Guest Editorial – FGCS Special Issue on Cluster Computing

Cluster computing can be described as a fusion of the fields of parallel, high-performance, distributed, and high-availability computing. Cluster computing has become a hot topic of research among academic and industry community including system designers, network developers, language designers, standardizing forums, algorithm developers, graduate students and faculties. The use of clusters as computing platform is not just limited to scientific and engineering applications; there are many business applications that can benefit from the use of clusters. There are many exciting areas of development in cluster computing with new ideas as well as hybrids of old ones being deployed for production as well as research systems. The aim of this special issue is to bring together original and latest work from both academia and industry on various issues related to cluster computing.

This research, and development, is being done in many areas but there are a few that are of special interest. The first one is, with no doubt, the network. Clusters are based on the communication between nodes and designing fast and low latency networks is a must for Clusters to become the configuration of the future. Plenty of work is being done in this are in both hardware (Infiniband, SCI, Myrinet, QSnet) and software (Low latency protocols such as VIA).

Allowing heterogeneity in both hardware and software (OS) is extremely important and is becoming one of the key issues in Cluster research. This heterogeneity allows a better expandability (i.e. use old and new components) and scalability and has to be addressed at many levels. We need tools that allow us to develop applications for these configurations, as well as scheduling systems that offer mechanisms for exploitation of heterogeneity and that deliver better application performance. The papers selected for inclusion in this special issue address these issues.

File systems are also of great interest as clusters are being used for I/O intensive applications such as e-commerce, WEB servers or databases. This kind of file systems tight the I/O devices as well as offer high performance and a single-system image. Achieving these objectives is not easy and plenty of work is being devoted.

Clusters also provide an excellent platform for solving a range of parallel and distributed applications in both scientific and commercial areas. For scientific application, clusters can be used in grand challenge or supercomputing applications, such as earthquakes or hurricanes predication, complex crystallographic and microtomographic structural problems, protein dynamics and biocatalysis, relativistic quantum chemistry of actinides, virtual materials design and processing including crash simulations, and global climate modeling. For the commercial applications, cluster can be best used in e-commerce as superserver, which consolidate web server, ftp server, email server, database server, etc. Clusters can also be used in data mining applications, to provide the storage and data management services for the data set being mined and computational services required by the data filtering, preparation and mining tasks. Other commercial applications includes image rendering, network simulation, etc.

Due to the growing interest in cluster computing, IEEE Task Force on Cluster Computing (TFCC) was formed in early 1999. TFCC is acting as a focal point and guide to the current cluster computing community and has been actively promoting the field of cluster computing with the aid of a number of novel projects. With the support of TFCC, a series of special sessions, workshop, symposiums and conferences were held to guide R&D work both in academic and industrial settings. Further information on TFCC activities can be accessed from the web site: http://www.ieeetfcc.org/.

This special issue emerges from the best seven papers that we selected from the three conferences in the area of cluster computing. They are the first IEEE International Workshop/Conference on Cluster Computing[1] (Melbourne, Australia), CC-TEA 2001 (Las Vegas, USA), and the Asia-Pacific International Symposium on Cluster Computing (Beijing, China). The last two meetings have been merged to create a major IEEE international conference series (called CCGRID[2]) that focus on both cluster computing and grid[3] technologies. Grid technology will help in coupling multiple clusters in the same or different organizations to create computational grids (federated or hyper clusters) for solving large-scale multidisciplinary problems.

---

[1] http://www.clustercomp.org
[2] http://www.ccgrid.org
[3] http://www.gridcomputing.com

The papers that we have selected for inclusion in this special issue have received the higher review ratings for their research contributions. All the papers are extended and revised to reflect the latest research achievements in various areas of cluster computing. The topics of the papers cover different aspects of cluster computing ranging from lightweight (low-latency and high-bandwidth) communication protocols to highly optimized sorting algorithms implemented on clusters. Topic areas covered include, rapid application development environments, operating system kernels, file systems, load-balancing mechanisms, communication subsystems, industry standard oriented user-level communication interfaces, and a number of parallel sorting schemes. The works presented have addressed mechanisms for handling heterogeneity in clusters, as this is an important issue for handling design issues such as expandability of clusters.

The first paper entitled "*MPI-Delphi: An MPI Implementation for Visual Programming Environments and Heterogeneous Computing*" presents authors' efforts to integrate MPI for parallel programming with Delphi visual programming environment. MPI-Delphi interface makes it possible to manage a cluster of heterogeneous PCs. This interface is also suitable for some specific kind of programs, such as monitoring long execution parallel programs, or computational intensive graphical simulations.

"*PODOS - The Design and Implementation of a Performance Oriented Linux Cluster*" reports the design issues of a performance-oriented operating system, PODOS, which harness the performance capabilities of a cluster-computing environment. PODOS added four new components to existing Linux operating system, they are Communication Manager, PODOS File System, Resource Manager, and Global Inter-Process Communication. The custom designed communication protocol uses round-robin Transmission-Groups mechanism to multiplex packets across multiple network interfaces. PODOS file system builds an efficient file sharing environments on top of high-speed communication subsystem.

