

Preface

Cloud computing is a novel computing paradigm which has changed the way enterprise or Internet computing is performed. Today, for almost all the sectors in the world, cloud computing is synonym to on-demand provisioning and delivery of IT services in a pay-as-you-go model. The success story of cloud computing as a technology is credited to the long-term efforts of computing research community across the globe. Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) are the three major cloud product sectors. Each one of these product sectors has their effects and reaches to various industries. If forecasts are to be believed, then more than two-third of all the enterprises across the globe will be entirely run in cloud by 2026. These enthusiastic figures have led to huge funding for research and development in cloud computing and related technologies. University researchers, research labs in industry, and scholars across the globe have recreated the whole computing world into a new cloud enabled world. This has been only possible by coordinated efforts into this direction. Today, almost every university across the globe has cloud computing and its related technologies included in their computer science curriculum. Additionally, there are extensive efforts on innovation and technology creation in the direction of cloud computing. These efforts are much visible in the reputed cloud computing research platforms like international conferences and journals.

We feel that there is a significant need to systematically present quality research findings of recent advances in cloud computing for the benefit of community of researchers, educators, practitioners, and industries. Although there are large numbers of journals and conferences available, there is a lack of comprehensive and in-depth tutored analysis on various new developments in the field of cloud computing. This book on “Research Advances in Cloud Computing” discusses various new trends, designs, implementations, outcomes, and directions in the various areas of cloud computing. This book has been organized into three sections:

1. Programming model, infrastructure, and runtime
2. Resource Management
3. Security.

The first chapter on “[Serverless Computing: Current Trends and Open Problems](#)” covers various serverless platforms, APIs, their key characteristics, technical challenges, and related open problems. Recently, enterprise application architectures are shifting to containers and micro-services, and it provides enough reasons for serverless computing. The chapter provides detailed requirements of different programming models, platforms, and the need of significant research and development efforts to make it matured enough for widespread adoption.

Cloud providers face the important challenge regarding resource management and aim to provide services with high availability relying on finite computational resources and limited physical infrastructure. Their key challenge is to manage resources in an optimal way and to estimate how physical and logical failures can impact on users’ perception. The second chapter on “[Highly Available Clouds: System Modeling, Evaluations and Open Challenges](#)”, presents literature survey on high availability of cloud and mentions the main approaches for it. It explores computational modeling theories to represent a cloud infrastructure focusing on how to estimate and model cloud availability.

The third chapter on “[Big Data Analytics in Cloud—A Streaming Approach](#)” discusses streaming approach for data analytics in cloud. Big data and cloud have become twin words—used sometimes interchangeably. Interpretation of big data brings in idea of mining and analytics. There is significant literature on cloud that discusses infrastructure and architecture but a very little literature for algorithms required for mining and analytics. This chapter focuses on online algorithms that can be used for distributed, unstructured data for learning and analytics over Cloud. It also discusses their time complexity, presents architecture for deploying them over cloud, and concludes with presenting relevant open research directions.

Cloud data centers must be capable to offer scalable software services, which require an infrastructure with a significant amount of resources. Such resources are managed by specific software to ensure service-level agreements based on one or more performance metrics. Within such infrastructure, approaches to meet non-functional requirements can be split into various artifacts, distributed across different operational layers, which operate together with the aim of reaching a specific target. Existing studies classify such approaches using different terms, which usually are used with conflicting meanings by different people. Therefore, it is necessary a common nomenclature defining different artifacts, so they can be organized in a more scientific way. The fourth chapter on “[A Terminology to Classify Artifacts for Cloud Infrastructure](#)” proposes a comprehensive bottom-up classification to identify and classify approaches for system artifacts at the infrastructure level, and organize existing literature using the proposed classification.

The fifth chapter focuses on “[Virtual Networking with Azure for Hybrid Cloud Computing in Aneka](#)”. It provides a discussion on the need of inter-cloud communication in the emerging hybrid, public, or federated clouds. Later, they provide

a case of Azure Virtual Private Network (VPN) services to establish such inter-cloud connections using an overlay network for hybrid clouds in Aneka platform. It also presents a functional evaluation of the proposed approach with the help of experiments.

The sixth chapter on “[Building Efficient HPC Cloud with SR-IOV Enabled InfiniBand: The MVAPICH2 Approach](#)” presents a detailed case of high-performance computing in cloud. It discusses the single-root I/O virtualization performance in the InfiniBand interconnects and provides locality aware communication designs to optimize the overall performance using MVAPICH2 library. It also proposed advanced designs to support the HPC in cloud computing environments along with open research problems.

To facilitate effective resource allocation, cloud providers should allocate resources ahead of service demands, in a way that does not waste resources. The calculation of optimal allocations requires integer programming, which is computationally difficult to accomplish. The seventh chapter on “[Resource Procurement, Allocation, Metering, and Pricing in Cloud Computing](#)” proposes an approach using the uncertainty principle of game theory which achieves close to optimal results. An approach for time-varying tariffs for cloud services, considering varying load levels on the cloud provider’s infrastructure, and the time-varying pricing of electricity from a smart grid, is also proposed. The chapter involves the creation of a per-instance power consumption model for VMs on a cloud and a power-aware cloud metering architecture.

