

## Preface

Three IT initiatives are currently popular in every organization across the globe: cloud computing, big data analytics and Internet of Things (IoT). Cloud computing and big data are complementary, forming a dialectical relationship. Cloud computing and the Internet of Things' widespread application is people's ultimate vision, and the rapid increase in big data is a thorny problem that is encountered during development. Cloud computing is a trend in technology development, while big data is an inevitable phenomenon of the rapid development of a modern information society. In the other hand, big data analytics offers the promise of providing valuable insights that can create competitive advantage, spark new innovations, and drive increased revenues. As a delivery model for IT services, cloud computing has the potential to enhance business agility and productivity while enabling greater efficiencies and reducing costs. But, Internet of Things refers to scenarios where network connectivity and computing capability extends to objects, sensors and everyday items not normally considered computers, allowing these devices to generate, exchange and consume data with minimal human intervention. There is, however, no single, universal definition. The concept of combining computers, sensors, and networks to monitor and control devices has existed for decades. The recent confluence of several technology market trends, however, is bringing the Internet of Things closer to widespread reality. These include Ubiquitous Connectivity, Widespread Adoption of IP-based Networking, Computing Economics, Miniaturization, Advances in Data Analytics, and the Rise of Cloud Computing. IoT implementations use different technical communications models, each with its' own characteristics. Four common communications models described by the Internet Architecture Board include: *Device-to-Device*, *Device-to-Cloud*, *Device-to-Gateway*, and *Back-End Data-Sharing*. These models highlight the flexibility in the ways that IoT devices can connect and provide value to the user. However, this edited book deals with cloud database systems with capacity planning and big data/big data analytics. And cloud computing is core to cloud database systems.

Cloud computing and big data analytics, both technologies continue to evolve. Organizations are moving beyond questions of what and how to store big data to address how to derive meaningful analytics that respond to real business needs. As cloud computing continues to mature, a growing number of enterprises are building efficient and agile cloud environments, and cloud providers continue to expand service offerings. Speculation says that all IT organizations look for cloud computing as the structure to support their big data projects. Big data environments require clusters of servers to support the tools that process the large volumes, high velocity, and varied formats of big data. The outbreak of big data is a thorny problem encountered in social and informatization development. Because of the growth of data traffic and data volume, data formats are now multisource and heterogeneous, and they require real-time and accurate data processing. Big data can help us discover the potential value of large amounts of data. Traditional IT architecture is incapable of handling the big data problem, as there are many

## Preface

bottlenecks, such as: poor scalability; poor fault tolerance; low performance; difficulty in installation, deployment, and maintenance; and so on. Because of the rapid development of the Internet of Things, the Internet, and mobile communication network technology in recent years, the frequency and speed of data transmission has greatly accelerated. This gives rise to the big data problem, and the derivative development and deep recycling use of data make the big data problem even more prominent. Clouds are already deployed on pools of server, storage, and networking resources and can scale up or down as needed. Cloud computing offers a cost-effective way to support big data technologies and the advanced analytics applications that can drive business value.

As a new computing model, Cloud computing has gained great momentum in both academia and industry. Governments, research institutions, and industry leaders are actively trying to solve the growing computing and storage problems in the Internet age using Cloud computing. In addition to Amazon Web Services (AWS), Google's App Engine, and Microsoft's Windows Azure Services—along with other commercial cloud platforms—there are also many open-source Cloud computing platforms, such as: OpenNebula, Eucalyptus, Nimbus, and OpenStack. Each platform has its own significant features and constantly evolving community. AWS is the most popular Cloud computing platform. The most distinct features of its system architecture are open data, functioning via Web Service interfaces, and the achievement of loose-coupling via Service Oriented Architecture.

At the same time, big data processing needs the support of cloud data centers that have large-scale physical resources and Cloud computing platforms that have efficient resource scheduling and management. Cloud computing management platforms can: provide flexible and efficient deployment, operation, and management environments for large data centers and enterprises; support heterogeneous underlying hardware and operating systems with virtualization technology; provide applications with cloud resource management solutions that are secure, high performance, highly extensible, highly reliable, and highly scalable; reduce the costs of application development, deployment, operation, and maintenance; and improve the efficiency of resource utilization. This book focuses cloud database system taking big data into consideration.

