A Proposal for Creating a Computing Research Repository (CoRR, http://www.arXiv.org/) on Cluster Computing

Rajkumar Buyya
Monash University, Melbourne, Australia

Introduction to Cluster Computing

Cluster computing is emerged as results of availablity of powerful computers and high-performance networks as commodity components and a fusion/*unification of the fields of parallel, high-performance, distributed, high-availability computing, and internet technologies*. It has become a popular topic of research among the academic and industrial communities, including system designers, network developers, algorithm developers, as well as faculty and graduate researchers. The use of clusters as a application platform is not just limited to the scientific and engineering area; there are many business applications, including Internet and E-commerce are benefiting from the use of clusters.

There are many exciting areas of development in cluster computing. These include new ideas as well as hybrids of old ones that are being deployed in production and research systems. There are attempts to couple multiple clusters, either located within one organisation or situated across multiple organisations forming what is known as a *federated clusters* or *hyperclusters*. The exploitation of federated clusters (clusters of clusters) as an infrastructure can seem to be approaching the area of the increasingly popular grid infrastructure.

The motivations and arguments for establishing a Computing Research Repository (CoRR) on Cluster Computing topic will be the same as those taken into consideration while forming the IEEE Task Force on Cluster Computing. Hence, we present them below.

IEEE Task Force on Cluster Computing (TFCC)[1]

Recognizing the trend toward clusters for high-performance computing, the IEEE Computer Society has approved a Task Force on Cluster Computing (TFCC). You may ask, "What's so special about that? People have been using clusters of computers for years. It's a bit like Microsoft realizing that the Internet may be a big thing." But think again. The overwhelming number of cluster-related projects and products appearing in the development arena and the commercial marketplace means that a focused group can help lead international efforts in cluster-based computing.

Why not place cluster-based activities under the umbrella of another related Technical Committee, such as the committees on the Internet, Supercomputing Applications, or Distributed or Parallel Processing? After all, what is cluster computing but a mixture of these disciplines? But cluster computing combines so many computing concepts and technologies that placing it under an existing banner would dilute the attentions of individuals interested in all the aspects that come together in this field.

With the advent of the TFCC, interested Computer Society members can participate in one focused group to champion the cause of cluster computing by sponsoring workshops, conferences, projects, and standards.

The TFCC has been in existence since early 1999 and within its short life it has started to have an impact on the cluster computing community in both academia and industry. The number of volunteers willing to be involved in the TFCCs activities and the amount of international cluster-related events being put on can be seen as some evidence of this success.

[1] IEEE Task Force Home Page: `http://www.ieeetfcc.org`

The TFCC is committed to the development of cluster computing research, education, and industry. Towards this end, we have conducted a number of activities including promoting open community discussion, holding events that bring together leading experts, the instigation of a popular book donation programme and the provision of advise to commercial bodies. One particularly successful activity has been the TFCCs educational programme. Here we are attempting to promote the take up of cluster-related technologies in the core curriculum of educational institutions around the world. Another associated effort is the TFCCs book donation programme. In conjunction with influential international authors and publishers the TFCC have donated more than five hundred books to academic institutions around the world. In order to help underdeveloped countries, TFCC has reserved 50% of books that we donate for faculty members working in developing/underdeveloped as it is very hard for them to afford to buy expensive books published international. The impact of this program is clear when we noticed that faculties in number of educational institutions have started teaching courses on cluster computing.

Why Cluster Computing Repository?

As we pointed earlier, the motivations and arguments for establishing a Computing Research Repository (CoRR) on Cluster Computing topic will be the same as those taken into consideration while forming the IEEE Task Force on Cluster Computing. Another additional reason for the repository is that "It helps in disseminating information freely and making it easily accessible to whole community including our colleagues working in developing/underdeveloped countries. Our interest in creating such repository for "cluster computing" basically stem from their request. For your information, we have enclosed one/two of the email messages that we received from them.

As we mentioned earlier cluster computing is an interdisciplinary field where we look into all aspect of this technology. Cluster-based systems are not only used in high performance computing, they are extensively used in high-availability, mission critical application areas. Cluster systems differ from distributed systems, as nodes in the distributed systems do not have strong sense of membership. Whereas, cluster nodes have strong sense of membership and all of them work as a single resource (exhibits single system image). Due to its interdisciplinary nature, we can better serve cluster-computing community if there a single focused repository. This in fact compliments IEEE Task Force's efforts as its members strong the feel the need of a repository. Instead of creating a separate TFCC's repository, we feel that it will be good for the whole community of computer science, if we can create it as part of CoRR and task force contributes for the same.

On advised from Joe Halpern (CoRR administrator), we have also discussed about the proposal for creating a repository and we received many positive comments and number of IEEE task force volunteers and members are willing to help us out in this regard.

Cluster Computing Conferences

A number of conferences and workshops are devoted to cluster computing each year. IEEE Task Force has already established two major international conferences in this area:

- IEEE International Conference on Cluster Computing (CLUSTER)
- IEEE International Symposium on Cluster Computing and the Grid (CCGrid)

These two conference series have attracted hundreds of participants every year and the IEEE Computer Society Press has published the proceedings. In addition, task force has already sponsored a number of other workshops and symposiums in this area around the world. For detailed information on these can be found at the task force web site.

