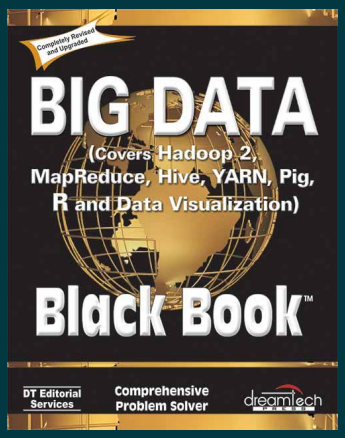
WILEY



Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization

By DT Editorial Services

Paperback 9789351199311 ₹1,299.00

Pages: 1008

Description

Big Data is one of the most popular buzzwords in technology industry today. Organizations worldwide have realized the value of the immense volume of data available and are trying their best to manage, analyse and unleash the power of data to build strategies and develop a competitive edge. At the same time, the advent of the technology has led to the evolution of a variety of new and enhanced job roles.

The objective of this book is to create a new breed of versatile Big Data analysts and developers, who are thoroughly conversant with the basic and advanced analytic techniques for manipulating and analysing data.

About the Author

DT Editorial Services

DT Editorial Services has seized the market of computer books

Table of Contents

Introduction

Book Preview

Chapter 1: Getting an Overview of Big Data

- What is Big Data?
- History of Data Management – Evolution of Big Data
- Structuring Big Data
- Elements of Big Data
- Big Data Analytics
- Careers in Big Data
- Future of Big Data
- Summary
- Quick Revise

Chapter 2: Exploring the Use of Big Data in Business Context

- Use of Big Data in Social Networking
- Use of Big Data in Preventing Fraudulent Activities
- Use of Big Data in Detecting Fraudulent Activities in Insurance Sector
- Use of Big Data in Retail Industry
- Summary

- Quick Revise

Chapter 3: Introducing Technologies for Handling Big Data

- Distributed and Parallel Computing for Big Data
- Introducing Hadoop
- Cloud Computing and Big Data
- In-Memory Computing Technology for Big Data
- Summary
- Quick Revise

Chapter 4: Understanding Hadoop Ecosystem

- Hadoop Ecosystem
- Hadoop Distributed File System
- MapReduce
- Hadoop YARN
- Introducing HBase
- Combining HBase and HDFS
- Hive
- Pig and Pig Latin
- Sqoop
- ZooKeeper
- Flume
- Oozie
- Summary
- Quick Revise

Chapter 5: Understanding MapReduce Fundamentals and HBase

- The MapReduce Framework
- Techniques to Optimize MapReduce Jobs
- Uses of MapReduce
- Role of HBase in Big Data Processing
- Summary
- Quick Revise

Chapter 6: Understanding Big Data Technology Foundations

- Exploring the Big Data Stack
- Virtualization and Big Data
- Virtualization Approaches
- Summary
- Quick Revise

Chapter 7: Storing Data in Databases and Data Warehouses

- RDBMS and Big Data
- Non-Relational Database
- Polyglot Persistence
- Integrating Big Data with Traditional Data Warehouses
- Big Data Analysis and Data Warehouse

- Changing Deployment Models in Big Data Era
- Summary
- Quick Revise

Chapter 8: Processing Your Data with MapReduce

- Recollecting the Concept of MapReduce Framework
- Developing Simple MapReduce Application
- Points to Consider while Designing MapReduce
- Summary
- Quick Revise

Chapter 9: Customizing MapReduce Execution and Implementing MapReduce Program

- Controlling MapReduce Execution with InputFormat
- Reading Data with Custom RecordReader
- Organizing Output Data with OutputFormats
- Customizing Data with RecordWriter
- Optimizing MapReduce Execution with Combiner
- Controlling Reducer Execution with Partitioners
- Customizing the MapReduce Execution in Terms of YARN
- Implementing a MapReduce Program for Sorting Text Data
- Summary
- Quick Revise

Chapter 10: Testing and Debugging MapReduce Applications

- Debugging Hadoop MapReduce Locally
- Performing Unit Testing for MapReduce Applications
- Performing Local Application Testing with Eclipse
- Logging for Hadoop Testing
- Application Log Processing
- Defensive Programming in MapReduce
- Summary
- Quick Revise

Chapter 11: Understanding Hadoop YARN Architecture

- Background of YARN
- Advantages of YARN
- YARN Architecture
- Working of YARN
- YARN Schedulers
- Backward Compatibility with YARN
- YARN Configurations
- YARN Commands
- YARN Containers
- Registry
- Log Management in Hadoop 1
- Summary
- Quick Revise

Chapter 12: Exploring Hive

- Introducing Hive
- Getting Started with Hive
- Hive Services
- Data Types in Hive
- Built-In Functions in Hive
- Hive DDL
- Data Manipulation in Hive
- Data Retrieval Queries
- Using JOINS in Hive
- Summary
- Quick Revise

Chapter 13: Analyzing Data with Pig

- Introducing Pig
- Running Pig
- Getting Started with Pig Latin
- Working with Operators in Pig
- Debugging Pig
- Working with Functions in Pig
- Error Handling in Pig
- Summary
- Quick Revise

