Preface

INTRODUCTION

Utility computing is basically a service provisioning model, enabling using computing a Utility. Computing systems starting from Mobile to Supercomputers are becoming heterogeneous using both multi-core CPUs and multi-thread Graphical Processing Units (GPUs) for higher power efficiency and computational throughput. In this book the authors have discussed various facets of Utility computing, GPU architecture and related technologies in details in the 20 chapters. The chapters are contributed by authors from various reputed IIT, NIT from India and reputed Universities from India and abroad.

The underlying design of utility computing is based on a service provisioning model, where Consumers pay providers for using computing power only when they need to. Our Industries like Banking, Healthcare and Education are growing with the rapid demand of resources. Thus, to reduce the load of resource consumption and improve its capacity with performance, is a major focus. This could be crafted with a policy based assignment of resources approach and adaptive self-learning with virtualization of resources for optimization. Using such approaches and methods, it helps in the quality of service with a higher availability, greater performance, and improved recoverability.

The mission “Saving Earth” has become need for all of us to sustain life on the earth. There are many holistic approaches for Green Computing which impact on stack holders of the computing system including hardware, software and people. There are many reasons to develop green computing like environmentally friendly, saving powers, long term profit, reduce pollution, power management and increasing performance, etc. Approach to develop green computing can be broadly divided into four parts such as Power efficient Hardware Device Manufacturing, Software Techniques, Peoples’ Awareness and Standard Policies.

Another important feature of the book is the inclusion of a chapter on Cloud Computing, the most popular form of Utility computing. Cloud Computing is an on Demand, Pay-per-use distributed computing service delivery model in which computing resources can be used as a Utility like other utilities such as water, electricity etc. as per the requirement. Cloud computing has made it possible to provide virtually unlimited computing infrastructure, i.e. IaaS, on demand using virtualization technology. The chapter on virtualization provides a quantitative and qualitative comparison of four popular virtualization platforms, VMware vSphere (ESXi), Microsoft Hyper-V and open-source Xen, KVM. Another very important topic on cloud computing “Load Balancing” which helps in achieving maximum resource utilization and user satisfaction. This mechanism transparently transfers a load from a heavily loaded process to an under loaded process. Various techniques for load balancing in cloud computing are discussed in the greater interest of the readers.
ORGANIZATION OF THE BOOK

All computing systems, from mobile to supercomputers, are becoming heterogeneous parallel computers using both multi-core CPUs and many-thread Graphical Processing Unit (GPUs) for higher power efficiency and computation throughput. While the computing community is racing to build tools and libraries to ease the use of these heterogeneous parallel computing systems, effective and confident use of these systems will always require knowledge about the low-level programming interfaces in these systems. In this book the authors have discussed various facets of Utility computing, GPU architecture and related technology in details. This book is having 20 chapters contributed by authors from various reputed Institutes from India and abroad.

The first chapter of the book is about Utility Computing. Utility computing is envisioned to be the next generation of Information Technology (IT) evolution that depicts how computing needs of users can be fulfilled in the future IT industry. Its analogy is derived from the real world where service providers maintain and supply utility services, such as electrical power, gas, and water to consumers. Consumers’ pay for the services based on their usage. Therefore, the underlying design of utility computing is based on a service provisioning model, where Consumers pay providers for using computing power only when they need to. This chapter first discusses some features, challenges and impacts of utility computing. Finally, this chapter points out the important, standards and recommendation of utility computing in the cloud platform with a suitable example.

Chapter 2 of the book is about the Optimization & Management of Resource in Utility Computing. In this modern era of computing, we are surrounded directly or indirectly related to the computer resources and services, and uses several programming languages, different database management systems like RD-BMS. At the same time, it needs respective compilers and editors for different languages and the most important resource is “storage”, which could be either in the form of primary or secondary space storage. Our Industries like banking, health and education are growing with the rapid demand of resources. Thus, to reduce the load of resources consumption and improves its capacity with performance, would be a major focus in this chapter. This could be crafted with the policy - base assignment of resources approach and adaptive self-learning with virtualization of resources for optimization. Using such approaches and methods, it helps in quality of service with higher availability, greater performance, and improved recoverability.

In traditional computing, hardware and software resources are configured to meet the single user’s requirements. Utility computing model assists to maximize utilization of resource, hence, minimize the cost associated with it. This approach allows a utility to gradually modernize computing service by enabling existing legacy applications and adapting standard interfaces. The Enterprise Architecture comprises the resource associated with facilitating services domain for Hardware and Application Monitoring via shared services. Chapter 3 of the book discusses about the several architectures that have been presented by different researchers and experts in the field of utility computing.

