

A Model for Cooperative Federation of Distributed Clusters

Rajiv Ranjan, Rajkumar Buyya and Aaron Harwood
Department of Computer Science and Software Engineering
University of Melbourne
Victoria, Australia
{rranjan,raj,aharwood}@cs.mu.oz.au

Abstract

Interest in Grid computing has grown significantly over the past five years. Management of distributed cluster resources is a key issue in Grid computing. Central to management of resources is the effectiveness of resource allocation, as it determines the overall utility of the system. In this paper, we propose a new Grid system that consists of Grid Federation Agents which couple together distributed cluster resources to enable a cooperative environment.

1 Introduction

Current approaches to resource allocation in a Grid environment are non-coordinated as application-level schedulers or brokers make scheduling decisions independent of the others in the system. Clearly, this can exacerbate the load sharing and utilization problems of distributed resources due to suboptimal schedules that are likely to occur.

To overcome these limitations, we propose a mechanism for cooperative and coordinated sharing of distributed clusters based on computational economy [2, 5, 6]. The resulting environment, called Grid Federation, allows the transparent use of resources from the federation when local resources are insufficient to meet its users requirements. At the same time, it provides services to other members of the federation when it has some spare resources. Fig.1 shows an abstract model of our grid federation over a shared federation directory. To enable policy based transparent resource sharing between these clusters, we define and model a new RMS system, which we call Grid Federation Agent (GFA). Individual GFAs access the directory information using the interface shown in Fig.1, i.e. subscribe, quote, unsubscribe, query. More comprehensive details about the Grid Federation can be found in [4].

2 General scheduling technique

We define our Grid-Federation (shown in Fig.1) as an architectural framework that enables logical coupling of

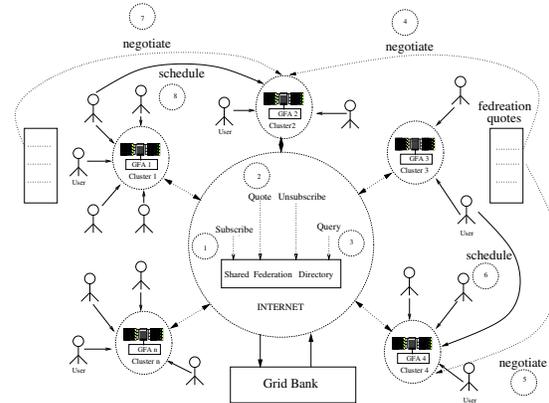


Figure 1. Grid-Federation

cluster resources. In Fig.1 a user who is local to GFA 3 is submitting a job. If the user's job QoS can't be satisfied locally then GFA 3 queries the federation directory to obtain the quote of the 1-st fastest or 1-st cheapest cluster. In this case, the federation directory returns the quote advertised by GFA 2. Following this, GFA 3 sends a negotiate message (enquiry about QoS guarantee in terms of response time) to GFA 2. If GFA 2 has too much load and cannot complete the job within the deadline then GFA 3 queries the federation directory for the 2-nd cheapest/fastest GFA and so on. The query-negotiate process is repeated until GFA 3 finds a GFA that can schedule the job (in this example the job is finally scheduled on cluster 4).

3 Performance Evaluation

We used trace based simulation to evaluate the effectiveness of the proposed system. The workload trace data was obtained from [1]. The trace contained real time workload of various resources/supercomputers that are deployed at the Cornell Theory Center (CTC SP2), Swedish Royal Institute of Technology (KTH SP2), Los Alamos National Lab (LANL CM5), LANL Origin 2000 Cluster (Nirvana) (LANL Origin), NASA Ames

(NASA iPSC) and San-Diego Supercomputer Center (SDSC Par96, SDSC Blue, SDSC SP2). The simulator was implemented using GridSim [3] toolkit that allows modeling and simulation of distributed system entities for evaluation of scheduling algorithms. To enable parallel workload simulation with GridSim, we extended existing GridSim's Alloc Policy and Space Shared entities. More detailed performance evaluation including economy based scheduling, scalability experiments can be found in [4].

3.1 Experiment - Quantifying Resource Utilization

In this experiment we simulate two scenarios

- independent resource;
- resource part of federation ;

We consider, if the user request can not be served within requested deadline, then it is rejected otherwise it is accepted. During scenario-1, all the workload submitted to a resource is processed and executed locally (if possible). Scenario-2 simulates various resources when they are part of the Grid Federation. In this case the workload assigned to a resource can be processed locally. In case a local resource is not available then online scheduling is performed that considers the resources in the federation in decreasing order of their computational speed.

3.2 Results and observations

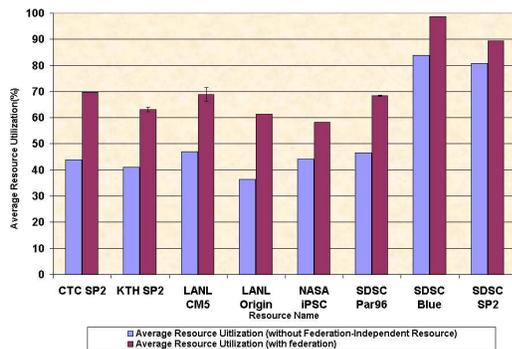


Figure 2. Average resource utilization (%) vs. resource name

During scenario-1 we observed that 6 out of 8 resources remained underutilized (less than 50%). While during scenario 2, we observed that overall resource utilization of most of the resources increased as compared to scenario 1 (when they were not part of the federation), for instance resource utilization of CTC SP2 increased

from 43.88% to 69.74%. Same trends can be observed for other resources as well (refer to Fig.2). Thus, we can see that, for the given job trace, it is preferable to make use of more resources (i.e. migrate jobs). In other words, the job trace shows the potential for resource sharing to increase utilization of the system.

4 Conclusion

We proposed a new distributed cluster resource management system called Grid Federation. The federation uses agents that maintain and access a shared federation directory of resource information. Our simulation results proves the feasibility of the proposed framework. Overall, the proposed Grid Federation, in conjunction with a scalable, shared, federation directory, is a favourable model for building large scale grid systems.

References

- [1] <http://www.cs.huji.ac.il/labs/parallel>.
- [2] D. Abramson, R. Buyya, and J. Giddy. A computational economy for grid computing and its implementation in the Nimrod-G resource broker. *Future Generation Computer Systems (FGCS) Journal, Volume 18, Issue 8, Pages: 1061-1074, Elsevier Science, The Netherlands, October, 2002*.
- [3] R. Buyya and M. Murched. Gridsim: A toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing. *Journal of Concurrency and Computation: Practice and Experience;14(13-15), Pages:1175-1220, 2002*.
- [4] R. Ranjan, A. Harwood, and R. Buyya. Grid federation: An economy based distributed resource management system for large-scale resource coupling. *Technical Report, GRIDS-TR-2004-10, Grid Computing and Distributed Systems Laboratory, University of Melbourne, Australia, 2004*.
- [5] M. Stonebraker, R. Devine, M. Kornacker, W. Litwin, A. Pfeffer, A. Sah, and C. Staelin. An economic paradigm for query processing and data migration in maiposa. *Proceedings of 3rd International Conference on Parallel and Distributed Information Systems, Austin, TX, USA, September 28-30, IEEE CS Press, 1994*.
- [6] C. Waldspurger, T. Hogg, B. Huberman, J. Kephart, and W. Stornetta. Spawn: A distributed computational economy. *IEEE Transactions on Software Engineering , Vol. 18, No.2, IEEE CS Press, USA, February, 1992*.