Cluster Computing Journal

Task force volunteers have brought out special issues on Cluster Computing along with many international journals. Our volunteers have also established a new dedicated Journal for this area called:

- Cluster Computing: the Journal of Networks, Software Tools and Applications
  Baltzer Science Publishers, The Netherlands.
  http://www.baltzer.nl/cluster/cluster.asp

- A number of other journals have brought out special issues on Cluster Computing.

Proposed Moderators

- Rajkumar Buyya, Monash University, Melbourne, Australia
- Mark Baker, University of Portsmouth, UK
- Xiannong Meng, University of Texas, USA
- Ivan Tanev Tanev, Muroran Institute of Technology, Japan
- Jesús Carretero, Carlos III University in Madrid, Spain

Papers for inclusion in Repository

The task force (TFCC) has already offering such a repository serve to industries as part of educational program. Many industries such as Sun and IBM have already posted their whitepapers on TFCC Education resources page. We will move all these papers to this proposed repository.

The task force has over five hundred members on discussion forum and over two hundred volunteers all over the world. Once repository is formed, within a few days we will be able to create a repository of hundreds of papers. Task force itself has published white papers and number of reports. All these publications can become part of the CoRR for the benefit of whole community and the world.

A D D I T I O N A L I N F O R M A T I O N

The comments of one of our colleagues (Xiannong Meng): "Clsuter computing is related, but different from parallel computing, distributed computing, or reliable computing. Traditional parallel computing concentrated on tightly coupled algorithms and architectures. Cluster computing focuses on single system image while the underlining support architecture can vary from a collection of loosely coupled workstations to massively parallel processor architecture. We should view cluster computing as a natural extension to parallel/distributed/reliable computing. The term 'parallel/distributed/reliable computing' can no longer define the nature, the meaning, and the technology used in cluster computing. Thus a separate subject area is needed.

FYI, a copy two Messages of Task Force on Cluster Computing Members

```
-------- Original Message --------
Subject: about cluster computing and TFCC
Date: Wed, 16 Aug 2000 09:21:07 +0800
From: Charles Zhuang <charles@mail.hf.ah.cn>
To: "rajkumar@dgs.monash.edu.au" <Rajkumar.Buyya@infotech.monash.edu.au>

rajkumar£¬

      Hi, I am a Ph.D candidate of USTC(University of Science and Technology of
China).
 I am intrested in cluster computing area. I hear about your book "High Performance
Cluster Computing". However I can not afford to it and I have no chance to see it in
China too. Because lagged in E-commerce, I can not make use of Internet to buy it .
Can
you provide me some chapters about SSI or Process Scheduling, Load Sharing, and
Balancing?
        Additionally, It is best that TFCC can establish a free E-print site like
xxx.lanl.gov. After all, we can find many projects in TFCC but few papers.


                        Please mail to:
                                zxn@ahpost.com.cn
                                                Ding qing
```

-------- Original Message --------
Subject: Re: FREE CLUSTER COMPUTING research/papers Repository (CoRR),do we need it ?
Date: Thu, 24 Aug 2000 14:15:39 -0700
From: Ivan Tanev <tanev@surgeonc.csse.muroran-it.ac.jp>
To: rajkumar@csse.monash.edu.au
References: <39A394F2.1F522B65@csse.monash.edu.au>

Dear Dr.Buyya,

Thank yo very much for posting the info about the possibility of establishing the CoRR sub-Section on "CLUSTER COMPUTING". In my opinion it would be quite useful for many researchers due to the following reasons:

- Indeed to some extend cluster computing (CC) can be viewed as a special case of parallel computing, or distributed computing, or internet-based computing, or grid-computing, etc. However CC features specific characteristics and faces specific problems which could be figured out with some difficulties if the matter is dispersed among these areas.

- Moreover, I believe that current technological trends (fast and cheap commodity off the shelf PCs, >1GHz Pentiums, >Gbit/s LANs, NT2000, Linux, etc.) in computer science would lead to establishing a separate, and relatively autonomous (with respect to the parallel computing, or distributed computing, or internet-based computing, or grid-computing) area which should be addressed adequately by researchers.

- Social point of view: I see some discrepancy from economical point of view - we are talking about using commodity off the shelf components for building clusters, which many research institutes can afford, however, doing research could be in the same (or even greater) order of expenses, in terms of cash. Indeed, studying the matter it is next to impossible for many researchers from many countries - to buy books, attend conferences, subscribe to journals, etc., in order to become familiar with the matter (I am from Bulgaria, and I thing this is the case for many of my compatriots right now). I thing, these researchers would be extremely grateful to you for these efforts!

As for myself, I would like, but I do not know if I can contribute to this great idea, since I am doing some work which is not yet recognized as a true cluster computing - using cluster of NT2000 on switched Fast Ethernet for parallel implementations of evolutionary computations (genetic programming, genetic algorithms), using (and this is the controversial point) component object-based approaches (distributed component object model - DCOM). DCOM offers location-, platform-, language-, network protocol-neutrality, it is free (embedded in any of the current Windows'), offers generic support for naming, locating, and protecting the distributed entities. It is fine for now... However, the cost of these benefits is that, compared with MPI, the "raw" DCOM features significant software overheads, and less data transmission ratio.

Thank you.

Sincerely yours,

Ivan Tanev Tanev, Ph.D.
Dept.of Comp.Sci. and System Engineering,
Muroran Institute of Technology,
Mizumoto-cho 27-1,
Muroran, Hokkaido 050,
JAPAN

tel.: (+81 143) 465955
fax.:(+81 143) 465499
e-mail:
tanev@surggw.csse.muroran-it.ac.jp
i.tanev@computer.org