Preface

In recent years, volumes of data produced by a variety of heterogeneous sources have increased around the world. As a result, both Information Technology and business users need to efficiently collect and process this huge amount of data in real time to discover relevant situations which will allow driving successful business decisions or actions. In this regard, big data management is an approach which helps to process this huge amount of data and is a reality for an increasing number of organizations in many areas and represents a set of challenges involving big data modeling, storage and retrieval, analysis and visualization. However, technological resources, people and processes are crucial to facilitate the management of big data in any kind of organization, allowing information and knowledge from a large volume of data to support decision-making. Big data management can be supported by these three dimensions: technology, people and processes. It normally focuses on database Management system of Big Data.

The main aim of this publication is to collect and compile the most representative approaches in current research which tackle the different faces of big data management. In this regard, readers will be able to acquire a panoramic overview of existing solutions for dealing with big amounts of data in real time in different development scopes. In this sense, the book would have a twofold purpose, not only as background reading on big data management, but also as a reference book in the search for dealing with big data management solutions to managing massive datasets problems.

ORGANIZATION OF THE BOOK

This book consists of 15 peer-reviewed invited chapters authored by several international researchers around the world.

Chapters

Chapter 1: Big Data – Challenges, Opportunities and Realities

This chapter discusses the emerging technology of modern era: Big Data with detailed description of the three Vs (Variety, Velocity and Volume). Further sections of this chapter will enable to understand the concepts of big data analytics, its application, advantages and limitations how this large amount of data is increasing rapidly as compared to the advancement in computing resources and which new

technologies and architectures are needed to extract value from it by capturing and analysis processes. The chapter also raised few research issues and future directions about Big Data.

Chapter 2: Introducing Data Structures for Big Data

This chapter covers very important and new aspects so called data structures for Big Data and to be regarded as a new subject in Big Data Science, not just as a new topic, considering the explosive momentum of the big data. Based upon the notion of the big data structures train and atrain, this chapter introduces the useful data structures for the programmers working with big data which are**:** homogeneous stacks 'train stack' and 'rT-coach stack', heterogeneous stacks 'atrain stack' and 'rA-coach stack', homogeneous queues 'train queue' and 'rT-coach queue', heterogeneous queues 'atrain queue' and 'rA-coach queue', homogeneous binary trees 'train binary tree' and 'rT-coach binary tree', heterogeneous binary trees 'atrain binary tree' and 'rA-coach binary tree', homogeneous trees 'train tree' and 'rT-coach tree', heterogeneous trees 'atrain tree' and 'rA-coach tree', to enrich the subject 'Data Structures for Big Data' for implementation of Big Data Management.

Chapter 3: Big Data Mining

This chapter explores various schemes that have been used to tackle the big databases. In this context, this chapter present big data mining plays a major role which is different from the traditional data mining as well as providing the process of extracting useful information from large datasets or streams of data, due to its volume, velocity, variety, validity, veracity, value and visibility. This chapter also discusses the big data mining tools and applications issues concerning big data Management.

Chapter 4: An Empirical Study of NoSQL Databases for Big Data

This chapter provides a demonstration of the ways in which author introduce NoSQL databases by using an empirical approach with MongoDB. The chapter is organized as follows:It starts with an introduction followed by Features and types of NoSQL Databases, big data management, technologies, and applications, big data generation, capturing, and collection, Big Data Storage and Preservation, Big Data Analytics, Management, Visualization, and Sharing. It also discusses the practical implementation of MongoDB database to manage Big Data Management.

Chapter 5: The Challenges of Data Cleansing with Data Warehouses

This chapter describes how data cleansing methods and related technologies can be used to which every organisation that incorporates a form of data processing or data mining. The chapter also discusses the necessary methods needed to process data at a high quality. It also classifies common problems which organisations face when cleansing data from a source or multiple sources while evaluating methods which aid in this process. Second, the new and different challenges faced at schema-level and instance-level are also explained and how they can be overcome. The chapter also introduces tools which provide data

cleansing, but are limited due to the uniqueness of every data source and data warehouse. Outlined are the limitations of these tools and how human interaction (self-programming) may be needed to ensure vital data is not lost as well as discussing the importance of maintaining and removing data which has been stored for several years and may no longer have any value in perspectives of big data management.

