

A. Karunamurthy, R. Ramakrishnan ,  
V. Udhayakumar , and P. Rajapandian *Editors*

# Cloud Computing and Big Data



**DeepScience**

# Cloud Computing and Big Data

**A. Karunamurthy**  
**R. Ramakrishnan**  
**V. Udhayakumar**  
**P. Rajapandian**

Department of Master of Computer Applications, Sri  
Manakula Vinayagar Engineering College Puducherry,  
(Autonomous) Madagadipet, Puducherry 605107- India



**DeepScience**

*Published, marketed, and distributed by:*

Deep Science Publishing, 2025  
USA | UK | India | Turkey  
Reg. No. MH-33-0523625  
[www.deepscienceresearch.com](http://www.deepscienceresearch.com)  
[editor@deepscienceresearch.com](mailto:editor@deepscienceresearch.com)  
WhatsApp: +91 7977171947

ISBN: 978-93-7185-300-2

E-ISBN: 978-93-7185-679-9

<https://doi.org/10.70593/978-93-7185-679-9>

Copyright © A. Karunamurthy, R. Ramakrishnan, V. Udhayakumar, P. Rajapandian, 2025.

**Citation:** Karunamurthy, A., Ramakrishnan, R., Udhayakumar, V., & Rajapandian, P. (2025). *Cloud Computing and Big Data*. Deep Science Publishing. <https://doi.org/10.70593/978-93-7185-679-9>

This book is published online under a fully open access program and is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). This open access license allows third parties to copy and redistribute the material in any medium or format, provided that proper attribution is given to the author(s) and the published source. The publishers, authors, and editors are not responsible for errors or omissions, or for any consequences arising from the application of the information presented in this book, and make no warranty, express or implied, regarding the content of this publication. Although the publisher, authors, and editors have made every effort to ensure that the content is not misleading or false, they do not represent or warrant that the information-particularly regarding verification by third parties-has been verified. The publisher is neutral with regard to jurisdictional claims in published maps and institutional affiliations. The authors and publishers have made every effort to contact all copyright holders of the material reproduced in this publication and apologize to anyone we may have been unable to reach. If any copyright material has not been acknowledged, please write to us so we can correct it in a future reprint.

# Preface

Cloud computing and big data have emerged as two of the most transformative technologies in the 21st century, revolutionizing the way organizations store, manage, and process information. This book, "Cloud Computing and Big Data," has been designed to serve as a comprehensive resource for students, educators, and professionals who seek to understand the fundamental concepts, architectures, and evolving paradigms of cloud technologies and big data frameworks.

Chapter I, Introduction to Cloud Computing, lays the foundation by tracing the evolution of cloud computing, including developments in hardware, internet software, and virtualization. It introduces essential service models such as IaaS, PaaS, SaaS, and newer concepts like Communication-as-a-Service and Monitoring-as-a-Service. This chapter also focuses on building and understanding cloud network infrastructure.

Chapter II, Cloud Information Systems, explores the integration and management of information within the cloud. It emphasizes the significance of cloud federation, presence, privacy, and security, while also addressing standardization and end-user access—critical for modern cloud-based information systems.

Chapter III, Cloud Infrastructure, transitions into the structural aspects of cloud environments. It discusses utility models, evolving software applications, and IT infrastructure, highlighting key technologies such as virtualization, hyper-threading, and blade servers. Standards, automation, and application management are also addressed to provide a thorough understanding of data center optimization and challenges.

Chapter IV, Introduction to Big Data and Hadoop, introduces the concept of big data and its analytics. It traces the historical development of Hadoop, its core components, and tools like Hadoop Streaming and IBM's Big Insights. This chapter serves as a bridge between cloud computing and big data analytics, offering hands-on approaches to data analysis using both Unix tools and the Hadoop framework.

Chapter V, HDFS and MapReduce, dives deep into Hadoop's architectural components. It covers HDFS design, data ingest tools like Flume and Sqoop, data formats, and the MapReduce programming model. Students will learn how data flows in Hadoop, how jobs are scheduled and executed, and how to optimize MapReduce applications for real-world data processing scenarios.

This book aligns with the latest curriculum and learning outcomes (CO1 to CO4) and is intended not only to build theoretical knowledge but also to offer practical insights and applications of cloud and big data technologies. We hope this book empowers learners to confidently navigate and contribute to the evolving digital landscape.

Dr. A. Karunamurthy  
Mr. R. Ramakrishnan  
Mr. V. Udhayakumar  
Mr. P. Rajapandian

# Table of Content

CHAPTER I: INTRODUCTION.....1

CHAPTER II: CLOUD INFORMATION SYSTEMS.....39

CHAPTER III: CLOUD INFRASTRUCTURE.....51

CHAPTER IV: INTRODUCTION TO BIG DATA AND HADOOP.....113

CHAPTER V: HADOOP DISTRIBUTED FILE SYSTEM (HDFS) & MAP

REDUCE .....152