The area cloud computing and big data analytics play a major role in daily life and it's importance is increasing day by day. Without research, technology does not carry any meaning. Similarly, without information and engineering the word "grow" has no existence in every field of life. Technology makes life better and smoother. To achieve that objective we have to value the potential global contribution of our researchers. Every day new inventions are coming to limelight enriching human life. The above topic tells about the latest development of cloud database systems which is essential to inform and educate the demand of global scientists and human being as well. Hence, our endeavor is to capture new inventions and present those to the researchers at large.

It gives me immense pleasure to introduce this collection of chapters, *Advancing Cloud Database Systems and Capacity Planning With Dynamic Applications*, to the readers of the book series. The objective of this book is to bridge the existing gap in literature and comprehensively cover the system, processing and application aspects of cloud computing and big data analytics. Due to rapid developments in specialized areas of cloud computing and big data, this book takes on the form of a contributed volume where well known experts address specific research and application problems. It presents the state of the art as well as the most recent trends both in theory and applications. It serves the needs of different readers at different levels. It can be used as stand-alone reference for masters, researchers and practitioners. For example, the researcher can use it as an up-to-date reference material since it offers a

broad survey of the relevant literature. Finally, practicing engineers may find it useful in designing and implementing various cloud database system tasks.

This book purports to serve as a research reference book in the area of cloud computing and big data analytics by providing useful cutting edge research information to the students, researchers, scientists, engineers and other working professionals in this area. The book provides the latest research trends and concepts to develop new methodologies and applications in the area of resource allocation, resource management, dynamic virtual machine management, cloud platforms and infrastructure, energy and carbon-aware management of geo-distributed cloud data centers, big data optimization in cloud environment and its applications. In addition, the book also incorporates chapters related to new challenging application area of cloud computing and big data analytics. Above all, each and every chapter is designed in such a way as to incorporate the latest literature review, methods and models, implementation, experimental results, performance analysis, conclusion, future work and the latest relevant references.

Besides huge applications of the subject cloud computing and big data analytics in diversified fields, several challenges are coming into picture in the new millennium based on different applications and geographical locations. Such challenges are of different types like privacy and security, data integration, cloud capacity planning, data virtualization, mining and visualization, predictive analytics in the cloud, big data processing algorithms and machine learning for information extraction, etc.

Theory and applications are both important in cloud computing and cloud database systems with capacity planning. They are treated equally well in this book on a pragmatic basis. Here different types of problems of scientists and engineers are addressed concerning cloud database systems with cloud computing. The book comprises chapters contributed by highly qualified and diverse group of authors.

I am very grateful to the researchers who offered their services in making the book effective and scientific.

It is my pleasure to present this book which includes selected chapters of internationally recognized authors on cloud database systems, cloud computing and related areas. The book is intended to provide a forum for researchers, educators and professionals to share their discoveries and innovative practices with others and to explore future trends and applications in the field of cloud database systems and its future trends. However, this book will also provide a forum for dissemination of knowledge on both theoretical and applied research on the above areas with an ultimate aim to bridge the gap between these coherent disciplines of knowledge. This forum accelerates interaction between the above bodies of knowledge, and fosters a unified development in the next generation cloud database systems and application.

The broad spectrum of this book includes the topics:

- Information Systems Applications,
- Artificial Intelligence,
- Database Management,
- Information Storage and Retrieval,
- Data Mining and Knowledge Discovery,
- Algorithm Analysis and Problem Complexity,
- Big Data Architectures,
- Cloud Capacity Planning,
- Cloud Database Systems (NoSQL, NewSQL, and Hybrid),
- Cloud Platforms and Infrastructures,
- Data Integration as a Service,

## Preface

- Data Mining and Predictive Analytics in the Cloud,
- Data Virtualization and Governance,
- Statistical, Mathematical, and Dynamical Properties of Big Data,
- Machine Learning for Information Extraction,
- Distributed Systems and Cloud Computing for Big Data,
- Big Data Applications,
- Big Data Security and Privacy,
- Big Data Management,
- Big Data Processing Algorithms,
- Big Data Search, Mining and Visualization,
- Big Data Applications for Business, Government, and Society,
- Multimedia Big Data,
- Big Data Communication,
- And Related Topics.

