

Preface

Rapid developments in communication and computing technologies have been the driving factors in the spread of the internet technology. This technology is able to scale up and reach out to more and more people. People at opposite sides of the globe are able to remain connected to each other because of the connectivity that the internet is able to provide now. Getting people together through the internet has become more realistic than getting them together physically at one place. This has led to the emergence of cyber society, a form of human society that we are heading for with great speed. As is expected, this has also affected different activities from education to entertainment, culture to commerce, goodness (ethics, spiritual) to governance. The internet has become a platform of all types of human interactions. Services of different domains, designed for different walks of people, are being provided via the internet. Success of these services decisively depends on understanding people and their behaviour over the internet. For example, people may like a particular kind of service due to many desired features the service has. Features could be quality of service like response time, average availability, trust and similar factors. So service providers would like to know of consumer preferences and requirements for designing a service, so as to get maximum returns on investment. On the other side, customers would require enough information to select the best service provider for their needs. Thus, decision-making is key to cyber society. And, informed decisions can only be made on the basis of good information, i.e. information that is both qualitatively and quantitatively sufficient for decision-making.

Fortunately for cyber society, through our presence on the internet, we generate enough data to garner a lot of meaningful information and patterns. This information is in the form of metaphorical due to footsteps or breadcrumbs that we leave on the internet through our various activities. For example, social networking services, e-businesses and search engines generate huge data sets every second of the day. And these data sets are not only voluminous but also in various forms such as picture, text and audio. This great quantum of data sets is collectively christened *big data* and is identified by its three special features *velocity*, *variety* and *volume*.

Collection and processing of big data are topics that have drawn considerable attention of concerned variety of people ranging from researchers to business makers. Developments in infrastructure such as grid and cloud technology have given a great impetus to big data services. Research in this area is focusing on big data as a service and infrastructure as a service. The former looks at developing algorithms for fast data access, processing as well as inferring pieces of information that remain hidden. To make all this happen, internet-based infrastructure must provide the backbone structures. It also needs an adaptable architecture that can be dynamically configured so that fast processing is possible by making use of optimal computing as well as storage resources. Thus, investigations on big data encompass many areas of research, including parallel and distributed computing, database management, software engineering, optimization and artificial intelligence. The rapid spread of the internet, several governments' decisions in making of smart cities and entrepreneurs' eagerness have invigorated the investigation on big data with intensity and speed. The efforts made in this book are directed towards the same purpose.

Goals of the Book

The goal of this book is to highlight the issues related to research and development in big data. For this purpose, the chapter authors are drawn from academia as well as industry. Some of the authors are actively engaged in the development of products and customized big data applications. A comprehensive view on six key issues is presented in this book. These issues are big data management, algorithms for distributed processing and mining patterns, management of security and privacy of big data, SLA for big data service and, finally, big data analytics encompassing several useful domains of applications. However, the issues included here are not completely exhaustive, but the coverage is enough to unfold the research as well as development promises the area holds for the future. Again for the purpose, the Introduction provides a survey with several important references. Interested readers are encouraged to take the lead following these references.

Intended Audience

This book promises to provide insights to readers having varied interest in big data. It covers an appreciable spread of the issues related to big data and every chapter intends to motivate readers to find the specialities and the challenges lie within. Of course, this is not a claim that each chapter deals an issue exhaustively. But, we sincerely hope that both conversant and novice readers will find this book equally interesting.

In addition to introducing the concepts involved, the authors have made attempts to provide a lead to realization of these concepts. With this aim, they have presented algorithms, frameworks and illustrations that provide enough hints towards system realization. For emphasizing growing trends on big data application, the book includes a chapter which discusses such systems available on the public domain. Thus, we hope this book is useful for undergraduate students and professionals looking for an introduction to big data. For graduate students intending to take up research in this upcoming area, the chapters with advanced information will also be useful.

Organization of the Book

This book has seven chapters. Chapter “[Big Data: An Introduction](#)” provides a broad review of the issues related to big data. Readers new to this area are encouraged to read this chapter first before reading other chapters. However, each chapter is independent and self-complete with respect to the theme it addresses.

