Preface

Management, storage and information retrieval from huge amount of data require a paradigm shift in the way data is handled. This huge amount of structured, semi-structured or unstructured data can be termed as Big Data. Big Data can be characterized by 5V and these V’s pose ultimate challenge in terms of storage, processing, computation, analysis, and management.

- **Volume**: This refers to the amount of data been generated from different sources such as data logs from twitter, click streams of web pages and mobile apps, sensor-enabled equipment capturing data, etc.
- **Velocity**: This refers to the rate at which data is generated and received. For example for an effective marketing offer to a consumer, ecommerce applications combine mobile GPS location and personal preferences.
- **Variety**: This refers to various types of structured, unstructured and semi-structured data types. Unstructured data consist of files such as audio and video. Unstructured data has many of the requirements similar to that of structured data, such as summarization, audit ability, and privacy.
- **Value**: This refers to the intrinsic value that the data may possess, and must be discovered. There are various techniques to derive value from data. The advancement in the recent years have led to exponential decrease in the cost of storage and processing of data, thus providing statistical analysis on the entire data possible, unlike the past where random samples were analyzed to draw inferences.
- **Veracity**: This refers to the abnormality in data. Veracity in data analysis is one of the biggest challenges. This is dealt with by properly defining the problem statement before analysis, finding relevant data and using proven techniques for analysis so that the result is trustworthy and useful. There are various tools and techniques in the market for big data analytics. Hadoop is Java-based programming framework that supports processing of large data sets. It was started out as a project by Yahoo to analyze its data.
Operations management can be termed as the area of management concerned with designing, controlling and supervising the process of production and reconstruction of business operation in the production of goods and services. The same management process is also applicable to the software development. The data generated is huge in volume and can be structured or unstructured and satisfies all the characteristics of big data. Organizations need big data analytics for better decision making and sustain competitive advantage in the market.

There is need of analyzing huge amount of data for better decision making. There can be predictive analysis, classification, clustering, or some other analytical techniques applied to this data. Big Data management requires a scalable architecture for performing the analytical tasks on it. There are some solutions available for management, storage and information retrieval from this huge amount of data like:

- Hadoop Distributed File System,
- MapReduce,
- Apache Spark, and
- HBase.

The objective of this book is to address different challenges in operation management and how to overcome these challenges using Big Data analytics. The application of the tools, techniques discussed in this book is related to the following areas:

- Software development process,
- Social networks analysis,
- Semantic analysis,
- Predictive analytics,
- Education System Analysis,
- Transport analysis, and
- Cloud computing.

The impact of this edited book is focused on big data analytics and its application in operation management. It is evident that data rate is very high in today’s world and growing need of analysis in decision making is important area to focus upon. The content of this book will benefit the researchers in the area of information technology and operations management. The potential readers of this book are

1. The researchers, academician and engineers will be able to know the growing applications of big data analytics in operation management.
2. The industry persons dealing in data storage and its management will get to know how to analyze the big data.
3. The students of various universities studying information technology, computer science and management.

The intended audiences include:

1. Academician,
2. Researchers,
3. Engineers,
4. Industry professionals, and

Chapter 1 introduces big data and its 5V. It focuses on real time data collection, real time event processing, comprehensive data collection and deterministic data collection. The chapter discusses various approaches with use-cases to big data operations management. Authors present different approaches for big data operations management: SPLUNK approach, Reflex system approach, Cloud physics approach and XANGATI approach. These approaches perform the collection of heterogeneous data from various resources and perform Map reduce task on this data. SPLUNK has indexing, logs, drill down as its features. Reflex approach can be used for real time analysis by using virtualization query language.

Chapter 2 covers the application of analytical techniques on transport organization data. It discusses Artificial Neural Networks and a model to predict Gauge Widening. The chapter explains preprocessing of data and techniques for obtaining insight of various factors resulting track degradation and track maintenance. The problem of efficient maintenance of track is proposed to solve by a model which is required to predict the degradation of the tracks. The following is presented:

- Analyzing and trending the tram track attributes/variables over a period of time.
- The correlation between those variables and track deterioration will be identified.
- A model will be developed to predict the track deterioration/degradation based on tram track variables.
- The maintenance and replacement activities will be identified and a priority list for reparation of track rails will be prepared. This will lead to minimizing the maintenance costs and preventing unnecessary maintenance actions and therefore saving time.
Chapter 3 introduces ZAMBiDM which envisaged to collectively bringing all the resources from the ZAMREN member institutions so that big data can be managed. It discusses various Big Data technologies employed in industry. It proposes a ZAMBiDM model and explains the functions of the system components. It also includes ZAMBiDM’s operations and highlights the benefits of envisaging such a model. The ZAMBiDM is developed in-line with the following objectives:

- Design and build the ZAMBiDM architecture.
- Road map of ZAMBiDM.
- Acquire servers to accommodate large volumes of data.
- Collect and store large heterogeneous volume of data from academic and industry sectors including the NRENs worldwide.
- Manage the “V’s” of data and these are volume, velocity, variety, value and veracity.
- Build operational processes.
- Develop relevant strategies for the respective operation nodes.
- Elevate ZAMBiDM data to executive level.

There are several challenges while applying analytical and mining technique. Implementation of a mining algorithm in parallel requires certain modification in the basic sequential approach. Authors discuss the challenges like:

- Division of a problem into the sub problems.
- Deployment of the problem on a grid or cluster.
- Load balancing.

Chapter 4 discusses predictive analytical techniques and their implementations. Authors present regression technique, Machine learning technique, random forest etc. Chapter provides a detailed explanation of linear regression model, Parallel Back propagation, parallel support Vector Machine.

Chapter 5 discusses the pros and cons of applying opinion mining in operations management from the big data perspective. Semantic analysis also known as opinion mining is used on text generated by systems for finding the opinion of the users. In this analysis of opinionated text that contains people’s opinions toward entities such as products, organizations, individuals, and events. Opinion mining has been applied to a number of domains like hotels and restaurants, different products, movies and politics. Not only had this, but the ever growing growth of information on social media platforms influenced many companies to use this analysis in the operational management as well. Authors also include advantages and disadvantages of opinion mining in operational management.
Chapter 6 discusses the operational management in educational system and its benefits, it also explains some existing operational techniques to store and process big data. Authors propose a conceptualized framework of novel educational operational management. It is highly essential that operational management pertaining to educational system be studied very closely. The functions of operations may again include various essential activities like prediction, effective planning of capacity, data storage etc. Chapter explains the adaptation of information and communication technologies into the education system. Chapter also discusses complexities which exist in the education system in context of operation management. Chapter derives the relationship between operation management in education system and the big data analytics. It also provide information about the solving these problems using the big data analytical techniques.

Chapter 7 discusses a synthetic semantic data management approach for managing the data of small and medium size enterprises which can analyze customer and business intelligence. This chapter presents the survey of big data and semantic technology. It also discusses the semantic web languages, Synthetic Semantic Data Management (SSDM) and its implementation. This uncovers hidden facts, unknown relations, customer requirements and business needs. Small and medium size enterprise generates data which can be used for analysis so that better decision can be made. The chapter provides an explanation of using different scripting languages like OWL for managing small and medium size enterprise data.

Chapter 8 presents an overview of big data security referencing Hadoop framework. Authors discuss the possible solutions related to privacy and security of big data environment. Since security of data is main concern of every organization because of data sensitivity and its usefulness in decision making process. The hostile nature of digital data itself poses security challenges. This chapter tries to uncover big data security issues and to find the better solution for handling security challenges. The observation and analysis of different security mechanism are presented.

Manish Kumar
Indian Institute of Information Technology, Allahabad, India