## ORGANIZATION OF THE BOOK

The book, *Advancing Cloud Database Systems and Capacity Planning With Dynamic Applications*, provides an overview of recent research developments in the field of cloud database systems and its applications. This book contains 15 chapters starting from basic concept level to research and application level.

The first chapter discusses the fundamentals of data mining and data warehousing techniques, tools and methodologies to solve different kinds of problems being generated day by day in order to extract valuable information for future development of the mankind.

Atefeh and Buyya in second chapter titled “Energy and Carbon-Aware Management of Geo-Distributed Cloud Data Centers: State of the Art and Future Directions” have addressed a taxonomy and classified the existing research works based on their target system, objective, and the technique they use for resource management in achieving a green cloud computing environment. Finally, this chapter discusses how each work addresses the issue of energy and carbon-efficiency and also provides an insight into future directions. These challenges have motivated many researchers to work in this area.

The third chapter talks about Cloud service models as designed to power storage platforms, infrastructure solutions, provisioning and virtualization. Cloud computing services are developed to support shared network resources, provisioned between physical and virtual networks. These solutions are offered to organizations and consumers as utilities, to support dynamic, static, network and database provisioning processes. Vendors offer these resources to support day-to-day resource provisioning amid physical and virtual machines.

In contrast, Chapter 4 gives overview about the consumption of huge amount of power by large scale data centers and generation of considerable amount of carbon dioxide. The authors have tried to solve this problem by using utility functions, which are widely used in economic modeling for representing user preferences. This approach also uses Meta heuristic genetic algorithm and the fitness is evaluated with the utility function to consolidate virtual machine migration within cloud environment. Their results show the significant improvement in energy consumption over the existing methodologies.

Chapter 5, “Resource Allocation Policies in Cloud Computing Environment,” attempts to discuss the task scheduling algorithms for better resource utilization without violating service level agreement

(SLA). The authors have developed different backfilling algorithms for task scheduling scalable to cloud computing environment and implemented. Based on the result it works well.

In Chapter 6, the authors have demonstrated an overview on public auditing schemes to explore the system models designed by various researchers to provide security when data outsourced. This book chapter aims to discuss some security issues concerned to cloud data storage in cloud computing environment. The cloud service provider has to verify the data correctness by trusted third party auditor (TTPA) while outsourcing takes place. Although it leads to computation cost and communication costs in huge amount, it is essential due to open nature of cloud. Moreover, the authors have tried their best to discuss different standards.

In Chapter 7, Sangulagi et al. have discussed about the processing and storing of sensed data in cloud. In general due to the lack of battery energy and bandwidth the sensor nodes are incapable to store and process large amount of data which is a challenging task. The sensor cloud comes into existence to accomplish multiple tasks that are not possible with existing sensor network. In sensor cloud the sensed data are processed and stored in the clouds and data can be accessed anywhere and anytime. Maintaining the resources and providing the resources to end users is a challenging task in sensor cloud. This chapter talks about the architecture of sensor cloud, application of sensor cloud in various sectors, advantages of using sensor cloud compared to existing networks and management of resources in sensor cloud.

Chapter 8 presents challenges and issues in information retrieval system. When the average speed of computers has almost doubled in a span of around eighteen months, the average speed of the network has doubled merely in a span of just eight months. In order to improve the performance, the author has used different techniques for web based information retrieval system. The author has also tried his best to show the pros and cons of the techniques used with sufficient justification.

Chapter 9 intends to evaluate whether it is possible to optimize and extend an existing sales system called *Patent Customer Treaty* (PCT), which is currently suffering from unacceptably high running times in its simulation process. This is done through analysis of the current implementation, followed by optimization of its models and development of efficient algorithms. The performances of these optimized and extended models are compared to the existing one in order to evaluate their improvement. The conclusion of this chapter is that the simulation process in PCT can indeed be optimized and extended.

Whereas, Chapter 10 presents Cloud Database Systems- NoSQL, NewSQL and Hybrid. In this chapter, the authors have discussed the limitations of relational database over cloud database. A cloud database is a database that typically runs on a cloud computing platform. Cloud databases are on the rise as more and more businesses look to capitalize on the advantages of cloud computing to power their business applications. Cloud databases are mainly used in data mining, data warehousing and business intelligence. This chapter deals with different types of cloud database and how database influence capacity planning.

