

# Preface

Big Data applications are growing very rapidly around the globe. This new approach to decision making takes into account data gathered from multiple sources. Here my goal is to show how these diverse sources of data are useful in arriving at actionable information. In this collection of articles the publisher and I are trying to bring in one place several diverse applications of Big Data. The goal is for users to see how a Big Data application in another field could be replicated in their discipline. With this in mind I have assembled in the “Guide to Big Data Applications” a collection of 19 chapters written by academics and industry practitioners globally. These chapters reflect what Big Data is, how privacy can be protected with Big Data and some of the important applications of Big Data in science, medicine and business. These applications are intended to be representative and not exhaustive. For nearly two years I spoke with major researchers around the world and the publisher. These discussions led to this project. The initial Call for Chapters was sent to several hundred researchers globally via email. Approximately 40 proposals were submitted. Out of these came commitments for completion in a timely manner from 20 people. Most of these chapters are written by researchers while some are written by industry practitioners. One of the submissions was not included as it could not provide evidence of use of Big Data. This collection brings together in one place several important applications of Big Data. All chapters were reviewed using a double-blind process and comments provided to the authors. The chapters included reflect the final versions of these chapters.

I have arranged the chapters in four parts. Part I includes four chapters that deal with basic aspects of Big Data and how privacy is an integral component. In this part I include an introductory chapter that lays the foundation for using Big Data in a variety of applications. This is then followed with a chapter on the importance of including privacy aspects at the design stage itself. This chapter by two leading researchers in the field shows the importance of Big Data in dealing with privacy issues and how they could be better addressed by incorporating privacy aspects at the design stage itself. The team of researchers from a major research university in the USA addresses the importance of federated Big Data. They are looking at the use of distributed data in applications. This part is concluded with a chapter that

shows the importance of word embedding and natural language processing using Big Data analysis.

In Part II, there are eight chapters on the applications of Big Data in science. Science is an important area where decision making could be enhanced on the way to approach a problem using data analysis. The applications selected here deal with Environmental Science, High Performance Computing (HPC), friendship paradox in noting which friend's influence will be significant, significance of using encrypted search with Big Data, importance of deduplication in Big Data especially when data is collected from multiple sources, applications in Oil & Gas and how decision making can be enhanced in identifying bridges that need to be replaced as part of meeting safety requirements. All these application areas selected for inclusion in this collection show the diversity of fields in which Big Data is used today. The Environmental Science application shows how the data published by the National Oceanic and Atmospheric Administration (NOAA) is used to study the environment. Since such datasets are very large, specialized tools are needed to benefit from them. In this chapter the authors show how Big Data tools help in this effort. The team of industry practitioners discuss how there is great similarity in the way HPC deals with low-latency, massively parallel systems and distributed systems. These are all typical of how Big Data is used using tools such as MapReduce, Hadoop and Spark. Quora is a leading provider of answers to user queries and in this context one of their data scientists is addressing how the Friendship paradox is playing a significant part in Quora answers. This is a classic illustration of a Big Data application using social media.

Big Data applications in science exist in many branches and it is very heavily used in the Oil and Gas industry. Two chapters that address the Oil and Gas application are written by two sets of people with extensive industry experience. Two specific chapters are devoted to how Big Data is used in deduplication practices involving multimedia data in the cloud and how privacy-aware searches are done over encrypted data. Today, people are very concerned about the security of data stored with an application provider. Encryption is the preferred tool to protect such data and so having an efficient way to search such encrypted data is important. This chapter's contribution in this regard will be of great benefit for many users. We conclude Part II with a chapter that shows how Big Data is used in noting the structural safety of nation's bridges. This practical application shows how Big Data is used in many different ways.

Part III considers applications in medicine. A group of expert doctors from leading medical institutions in the Bay Area discuss how Big Data is used in the practice of medicine. This is one area where many more applications abound and the interested reader is encouraged to look at such applications. Another chapter looks at how data scientists are important in analyzing medical data. This chapter reflects a view from Asia and discusses the roadmap for data science use in medicine. Smoking has been noted as one of the leading causes of human suffering. This part includes a chapter on comorbidity aspects related to smokers based on a Big Data analysis. The details presented in this chapter would help the reader to focus on other possible applications of Big Data in medicine, especially cancer. Finally, a chapter is

included that shows how scientific analysis of Big Data helps with epileptic seizure prediction and control.

Part IV of the book deals with applications in Business. This is an area where Big Data use is expected to provide tangible results quickly to businesses. The three applications listed under this part include an application in banking, an application in marketing and an application in Quick Serve Restaurants. The banking application is written by a group of researchers in Europe. Their analysis shows that the importance of identifying financial fraud early is a global problem and how Big Data is used in this effort. The marketing application highlights the various ways in which Big Data could be used in business. Many large business sectors such as the airlines industry are using Big Data to set prices. The application with respect to a Quick Serve Restaurant chain deals with the impact of Yelp ratings and how it influences people's use of Quick Serve Restaurants.

As mentioned at the outset, this collection of chapters on Big Data applications is expected to serve as a sample for other applications in various fields. The readers will find novel ways in which data from multiple sources is combined to derive benefit for the general user. Also, in specific areas such as medicine, the use of Big Data is having profound impact in opening up new areas for exploration based on the availability of large volumes of data. These are all having practical applications that help extend people's lives. I earnestly hope that this collection of applications will spur the interest of the reader to look at novel ways of using Big Data.

This book is a collective effort of many people. The contributors to this book come from North America, Europe and Asia. This diversity shows that Big Data is a truly global way in which people use the data to enhance their decision-making capabilities and to derive practical benefits. The book greatly benefited from the careful review by many reviewers who provided detailed feedback in a timely manner. I have carefully checked all chapters for consistency of information in content and appearance. In spite of careful checking and taking advantage of the tools provided by technology, it is highly likely that some errors might have crept in to the chapter content. In such cases I take responsibility for such errors and request your help in bringing them to my attention so that they can be corrected in future editions.

Houston, TX, USA  
December 15, 2016

S. Srinivasan



<http://www.springer.com/978-3-319-53816-7>

Guide to Big Data Applications

Srinivasan, S. (Ed.)

2018, XVII, 565 p. 205 illus., 155 illus. in color.,

Hardcover

ISBN: 978-3-319-53816-7