Chapter “[Big Data Architecture](#)” lays out a universal data architecture for reasoning with all forms of data. Fundamental to big data analysis is big data management. The ability to collect, store and make available for analysis the data in their native forms is a key enabler for the science of analysing data. This chapter discusses an iterative strategy for data acquisition, analysis and visualization.

Big data processing is a major challenge to deal with voluminous data and demanding processing time. It also requires dealing with distributed storage as data could be spread across different locations. Chapter “[Big Data Processing Algorithms](#)” takes up these challenges. After surveying solutions to these problems, the chapter introduces some algorithms comprising random walks, distributed hash tables, streaming, bulk synchronous processing and MapReduce paradigms. These algorithms emphasize the usages of techniques, such as bringing application to data location, peer-to-peer communications and synchronization, for increased performance of big data applications. Particularly, the chapter illustrates the power of the Map Reduce paradigm for big data computation.

Chapter “[Big Data Search and Mining](#)” talks of mining the information that big data implicitly carries within. Often, big data appear with patterns exhibiting the intrinsic relations they hold. Unearthed patterns could be of use for improving enterprise performances and strategic customer relationships and marketing. Towards this end, the chapter introduces techniques for big data search and mining. It also presents algorithms for social network clustering using the topology discovery technique. Further, some problems such as sentiment detection on processing text streams (like tweets) are also discussed.

Security is always of prime concern. Security lapses in big data could be higher due to its high availability. As these data are collected from different sources, the vulnerability for security attacks increases. Chapter “[Security and Privacy of Big Data](#)” discusses the challenges, possible technologies, initiatives by stakeholders and emerging trends with respect to security and privacy of big data.

The world today, being instrumented by several appliances and aided by several internet-based services, generates very high volume of data. These data are useful for decision-making and furthering quality of services for customers. For this, data service is provided by big data infrastructure to receive requests from users and to accordingly provide data services. These services are guided by Service Level Agreement (SLA). Chapter “[Big Data Service Agreement](#)” addresses issues on SLA specification and processing. It also introduces needs for negotiation to avail data services. This chapter proposes a framework for SLA processing.

Chapter “[Applications of Big Data](#)” introduces applications of big data in different domains including banking and financial services. It sketches scenarios for the digital marketing space.

Acknowledgments

The genesis of this book goes to 11th International Conference on Distributed Computing and internet Technology (ICDCIT) held in February 2015. Big data was a theme for industry symposium held as a prelude to the main conference. The authors of three chapters in this book presented their ideas at the symposium. Editors took the feedback from participants and conveyed the same to the chapter authors for refining their contents.

In preparation of this book, we received help from different quarters. Hrushikesh Mohanty expresses his sincere thanks to the School of Computer and Information Sciences, University of Hyderabad, for providing excellent environment for carrying out this work. I also extend my sincere thanks to Dr. Achyuta Samanta, Founder KIIT University, for his inspiration and graceful support for hosting the ICDCIT series of conferences. Shri. D.N. Dwivedy of KIIT University deserves special thanks for making it happen. The help from ICDCIT organizing committee members of KIIT University is thankfully acknowledged. Deepak Chenthati and Prachet Bhuyan extend their thanks to their respective organizations Teradata India Pvt. Ltd. and KIIT University. Thanks to Shri Abhayakumar, graduate student of SCIS, University of Hyderabad, for his help in carrying out some pressing editing work.

Our special thanks to chapter authors who, despite their busy schedules, contributed chapters for this book. We are also thankful to Springer for publishing this book. In Particular, for their support and consideration for the issues we have been facing while preparing the manuscript.

Hyderabad
March 2015

Hrushikesh Mohanty
Prachet Bhuyan
Deepak Chenthati

Big Data

A Primer

Mohanty, H.; Bhuyan, P.; Chenthati, D. (Eds.)

2015, XIV, 184 p. 48 illus., 42 illus. in color., Hardcover

ISBN: 978-81-322-2493-8