In Chapter 11, the author has presented the challenges and opportunities of developing hybrid mobile devices. In recent decades, vendors developed technology infrastructure solutions to integrate with enterprises and consumers' mobile devices. Hybrid development platforms are solution architecture designed to enhance developers' capabilities to provide organizations and customers the level of services to support mobile devices capabilities. These solutions are easy to deploy at various enterprises. These capabilities can be distributed to/or integrated with mobile devices as agile applications and system interfaces. Hybrid mobile devices are designed to also provide users with enhanced technology solutions: cloud computing, big data, the Internet, physical and/or virtual network systems. In general, the development of hybrid mobile platforms provides developers with advanced technology capabilities, necessary for

## **Preface**

supporting mobile devices once deployed to the marketplace. Technical and security features affecting the development and security of mobile devices are also discussed in this chapter.

However, Chapter 12 presents a literature survey on denial of service attacks over cloud environment. Cloud services provide pay as go models on capacity or usage. For providing better cloud services, capacity planning is very important. Proper capacity planning maximizes efficiency and on the other side proper control over the resources helps to overcome from attacks. As the technology develops in one side, threats and vulnerabilities to security also increases on the other side. A complete analysis of Denial of Service (DOS) attacks in cloud computing and how are they done in the cloud environment and the impact of reduced capacity in cloud causes greater significance. Among all the cloud computing attacks, DOS is a major threat to the cloud environment. In this chapter, the authors have discussed DOS attack in the cloud and its types, tools used to perform DOS attack and how they are detected and prevented. Finally it deals with the measures to protect the cloud services from DOS attack and also penetration testing for DOS attack.

Chapter 13 tells us about security aspects in cloud. The goal of this chapter is to present an intensive study of various security concerns in cloud computing. Cloud computing is recently emerging technology which provides a way to access computing resources over Internet on demand and pay per use basis. Cloud computing enables access to shared pool of resources efficiently that are managed by third party cloud service providers. Despite of various advantages of cloud computing security is the biggest threat. This chapter comprises various security issues at different levels in environment which includes infrastructure level security, data level and storage security. It also deals with the concept of Identity and Access Control mechanism.

In Chapter 14, the author has discussed the recent trend in data mining include web mining where it discovers knowledge from web based information to improve the page layout, structure and its content thereby it reduces the user latency in accessing the web page and website performance.

But the objective of Chapter 15 is to know about session hijacking over cloud environment. Generally, a session is said to be the collective information of an ongoing transaction. This package is typically stored on the server as a temporary file and labeled with an ID, usually consisting of a random number, time and date the session was initiated. That session ID is sent to the client with the first response, and then presented back to the server with each subsequent request. This permits the server to access the stored data appropriate to that session. That, in turn allows each transaction to be logically related to the previous one. Session hijacking is the common problem that is experienced in the cloud environment in which the session id is gained and information is gathered using the session ID compromising its security. This chapter discusses session hijacking and the countermeasures to prevent session hijacking.

This edited book has specific salient features. They are:

- It deals with important and timely topic of emerging relationship between cloud computing technology and big data analytics in unattended area.
- It presents research findings and materials authored by global experts in the field.
- It serves as a comprehensive source of information and reference material on the topic cloud database systems.
- It presents latest development of the topic related to cloud computing and big data analytics.
- It presents the research findings in well organized and structured manner.
- Even though it is not a text book, it can serve as a complete reference material for both cloud computing and big data analytics.

- It can certainly be used as one for graduate courses and research oriented courses dealing with cloud database management system.
- It can serve as light house of knowledge in cloud computing research lab including data science lab.

This comprehensive and timely publication aims to be an essential reference source, building on the available literature in the field of cloud computing and big data analytics to boost further research in this dynamic and challenging field. It is expected that this text will provide the resources necessary for technology developers, scientists and manufacturer to adopt and implement new inventions across the globe.

In short, I am very happy with both experience and end product of our sincere efforts. It is certain that this book will continue as an essential and indispensable resource for all concerned for coming years.

*Narendra Kumar Kamila*

*C. V. Raman College of Engineering, India*