Chapter 6: Big Data Analysis – Big Data Analysis Pipeline and Its Technical Challenges

This chapter presents a new approach in order to analyze the big data. By this approach a new class of systems had to be designed and implemented, giving rise to the new phenomenon of "Big Data". It discusses how the data-driven world has the potential to improve the efficiencies of enterprises and improve the quality of our lives and also there are a number of challenges that addressed to allow exploiting the full potential of Big Data. This chapter also highlights the key technical challenges of Big Data.

Chapter 7: Possibilities, Impediments, and Challenges for Network Security in Big Data

This chapter presents an analytical approach regarding possibilities, impediments and challenges for network security in Big Data. This chapter covers big data security management from concepts to real-world issues. It identifies the major challenges, industry trends, security management framework, Possibilities for Network Security in Big Data, Impediments for Network Security in Big Data, Existing tools and Techniques best for Big Data Security solutions in solving security problems, current research results, and future research issues.

Chapter 8: Mastering Big Data in the Digital Age

This chapter explains the overview of big data; the volume, velocity, variety, veracity, and variability of big data; the privacy and security of big data applications; big data and multimedia utilization; the concept of MapReduce; the concept of Hadoop; big data and data mining; big data and cloud computing; the applications of big data in health care industry; the applications of big data analytics in tourism and hospitality industry; and the challenges and implications of big data in the digital age. The chapter also discusses that how big data has the potential to increase organizational performance and gain sustainable competitive advantage in the digital age.

Chapter 9: Legal Responses to the Commodification of Personal Data in the Era of Big Data – The Paradigm Shift from Data Protection towards Data Ownership

The chapter presents a succinct overview of the legal ownership of big data by examining the key players in control of the information at each stage of the processing and managing of big data. This chapter also covers and describes the current legislative framework with regard to data protection and addition, techno-legal solutions offered to complement the law of big data in this respect.

Chapter 10: Big Data and Business Decision Making

This chapter analyzes the possibilities of big data to improve the services offered by companies and the customer experience and increase the efficiency of these companies. It also examines some special aspects associated with the use of big data such as the issues of data privacy and compliance with the regulations on the use of the information.

Chapter 11: Using Big Data to Improve the Educational Infrastructure and Learning Paradigm

This chapter proposes a generic model that uses the techniques of educational data mining to explore and analyze Big Data being generated by the educational sector. This chapter also examines and discusses the various questions that can be answered using educational data mining methods and how the discovered patterns can be used to enrich the learning experience of a student as well as help teachers make pedagogical decisions.

Chapter 12: Vehicle to Cloud – Big Data for Environmental Sustainability, Energy, and Traffic Management

This chapter explains the combined version of IoT and vehicles to create a V2C vehicle to cloud system that will create the big data for environmental sustainability, energy and traffic management by different technical and political views and aspects. There is a need to elaborate about the characteristics of IoT based data to find out the available and applicable solutions. Such kind of study also directs to realize the need of new techniques to cope up with such challenges.

Chapter 13: Point Cloud Manager – Applications of a Middleware for Managing Huge Point Clouds

This chapter presents a series of applications built on top of Point Cloud Manager (PCM), a middleware that provides an abstraction for point clouds with arbitrary attached data and makes it easy to perform out-of-core operations on them on commodity CPUs and GPUs. The chapter covers different kinds of real world applications are tackled, showing both real-time and offline examples, and in addition, render-oriented and computation-related operations as well.

Chapter 14: Big Data Management in Financial Services

This chapter discusses and provides a demonstration of big data management in financial services. The chapter is organized as follows: introduction, benefit of Big Data for financial institutions, main challenges for implementation of Big Data, smart data, storage and processing, the risks of Big Data.

Chapter 15: Recommender System in the Context of Big Data – Implementing SVD-Based Recommender System using Apache Hadoop and Spark

This chapter surveys the literature of RSs, reviews the current state of RSs with the main concerns surrounding them due to Big Data, investigates thoroughly SVD and provides an implementation to it using Apache Hadoop and Spark. This chapter also intended to validate the applicability of, existing contributions to the field of, SVD-based RSs as well as validated the effectiveness of Hadoop and spark in developing large-scale systems. The results proved the scalability of SVD-based RS and its applicability to Big Data.

Manoj Kumar Singh
Adama Science and Technology University, Ethiopia

Dileep Kumar G.
Adama Science and Technology University, Ethiopia