Preface

In recent years, advances in Web technology and the proliferation of sensors and mobile devices connected to the Internet have resulted in the generation of immense data sets available on the Web that need to be processed and stored. Cloud computing has emerged as a paradigm that promises to meet these requirements. Cloud computing is an extremely successful paradigm of service oriented computing, and has revolutionized the way computing infrastructure is abstracted and used. Three most popular cloud paradigms include: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

Cloud computing is associated with service provisioning and these services are based on processing and analysis of huge volume of data. Big data management in cloud computing environments is hereby proposed in recent years. It should be noted that big data management in cloud computing environments need information interoperations in a right way. At this point, explicit knowledge management can play a critical role in cloud computing. Viewed from this point, a new cloud paradigm named Knowledge as a Service (KaaS) is emerging. It is very critical to deal with the worth information for effective problem solving and decision making. It is especially true when a variety of data types and users’ requirements as well as large volumes of data are available. The techniques of data and knowledge management in cloud computing environments are challenging today’s database and information systems, and meanwhile promote their evolution.

The research, development and application of data and knowledge management in cloud computing environments are receiving increasing attention. By means of cloud computing technology, massive data and knowledge can be managed effectively and efficiently to support various problem solving and decision making. Data and knowledge management in cloud computing environments are the fields which must be investigated by academic researchers together with developers. Nowadays the research and development of big data management in cloud computing environments are receiving increasing attention. Some key techniques of data management in cloud computing environments are investigated by researchers all around the world on the one hand. On the other hand, massive data in many application domains are managed and analyzed in cloud computing environments.

This book covers a fast-growing topic in great depth and focuses on technologies and applications of big data management in cloud computing environments. It aims to provide a single record of current advances in big data management in cloud computing environments. The objective of the book is to provide state of the art information to academics, researchers and industry practitioners who are involved or interested in the study, use, design and development of advanced and emerging information technologies with ultimate aim to empower individuals and organizations in building competencies for exploiting the opportunities of the data and knowledge society. The book presents the latest results in
technology research, design implementation and application development of big data management in cloud computing environments. The chapters of the book have been contributed by different authors and provide possible solutions for the different types of technological problems.

This book, which consists of eleven chapters, is organized into three major sections. The first section contains the first five chapters. The next four chapters comprise the second section. The third section consists of the final two chapters.

The first section discusses several foundational issues in big data management in cloud computing environments. Let us look at these issues.

The development and extensive use of highly distributed and scalable systems to process Big Data have been widely considered. As a result, new data management architectures (e.g. distributed file systems and NoSQL databases) are used in this context. However, features of Big Data like their complexity and data analytics demands indicate that these concepts solve Big Data problems only partially. A development of so called NewSQL databases is highly relevant and even special category of Big Data Management Systems is considered. Jaroslav Pokorny and Bela Stantic discuss these trends and evaluate some current approaches to Big Data processing. They further identify the current challenges and suggest possible research directions in Big Data processing.

Ahmet Artu Yildirim and Dan Watson concentrate on cloud-aware distributed object storages in Internet services. They first present overall information of the current state-of-the-art storage systems which are used for reliable, high performance and scalable storage needs in data centers and cloud. On the basis, they introduce a novel distributed object storage system (CADOS) for retrieving big data efficiently through HTML5-enabled web browsers over big data in cloud infrastructure. The objective of the system is to minimize latency and propose a scalable object storage system on the cloud using a thin RESTful web service and modern HTML5 capabilities.

The Cloud of Things (CoT) is used to connect heterogeneous physical things to the virtual domain of the cloud. Bashar Alohalı demonstrate the creation of CoT system faces many challenges and risks linked with specific technical and network characteristics of things. Bashar Alohalı particularly discusses the security threats and available solutions that are connected with resource-constrained nature of CoT devices, and further addresses new security requirements that tackle the issue of CoT network. Finally it is advocated to investigate the issue of CoT architecture deployment and implementation focusing on specific features of IoT (Internet of Things) integration with cloud computing services.

