

"This text should be required reading for everyone in contemporary business."

—Peter Woodhull, CEO, Modus21

"The one book that clearly describes and links Big Data concepts to business utility."

—Dr. Christopher Starr, PhD

"Simply, this is the best Big Data book on the market!"

—Sam Rostam, Cascadian IT Group

"...one of the most contemporary approaches I've seen to Big Data fundamentals..."

—Joshua M. Davis, PhD

Big Data Fundamentals

Concepts, Drivers & Techniques

Co-authored and Edited by Best-Selling Author Thomas Erl
Co-authored by Wajid Khattak and Paul Buhler, PhD

Big Data Fundamentals

Big Data Fundamentals: Concepts, Drivers & Techniques

Table of Contents

Cover

Title Page

Copyright Page

Contents

Acknowledgments

Reader Services

PART I: THE FUNDAMENTALS OF BIG DATA

CHAPTER 1: Understanding Big Data

Concepts and Terminology

Datasets

Data Analysis

Data Analytics

Descriptive Analytics

Diagnostic Analytics

Predictive Analytics

Prescriptive Analytics

Business Intelligence (BI)

Key Performance Indicators (KPI)

Big Data Characteristics

Volume

Velocity

Variety

Veracity

Value

Different Types of Data



Pearson

Table of Contents

Structured Data

Unstructured Data

Semi-structured Data

Metadata

Case Study Background

History

Technical Infrastructure and Automation Environment

Business Goals and Obstacles

Case Study Example

Identifying Data Characteristics

Volume

Velocity

Variety

Veracity

Value

Identifying Types of Data

CHAPTER 2: Business Motivations and Drivers for Big Data Adoption

Marketplace Dynamics

Business Architecture

Business Process Management

Information and Communications Technology

Data Analytics and Data Science

Digitization

Affordable Technology and Commodity Hardware

Social Media

Hyper-Connected Communities and Devices

Cloud Computing

Internet of Everything (IoE)

Case Study Example

CHAPTER 3: Big Data Adoption and Planning Considerations

Organization Prerequisites

Table of Contents

Data Procurement

Privacy

Security

Provenance

Limited Realtime Support

Distinct Performance Challenges

Distinct Governance Requirements

Distinct Methodology

Clouds

Big Data Analytics Lifecycle

- Business Case Evaluation

- Data Identification

- Data Acquisition and Filtering

- Data Extraction

- Data Validation and Cleansing

- Data Aggregation and Representation

- Data Analysis

- Data Visualization

- Utilization of Analysis Results

Case Study Example

- Big Data Analytics Lifecycle

- Business Case Evaluation

- Data Identification

- Data Acquisition and Filtering

- Data Extraction

- Data Validation and Cleansing

- Data Aggregation and Representation

- Data Analysis

- Data Visualization

- Utilization of Analysis Results

CHAPTER 4: Enterprise Technologies and Big Data Business Intelligence

Table of Contents

Online Transaction Processing (OLTP)

Online Analytical Processing (OLAP)

Extract Transform Load (ETL)

Data Warehouses

Data Marts

Traditional BI

Ad-hoc Reports

Dashboards

Big Data BI

Traditional Data Visualization

Data Visualization for Big Data

Case Study Example

Enterprise Technology

Big Data Business Intelligence

PART II: STORING AND ANALYZING BIG DATA

CHAPTER 5: Big Data Storage Concepts

Clusters

File Systems and Distributed File Systems

NoSQL

Sharding

Replication

Primary-Secondary

Peer-to-Peer

Sharding and Replication

Combining Sharding and Primary-Secondary Replication

Combining Sharding and Peer-to-Peer Replication

CAP Theorem

ACID

BASE

Case Study Example

CHAPTER 6: Big Data Processing Concepts

Table of Contents

Parallel Data Processing

Distributed Data Processing

Hadoop

Processing Workloads

Batch

Transactional

Cluster

Processing in Batch Mode

Batch Processing with MapReduce

Map and Reduce Tasks

Map

Combine

Partition

Shuffle and Sort

Reduce

A Simple MapReduce Example

Understanding MapReduce Algorithms

Processing in Realtime Mode

Speed Consistency Volume (SCV)

Event Stream Processing

Complex Event Processing

Realtime Big Data Processing and SCV

Realtime Big Data Processing and MapReduce

Case Study Example

Processing Workloads

Processing in Batch Mode

Processing in Realtime

CHAPTER 7: Big Data Storage Technology

On-Disk Storage Devices

Distributed File Systems

RDBMS Databases

NoSQL Databases

Characteristics

Table of Contents

Rationale

Types

Key-Value

Document

Column-Family

Graph

NewSQL Databases

In-Memory Storage Devices

In-Memory Data Grids

Read-through

Write-through

Write-behind

Refresh-ahead

In-Memory Databases

Case Study Example

CHAPTER 8: Big Data Analysis Techniques

Quantitative Analysis

Qualitative Analysis

Data Mining

Statistical Analysis

A/B Testing

Correlation

Regression

Machine Learning

Classification (Supervised Machine Learning)

Clustering (Unsupervised Machine Learning)

Outlier Detection

Filtering

Semantic Analysis

Natural Language Processing

Text Analytics

Sentiment Analysis

Visual Analysis

Table of Contents

Heat Maps

Time Series Plots

Network Graphs

Spatial Data Mapping

Case Study Example

Correlation

Regression

Time Series Plot

Clustering

Classification

APPENDIX A: Case Study Conclusion

About the Authors

Thomas Erl

Wajid Khattak

Paul Buhler

Index