Chapter 14: Using Oozie

- Introducing Oozie
- Installing and Configuring Oozie
- Understanding the Oozie Workflow
- Oozie Coordinator
- Oozie Bundle
- Oozie Parameterization with EL
- Oozie Job Execution Model
- Accessing Oozie
- Oozie SLA
- Summary
- Quick Revise

Chapter 15: NoSQL Data Management

- Introduction to NoSQL
- Types of NoSQL Data Models
- Schema-Less Databases
- Materialized Views
- Distribution Models
- Sharding
- Summary
- Quick Revise

Chapter 16: Data Movement with Flume and Sqoop

- Flume Architecture
- Sqoop
- Importing Data
- Sqoop2 vs Sqoop
- Summary
- Quick Revise

Chapter 17: Introduction to Mahout

- What is Mahout?
- Machine Learning
- Collaborative Filtering (Recommendation)
- Clustering
- Classification
- Mahout Algorithms
- Environment for Mahout
- Summary
- Quick Revise

Chapter 18: Understanding Analytics and Big Data

- Comparing Reporting and Analysis
- Types of Analytics
- Points to Consider during Analysis
- Developing an Analytic Team
- Understanding Text Analytics
- Summary
- Quick Revise

Chapter 19: Analytical Approaches and Tools to Analyze Data

- Analytical Approaches
- History of Analytical Tools
- Introducing Popular Analytical Tools
- Comparing Various Analytical Tools
- Installing R
- Installing RStudio
- Summary
- Quick Revise

Chapter 20: Exploring R

- Exploring Basic Features of R
- Exploring RGUI
- Exploring RStudio
- Handling Basic Expressions in R
- Variables in R
- Working with Vectors
- Storing and Calculating Values in R

- Creating and Using Objects
- Interacting with Users
- Handling Data in R Workspace
- Executing Scripts
- Creating Plots
- Accessing Help and Documentation in R
- Summary
- Quick Revise

Chapter 21: Reading Datasets and Exporting Data from R

- Using the c() Command
- Using the scan() Command
- Reading Multiple Data Values from Large Files
- Reading Data from RStudio
- Exporting Data from R
- Summary
- Quick Revise

Chapter 22: Manipulating and Processing Data in R

- Selecting the Most Appropriate Data Structure
- Creating Data Subsets
- Merging Datasets in R
- Sorting Data
- Putting Your Data into Shape
- Managing Data in R Using Matrices
- Managing Data in R Using Data Frames
- Summary
- Quick Revise

Chapter 23: Working with Functions and Packages in R

- Using Functions Instead of Scripts
- Using Arguments in Functions
- Built-in Functions in R
- Introducing Packages
- Working with Packages
- Summary
- Quick Revise

Chapter 24: Performing Graphical Analysis in R

- Using Plots
- Saving Graphs to External Files
- Advanced Features of R
- Summary
- Quick Revise

Chapter 25: Integrating R and Hadoop and Understanding Hive

- RHadoop—An Integration of R and Hadoop

- Text Mining in RHadoop
- Data Analysis Using the MapReduce Technique in RHadoop
- Data Mining in Hive
- Summary
- Quick Revise

Chapter 26: Data Visualization-I

- Ways of Representing Visual Data
- Techniques Used for Visual Data Representation
- Types of Data Visualization
- Applications of Data Visualization
- Visualizing Big Data
- Tools Used in Data Visualization
- Tableau Products
- Summary
- Quick Revise

Chapter 27: Data Visualization with Tableau (Data Visualization-II)

- Introduction to Tableau Software
- Tableau Desktop Workspace
- Data Analytics in Tableau Public
- Using Visual Controls in Tableau Public
- Overview of Tableau 9.0
- Summary
- Quick Revise

Chapter 28: Social Media Analytics and Text Mining

- Introducing Social Media
- Introducing Key Elements of Social Media
- Introducing Text Mining
- Understanding Text Mining Process
- Sentiment Analysis
- Performing Social Media Analytics and Opinion Mining on Tweets
- Summary
- Quick Revise

Chapter 29: Mobile Analytics

- Introducing Mobile Analytics
- Introducing Mobile Analytics Tools
- Performing Mobile Analytics
- Challenges of Mobile Analytics
- Summary
- Quick Revise

Chapter 30: Finding a Job in the Big Data Market

- Importance and Scope of Big Data Jobs
- Big Data Opportunities

- Skill Assessment for Big Data Jobs
- Roles and Responsibilities in Big Data Jobs
- Gaining a Foothold in the Big Data Market
- Basic Educational Requirements for Big Data Jobs
- Basic Technological Requirements for Big Data Jobs
- Tools Supporting Big Data
- Consultants and In-House Specialists in Big Data
- Tactics for Searching Big Data Jobs
- Preparing for Interviews
- Obtaining Big Data Jobs through Social Media
- Summary
- Quick Revise

Big Data Practical

Appendix A: Cassandra

Appendix B: ZooKeeper

Index

To purchase this product, please visit:
https://www.wileyindia.com/big-data-black-book-covers-hadoop-2-mapreduce-hive-yarn-pig-r-and-data-visualization-completely-revised-and-upgraded.html

Scan to buy