Auto-scaling is an important feature of cloud computing which allows flexible just-in-time allocation and release of computational resources in response to dynamic and often unpredictable workloads. The eighth chapter on “[Dynamic Selection of Virtual Machines for Application Servers in Cloud Environments](#)” covers the importance of auto-scaling for web applications whose workload is time dependent and prone to flash crowds. Reactive auto-scaling policies are successful, but here the authors are investigating the issue related to which VM type is the most suitable for the specific application and have proposed an approach for dynamic VM-type selection. It uses a combination of online machine learning techniques, works in real time, and adapts to changes in the users’ workload patterns, application changes as well as middleware upgrades and reconfigurations. The chapter has described a prototype, which is tested with the CloudStone benchmark deployed on AWS EC2 and it has achieved encouraging results.

One of the current concerns of systems designers is related to the growth of power consumption in cloud computing systems. The techniques to address this problem range from decisions on locations for data centers to techniques that enable efficient resource management. Resource allocation, as a process of resource management, distributes the workload throughout the data center in an efficient manner, minimizing the power consumption and maximizing the system performance. The ninth chapter on “[Improving the Energy Efficiency in Cloud Computing Data Centres Through Resource Allocation Techniques](#)” presents an overview of the resource management and resource allocation techniques, which contribute to the reduction of power consumption without compromising the cloud

user and provider constraints. It also covers two practical cases to illustrate the theoretical concepts of resource allocation as well as have discussed the open challenges that resource management will face in the coming years.

The tenth chapter on “[Recent Developments in Resource Management in Cloud Computing and Large Computing Clusters](#)” provides a comprehensive and detailed overview of overall cloud computing resource allocation framework with a focus on various resource scheduling algorithms. This chapter also provides a definitive direction toward cloud scheduling solutions, architectures, and fairness algorithms.

The eleventh chapter on “[Resource Allocation for Cloud Infrastructures: Taxonomies and Research Challenges](#)” provides a classification of VM placements solutions in the form of taxonomies. These taxonomies are prepared for conceptualization of VM placement problem as provider–broker setting, and framing it as an optimization problem. Authors also comment on the formation of cloud markets to provide a basis for multi-objective VM placement algorithms.

The twelfth chapter on “[Many-Objective Optimization for Virtual Machine Placement in Cloud Computing](#)” presents a comprehensive discussion on virtual machine placement problem and extends the discussion by proposing many objective VM placement algorithms for initial VM placement and reconfiguration. It also gives an overview of open research problems at the end of the chapter to provide the scope of future work toward fully dynamic multi-objective VM placement problems.

The thirteenth chapter on “[Performance Modeling and Optimization of Live Migration of Virtual Machines in Cloud Infrastructure](#)” is based on improvement of the pre-copy algorithm for live migration system. The improved pre-copy algorithm is developed by three models: (i) compression model, (ii) prediction model, and (iii) performance model. Each model is used to evaluate downtime and total migration time of different workloads. The first model performs migration of different sizes of VM with three workloads: (i) idle system, (ii) kernel compile, and (iii) static web server. Prediction model works with adaptive dirty rate and adaptive data rate to evaluate complex workloads running in a VM. The performance model is used to find dirty pages using dirty page rate model. It is observed that both prediction model and performance model work efficiently than the existing framework of Xen. It concludes that three proposed models are able to improve pre-copy and the results are tested for the same.

Security and privacy being a very active and hot topic of research and discussion these days, we have five chapters dedicated to the relevant issues associated with cloud computing security. Isolated containers are rapidly becoming a great alternative to traditional virtualized environments. The fourteenth chapter on “[Analysis of Security in Modern Container Platforms](#)” makes two important contributions. First, it provides a detailed analysis of current security arrangements in the container platforms. Second, it offers an experimental analysis of containers by providing details on common threat and Vulnerabilities Exposures (CVEs) exploits. This twofold analysis helps in comparing the CVE exploits to be able to compare with the state-of-the-art security requirements by the popular literature.

The fifteenth chapter on “[Identifying Evidence for Cloud Forensic Analysis](#)” discusses forensic analysis and post-attack evidence collection on the cloud

computing infrastructures. Authors describe the evidence collection activity at three different places which are at Intrusion Detection System (IDS), cloud provider API calls, and VM system calls. It shows a step-by-step attack scenario reconstruction using the proposed prolog-based tool following the proposed evidence collection approach. Forensic analysis of cloud computing infrastructures is still in its infancy and authors provide directions for data collection and forensically capable clouds.

The sixteenth chapter on “[An Access Control Framework for Secure and Interoperable Cloud Computing Applied to the Healthcare Domain](#)” addresses various health record security issues and provides an FSICC framework (Framework for Secure and Interoperable Cloud Computing) that provides a mechanism for multiple sources to register cloud, programming, and web services and security requirements for use by applications. Future research directions are provided at the end of this chapter to help the enthusiastic readers about the open areas.

The seventeenth chapter on “[Security and Privacy Issues in Outsourced Personal Health Record](#)” provides a detailed survey on existing personal health record management systems (PHRMSs) considering the security and privacy features provided by each one of them. This state-of-the-art survey is extended by giving pointers to multiple open research problems in the healthcare domain.

The last in the series of five chapters dedicated to cloud security is a chapter on “[Applications of Trusted Computing in Cloud Context](#)”. Trusted computing paradigm has been considered as one of the important security research milestones to leverage various security solutions. This chapter investigates applications of trusted computing in cloud computing areas where security threats exist, namely in live virtual machine migration.

Ahmedabad, India
Ajmer, India
Melbourne, Australia

Sanjay Chaudhary
Gaurav Somani
Rajkumar Buyya

Research Advances in Cloud Computing

Chaudhary, S.; Somani, G.; Buyya, R. (Eds.)

2017, XX, 465 p. 126 illus., 81 illus. in color., Hardcover

ISBN: 978-981-10-5025-1