"*On a Scheme for Parallel Sorting on Heterogeneous Clusters*" discusses the parallel sorting algorithms and their implementations suitable for cluster architectures in order to optimize cluster resources. By carefully studying the various sorting algorithms on heterogeneous cluster, the authors concluded that the software challenges to better utilize the heterogeneous cluster platform are data decomposition techniques, scheduling and load balancing methods. The algorithm described in this paper combines very good property for load balancing.

"*Cluster File Systems: A Case Study*" presents a scalable single-image file system, called COSMOS, designed for Dawning 2000 superserver. COSMOS provides location transparency and strong UNIX file-sharing semantics. COSMOS uses serverless design and introduces a dual-granularity cooperative caching. Other characteristics of COSMOS includes heuristic replacement algorithm, network disk striping, and distributed metadata management. COSMOS provides high I/O bandwidth for large files and good I/O performance for small files and directory reads.

"*Load Balancing for Heterogeneous Clusters of PCs*" discusses the authors' experience of using the asymmetric load balancing approach in heterogeneous cluster environments. By using the LU benchmark from the NPB family, authors find that asymmetric load balancing approach is a general-purpose tool that can be used with any data decomposed regular problem, and with some extensions, it can be also used in irregular problem.

"*Directed Point: A Communication Subsystems for Commodity Supercomputing with Gigabit Ethernet*" studies the practical issues on the design a new high performance communication subsystem, called Directed Point (DP). The DP abstraction model depicts the communication channels built among a group of communicating processes. It supports both point-to-point communication and various types of group operations. The API of DP combines features from BSD sockets and MPI to facilitate the peer-to-peer communication in a cluster. DP improves the communication performance by reducing protocol complexity through the use of directed message, by reducing the intermediate memory copies between protocol layers through the use of token buffer pool, and by reducing the context switching and scheduling overhead through the use of light-weight messaging calls.

"*VI Architecture Communication Features and Performance on the Giganet Cluster LAN*" presents the performance result study of Giganet cLAN VI architecture hardware implementation. VI architecture aims to close the performance gap between the bandwidths and latencies provided by the communication hardware and visible to the application by minimizing the software overhead on the critical path of communication. The focus of this study is to assess and compare the performance for different VIA data-transfer modes and specific features that are available to higher-level communication software like MPI. The features investigated in this paper include the use of send/receive vs. RDMA data transfers, polling vs. blocking to check the completion of communication operations, multiple VIs,

completion queues, and scatter capabilities of VIA.

From the above discussion, it is clear that the papers included in this special issue cover as broad a range of topics in cluster computing. We hope that the whole cluster computing community including researchers, developers, practitioners, and users find this special issue of use and interest.

We would like take this opportunity thank editors of the Future Generation Computer Systems, Peter Sloot and Doutzen Abma for inviting use to edit this special issue. In fact, we are honored by their invitation. We congratulate all the authors whose papers have been included in this special issue and we thank them for agreeing to extend their papers to reflect the latest state of the art in their topical area. In cluster resources section below, we have included pointers to further information, particularly a list of book that have been published recently. We hope that you will find this special issue interesting and useful. Happy reading!

Guest Editors

Rajkumar Buyya
School of Computer Science and Software Engineering
Monash University
Melbourne, Australia
Email: rajkumar@csse.monash.edu.au
http://www.buyya.com

Hai Jin
Internet and Cluster Computing Lab.
University of Southern California
Los Angeles, CA, 90089, USA
Email: hjin@ceng.usc.edu
http://ceng.usc.edu/~hjin/

Toni Cortes
Departament d'Arquitectura de Computadors
Universitat Polit`ecnica de Catalunya
Barcelona, Spain
Email: toni@ac.upc.es
http://www.ac.upc.es/homes/toni

Cluster Computing Resources

[1] R. Buyya (editor). *High Performance Cluster Computing: Systems and Architectures*. Vol. Prentice Hall, 1999.

[2] R. Buyya (ed.). *High Performance Cluster Computing: Programming and Applications*. Vol. 2, Prentice Hall, 1999.

[3] G. Pfister. *In Search of Clusters*. 2/e, Prentice Hall PTR, NJ, 1998.

[4] K. Hwang, and Z. Xu, *Scalable Parallel Computing - Technology, Architecture, Programming*. McGraw-Hill, 1998.

[5] T. Sterling, J. Salmon, D. Becker, and D. Savarrese. *How to Build a Beowulf*. The MIT Press, 1999.

[6] B. Wilkinson and M. Allen. *Parallel Programming Techniques and Applications Using Networked Workstations and Parallel Computers*. Prentice Hall, NJ, 1999.

[7] Cluster Computing Info Centre: http://www.buyya.com/cluster

[8] IEEE Task Force on Cluster Computing – http://www.ieeetfcc.org