Multiple-Choice Questions..... | 800 |
| Subjective Questions..... | 801 |

| | |
|--|----------------|
| Chapter 29: Mobile Analytics | 805 |
| Introducing Mobile Analytics | 806 |
| Define Mobile Analytics..... | 808 |
| Mobile Analytics and Web Analytics..... | 809 |
| Types of Results from Mobile Analytics..... | 809 |
| Types of Applications for Mobile Analytics..... | 810 |
| Introducing Mobile Analytics Tools..... | 812 |
| Location-based Tracking Tools | 814 |
| Real-time Analytics Tools..... | 814 |
| User Behavior Tracking Tools..... | 816 |
| Performing Mobile Analytics | 817 |
| Challenges of Mobile Analytics | 832 |
| Summary..... | 833 |
| Quick Revise..... | 833 |
| Multiple-Choice Questions..... | 833 |
| Subjective Questions..... | 835 |
| Chapter 30: Finding a Job in the Big Data Market..... | 837 |
| Importance and Scope of Big Data Jobs | 839 |
| Big Data Opportunities | 841 |
| Skill Assessment for Big Data Jobs | 842 |
| Roles and Responsibilities in Big Data Jobs | 847 |
| Business Analyst for Big Data..... | 847 |
| Big Data Scientist | 848 |
| Software Developer for Big Data | 849 |
| Gaining a Foothold in the Big Data Market | 849 |
| Take Your Time..... | 850 |
| Preparing the Big Data Skill Learning and Testing Mechanism..... | 851 |
| Basic Educational Requirements for Big Data Jobs | 851 |
| Basic Technological Requirements for Big Data Jobs..... | 853 |
| Tools Supporting Big Data..... | 854 |
| Consultants and In-House Specialists in Big Data | 855 |
| Big Data Consultants..... | 855 |
| In-House Big Data Experts | 857 |
| Tactics for Searching Big Data Jobs | 858 |
| Network Building Tactics..... | 858 |
| Resume Building Tactics..... | 859 |
| Preparing for Interviews..... | 860 |
| Obtaining Big Data Jobs through Social Media | 862 |

Table of Contents

| | |
|------------------------------------|------------|
| Summary | 863 |
| Quick Revise..... | 864 |
| Multiple-Choice Questions..... | 864 |
| Subjective Questions..... | 864 |
| Big Data Practical | 875 |
| Appendix A: Cassandra | 947 |
| Appendix B: ZooKeeper..... | 955 |
| Index | 961 |