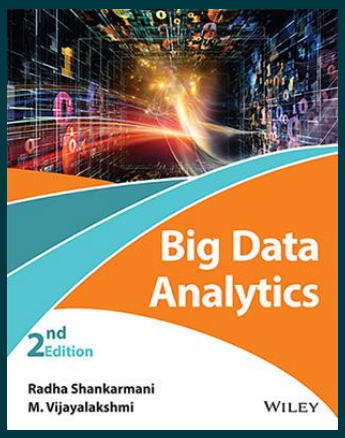
WILEY



Big Data Analytics, 2ed

By Radha Shankarmani, M. Vijayalakshmi

Paperback 9788126565757 ₹699.00

Pages: 336

Description

The goal of this book is to cover foundational techniques and tools required for Big Data Analytics. It focuses on concepts, principles and techniques applicable to any technology environment and industry and establishes a baseline that can be enhanced further by additional real-world experience. This book aims to be a ready reckoner to either a novice or a professional working in the field. Topics covered include Hadoop, MapReduce, Association Rules, Large-Scale Supervised Machine Learning, Data Streams, Clustering, NoSQL systems (Pig, Hive) and Applications including Recommendation Systems, Web and Security.

About the Author

Radha Shankarmani, M. Vijayalakshmi

Dr. Radha Shankarmani is currently working as Professor and Head at Department of Information Technology

Table of Contents

Preface

Acknowledgements

About the Authors

Chapter 1 Big Data Analytics

1.1 Introduction to Big Data

1.2 Big Data Characteristics

1.3 Types of Big Data

1.4 Traditional Versus Big Data Approach

1.5 Technologies Available for Big Data

1.6 Infrastructure for Big Data

1.7 Use of Data Analytics

1.8 Big Data Challenges

1.9 Desired Properties of a Big Data System

1.10 Case Study of Big Data Solutions

Chapter 2 Hadoop

2.1 Introduction

2.2 What is Hadoop?

2.3 Core Hadoop Components
2.4 Hadoop Ecosystem
2.5 Hive
2.6 Physical Architecture
2.7 Hadoop Limitations
Chapter 3 What is NoSQL?
3.1 What is NoSQL?
3.2 NoSQL Business Drivers
3.3 NoSQL Case Studies
3.4 NoSQL Data Architectural Patterns
3.5 Variations of NoSQL Architectural Patterns
3.6 Using NoSQL to Manage Big Data
Chapter 4 MapReduce
4.1 MapReduce and The New Software Stack
4.2 MapReduce
4.3 Algorithms Using MapReduce
Chapter 5 Finding Similar Items
5.1 Introduction
5.2 Nearest Neighbor Search
5.3 Applications of Nearest Neighbor Search
5.4 Similarity of Documents
5.5 Collaborative Filtering as a Similar-Sets Problem
5.6 Recommendation Based on User Ratings
5.7 Distance Measures
Chapter 6 Mining Data Streams
6.1 Introduction
6.2 Data Stream Management Systems
6.3 Data Stream Mining
6.4 Examples of Data Stream Applications
6.5 Stream Queries
6.6 Issues in Data Stream Query Processing
6.7 Sampling in Data Streams
6.8 Filtering Streams
6.10 Querying on Windows − Counting Ones in a Window
6.11 Decaying Windows
Chapter 7 Link Analysis
7.1 Introduction
7.2 History of Search Engines and Spam

7.3 PageRank
7.4 Efficient Computation of PageRank
7.5 Topic-Sensitive PageRank
7.6 Link Spam
7.7 Hubs and Authorities
Chapter 8 Frequent Itemset Mining
8.1 Introduction
8.2 Market-Basket Model
8.3 Algorithm for Finding Frequent Itemsets
8.4 Handling Larger Datasets in Main Memory
8.5 Limited Pass Algorithms
8.6 Counting Frequent Items in a Stream
Chapter 9 Clustering Approaches
9.1 Introduction
9.2 Overview of Clustering Techniques
9.3 Hierarchical Clustering
9.4 Partitioning Methods
9.5 The CURE Algorithm
9.6 Clustering Streams
Chapter 10 Recommendation Systems
10.1 Introduction
10.2 A Model for Recommendation Systems
10.3 Collaborative-Filtering System
10.4 Content-Based Recommendations
Chapter 11 Mining Social Network Graphs
11.1 Introduction
11.2 Applications of Social Network Mining
11.3 Social Networks as a Graph
11.4 Types of Social Networks
11.5 Clustering of Social Graphs
11.6 Direct Discovery of Communities in a Social Graph
11.7 SimRank
11.8 Counting Triangles in a Social Graph
Summary
Exercises
Programming Assignments
References
Appendix

Index