Omer K. Jasim et al. introduce an innovative cloud computing cryptographic environment (CCCE), which entails both Quantum Cryptography-as-service (QCaaS) and Quantum Advanced Encryption Standard (QAES). CCCE poses more secure data transmission channels by provisioning secret key among cloud’s instances and consumers. In addition, the QCaaS solves the key generation and key distribution problems that emerged through the online negotiation between the communication parties. Note that the CCCE solves the distance limitation coverage problem that is stemmed from the quantum state property. Also the CCCE reveals a higher secure data transmission comparing to others cloud security systems.

Big Data is driving radical changes in traditional data analysis platforms. Cloud Computing appears as an ideal solution for the efficient management and analysis of big data streams. However, there are many limitations and barriers in such effectiveness, especially in the light of the remote and local data movement overhead (network messages, memory and storage accesses). After critically analyzing the several cloud-based big data solutions available on the market, positioning them in the different phases of a well-established data value chain, Madhavi Arun Vaidya tries to highlight the potential, the urgent
Preface

challenges to be faced and the early results related to the development of data-intensive applications distributed across multiple cloud-based data centres.

The second section deals with the issues of managing big data of special types in cloud computing environments.

In order to store massive spatiotemporal datasets, Berkay Aydin, Vijay Akkineni and Rafal A. Angryk investigate the important aspects of non-relational (NoSQL) databases for storing large-scale spatiotemporal trajectory data. They propose two main data storage schemata for storing trajectories, which are called traditional and partitioned data models. In addition, to efficiently retrieve data under different usage scenarios, they design spatiotemporal and non-spatiotemporal indexing structures. The results of the experiments exhibit the advantages of utilizing the proposed data models and indexing structures for various query types.

Wei Yan concentrates on parallel queries of k Nearest Neighbor (kNN) for massive spatial data in cloud computing environments. They propose a parallel method of kNN queries based on clusters in MapReduce programming model. They first propose a partitioning method of spatial data using Voronoi diagram and then cluster the data point after partition using k-means method. They furthermore propose an efficient algorithm for processing kNN queries based on k-means clusters using MapReduce programming model. Extensive experiments evaluate the efficiency of the proposed approaches.

In the domain of disaster management, large volume of heterogeneous data is gathered and integrated by means of collaboration on a global scale and across governments, industries, and communities. Katarina Grolinger et al. propose a Knowledge as a Service (KaaS) framework for disaster cloud data management (Disaster-CDM), with the objectives of facilitating information gathering and sharing; storing large amounts of disaster-related data; and facilitating searching and supporting interoperability and integration. In the Disaster-CDM approach, NoSQL data stores provide storage reliability and scalability while service-oriented architecture achieves flexibility and extensibility. They demonstrate the contribution of Disaster-CDM by integration capabilities, on examples of full-text search and querying services.

The Resource Description Framework (RDF) is a model for representing information resources on the Web. With the widespread acceptance of RDF as the de-facto standard recommended by W3C (World Wide Web Consortium) for the representation and exchange of information on the Web, a huge amount of RDF data is being proliferated and becoming available. To manage massive RDF data, NoSQL (“not only SQL”) databases have been used for scalable RDF data store. Zongmin Ma and Li Yan provide an up-to-date overview of the current state of the art in RDF data storage in NoSQL databases. They aim at suggestions for future research.

The third section presents two application scenarios of big data management in cloud computing environments.

Aiming to examine the attitudes of Chinese Organizations towards Cloud Computing adoption, Tomayess Issa et al. provides an answer to the question: “What are the advantages and disadvantages of Cloud Computing adoption in Chinese organizations?” The answer is sought by means of an online survey of (N = 121) respondents. Their survey results reveal the Chinese position regarding the cloud movement, its strengths and weaknesses, the threats posed by Cloud Computing in China, and the specific advantages and disadvantages of this new technology that Chinese organizations and research communities should embrace for the realization of future cloud systems.

K. Palanivel proposes a cloud-oriented green computing architecture for eLearning applications (COGALA). The COGALA consists of client, client-oriented green cloud middleware and green broker. The green cloud middleware provides the client a tool to better manage the distribution of tasks to cloud
with the least carbon emission and other relevant decision criteria. The middleware is composed of a user interface application and a windows service. This architecture provides incentives to both users and providers to use and deliver the most “Green” services, respectively. K. Palanivel also discusses the implication of the proposed solution for future research directions to enable cloud-oriented green computing.