PREFACE

The massive amount of digital data currently being generated in all the human activities is a precious source of knowledge to both business and science. However, handling and analyzing huge datasets requires very large storage resources and scalable computing facilities. In fact, the large availability of big data sources demands for efficient data analysis tools and techniques for finding and extracting useful knowledge from them. Big data analysis today can be performed by storing data and running compute-intensive data mining algorithms on cloud computing systems to extract value from data in reduced time. Cloud computing systems can be used to run complex applications on dynamic computing servers and deliver them as services over the Internet. According to their elastic nature, cloud computing infrastructures can serve as effective platforms for addressing the computational and data storage needs of most big data analytics applications that are being developed nowadays. Coping with and gaining value from cloud-based big data, however, requires novel software tools and advanced analysis techniques. Indeed, advanced data mining techniques and innovative tools can help users to understand and extract what is useful in large and complex datasets; and the knowledge extracted from big data sources today is vital in making informed decisions in many business and scientific applications. This process, which constitutes the base for allowing the analysis of big data sources and repositories, must be implemented by combining big data analytics and knowledge discovery techniques with scalable computing systems such as clouds.

All these issues are discussed in this book. In fact, the main goal of the book is to introduce and present models, methods, techniques, and systems useful to analyze large digital data sources by using the computing and storage facilities of cloud computing systems. This book includes, as key topics, scalable data mining and knowledge discovery techniques, together with cloud computing concepts, models, and systems. After introducing these fields, this book focuses on scalable technologies for cloud-based data analysis such as MapReduce, workflows, and NoSQL models, and discusses how to design high-performance

distributed analysis of big data on clouds. Finally, this book examines research trends such as big data exascale computing, and massive social network analysis.

This book is for graduate students, researchers, and professionals in cloud computing, big data analysis, distributed data mining, and data analytics. Both readers who are beginners to the subjects and those experienced in the cloud computing and data mining domains will find many topics of interest. Researchers will find some of the latest achievements in the area and significant technologies and examples on the state-of-the-art in cloud-based data analysis and knowledge discovery. Furthermore, graduate students and young researchers will learn useful concepts related to parallel and distributed data mining, cloud computing, data-intensive applications, and scalable data analysis.

Other than introducing the key concepts and systems in the area of cloud-based data analysis, this book presents real case studies that provide a useful guide for developers on issues, prospects, and successful approaches in the practical use of cloud-based data analysis frameworks. In this book, the chapters are presented in a way so that the book could also be used as reference text in graduate and postgraduate courses, in parallel/distributed data mining and in cloud computing for big data analysis.

We would like to thank people from the publisher, Elsevier, particularly Lindsay Lawrence, for their support and work during the book publication process.

We hope readers will find this book's content interesting, attractive, and useful, as we found it stimulating and exciting to write.

<div align="right">

Domenico Talia
Paolo Trunfio
Fabrizio Marozzo

</div>