Chapter 4 is about the Green computing. The mission “Saving Earth” has become need for all of us to sustain life on the earth. There are many holistic approaches for Green Computing which impact on stock holders of the computing system including hardware, software and people. There are many reasons to develop green computing like environmentally friendly, saving powers, long term profit, reduce pollution, power management and increasing performance, etc. Approach to develop green computing can be broadly divided into four parts: hardware device manufacturing, software techniques, peoples’ awareness and standard policies.
Chapter 5 of the book is about the energy efficient computing. In recent years, companies in the computer industry realize that going green is in their best interest, both in terms of public relations and reduced costs. The principle behind energy efficient coding is to save power by getting software to make less use of the hardware, rather than continuing to run the same code on hardware that uses less power. This chapter first discusses features, challenges and impacts of green computing. Finally, this chapter points out the standard and recommendation of green computing with suitable example.

Chapter 6 of the book deals with the history of development of Graphics Processing Unit (GPU). The chronology of development of GPU is discussed in this chapter. It is forecasted that by 2020, 48 Core GPU will be available while by 2030 GPU with 3000 Core is likely to be available. This chapter describes the chronology of evolution of GPU hardware architecture and the future ahead.

Chapter 7 primarily discusses about GPU parallelism, applications, probable challenges and also highlights some of the GPU computing platforms, which includes CUDA, OpenCL (Open Computing Language), OpenMPC (Open MP extended for CUDA), MPI (Message Passing Interface), OpenACC (Open Accelerator), DirectCompute, and C++ AMP (C++ Accelerated Massive Parallelism). Each of these platforms is discussed briefly along with their advantages and disadvantages.

Chapter 8 is about CUDA (Compute Unified Device Architecture), a parallel computing platform and programming model created by NVIDIA and implemented by the graphics processing units (GPUs). CUDA gives program developers direct access to the virtual instruction set and memory of the parallel computational elements in CUDA GPUs. This chapter first discusses some features and challenges of GPU programming and the effort to address some of the challenges with building and running GPU programming in high performance computing (HPC) environment. Finally, this chapter point out the importance and standards of CUDA architecture.

Chapter 9 discusses about the design and implementation of an On-chip interconnection mechanism (INOC) for multicores processors. The benchmark results are presented by using a full system simulator.

Chapter 10 explores on the open source cloud storage service alternatives and their feature parity, running one of these services to set up a personal cloud server to protect data privacy and security. The authors also take a look at the security issues, data availability, cost effectiveness and viability of deploying a personal cloud when compared to some popular public cloud storage providers.

Chapter 11 provides a quantitative and qualitative comparison of four popular virtualization platforms i.e. VMware vSphere (ESXi), Microsoft Hyper-V and open-source Xen, KVM.

Chapter 12 presents aspect of today’s cyber world and its importance in concern to security with utility computing. This chapter also details about vulnerability of utility networks to cyber-attacks via external connections.

Chapter 13 deals with performance analysis of CUDA implementation of an image quality assessment tool based on the structural similarity index (SSI). Since it had been initially created at the University of Texas in 2002, the Structural SIMilarity (SSIM) image assessment algorithm has become a valuable tool for still image and video processing analysis. SSIM provided a big giant over MSE (Mean Square Error) and the PSNR (Peak Signal to Noise Ratio) techniques because it is way more closely aligned with the results that would have been obtained by subjective testing. For objective image analysis, this new technique represents as significant advancement over SSIM as the advancement that SSIM provided over PSNR.

Chapter 14 presents a novel approach on symmetric key cryptography technique using genetic algorithm that is implemented on CUDA architecture.
Chapter 15 describes the GPU implementation of a Friend recommender system which is based on content-based filtering mechanism. It has given significant speedups from its previous counterparts, thus making the whole process more efficient.

Chapter 16 discusses the reasons underlying the proposals and shows the pitfalls associated with software performance and its reliability.

Chapter 17 discusses about the various image features for indexing and retrieval of images so that they can retrieve image in consideration of texture, feature and color.

Chapter 18 proposes a hybrid technique for solving tasks assignment problem in cloud platform. PSO based heuristic has been developed to schedule random tasks in heterogeneous data centers. They used variants of Particle Swarm Optimization (PSO) which gives better result than PSO and other heuristics for load balancing in a cloud computing environment.

Chapter 19 describes about adaptability of mobile computing architecture with utility computing. How service governance could be beneficial in adaptive mobile computing services.

Chapter 20 outlines and examines the different components and computational requirements of a face recognition scheme implementing the Viola-Jones Face Detection Framework and an Eigen picture face recognition model. Face recognition can be separated into three distinct parts: face detection, eigenvector projections, and database search. For each, the detailed explanation of the exact process along with an analysis of the computational requirements and scalability of the operation is deliberated upon.

_Ganesh Chandra Deka_

*Regional Vocational Training Institute (RVTI) for Women, India*