# Cluster and Grid Integration Issues: A Review

Malkit Singh

**Abstract**— To resolve the problems today's Grid computing offers many solutions that are already addressed. Grid computing includes number of derivatives like data grids, cluster grids, compute grids, science grids, access grids, knowledge grids, terra grids and commodity grids. To support various services Grid need infrastructure like security, uniform access, resource management, scheduling, application composition, computational economy, and accountability. In this paper, I will discuss the several new issues and challenges related to DRMAA, GridBank, Grid programming, Chirp, Scheduling, Distributed Monitoring System, Grid data management, Sensor Grid and ASKALON. The toolsets and techniques are used to overcome the issues that are discussed in this paper.

**Index Terms**— Grid computing, Sensor Grid, Cluster computing, Grid data management, Chirp.

## 1 Introduction

Multiple solution producers and multiple users involved in Grid and Cluster computing systems. When a single resource provider is involved, in such case challenges for preventing vulnerability of confidential information is present [1]. In general, including processor cycles, data sources, special equipment, even people, and the electrical power grid in which the electricity is passed, the computing paradigm is formed [2] [3]. In a computational grid data-intensive jobs are not easy to run. In most systems, the exact set of files to be used by a grid job must define in advanced by the user [4].To undertake complex scientific or commercial problems distribution system is used to form a large scale aggregation of network-connected computer. Grid computing is a crucial arising computing first step [5]. Grid provides an extensible set of services in open grid services architecture (OGSA). For creating and composing sophisticated distributed system OGSA defines web services description language (WSDL) interfaces and associated conventions including lifetime management, change management, and notification [2].

## 2 Review Of Literature

### 2.1 DRMAA

The Distributed Resource Management Application API (DRMAA) specification is a Software standard developed in the Open GRID Forum (OGF). It defines a unified interface for monitoring, job submission and control in Distributed Systems. Using DRMAA, job submission rates can be doubled due to lower submission overhead. Different language bindings of independence parallel developments can identify general issues and lightweight API in a concurrent and non-reliable grid still demand continuous in-depth analysis [6].

### 2.2 Gridbank

In service oriented cluster and grid computing, a Gridbus project technologies are used. At cluster level, for economy-driven cluster scheduling, Libra technology has been developed for distributing computational tasks among resources that belong to a cluster within single administration domain. At grid level, to support Quality of Service (QOS) based schedule for both compute and data-intensive applications various tools are developed. GridBank is a secure grid wide accounting and payment handling system [7].

### 2.3 Grid Programming Issues

For many grid applications, Performance, Portability, Interoperability, Adaptivity, Resource Discovery, Fault Tolerance and Security will be an issue. Grid applications may want authentication, authorization, integrity checking and privacy. Reliable performance for many applications will be an equally important issue. To achieve reliable performance for a programming construct Quality of Service (QOS) will become increasingly necessary. At last, the issue of programming style. This evolution will come down how programming is done to solve computational problems in available computing platform, from single machines to parallel machine to grid [8].

### 2.4 Chirp

To meet the needs of grid computing, the Chirp distributed file system is designed. It provides strong and flexible security mechanism, tunable consistency semantics, clustering to increase capacity and throughput, and spread without particular perquisites. In grid computing environments traditional distributed file system designed for local and campus area network that do not adapt well so, they designed Chirp distributed file system for both cluster and grid computing. It provides services for grid application like deployment, naming, consistency, security and clustering [4].

### 2.5 Scheduling Issues

For grid system, scheduling is a very important mechanism. In a grid system there are many type of resources can be shared and used, and they can be accessed through an application running in the grid. Security is an important aspect in grid scheduling. Other issues are data-aware scheduling. Most of current grid applications are task oriented and resource-oriented approaches [9].

- *Student, Malkit Singh is currently pursuing master's degree program in Computer Science and Engineering in Lovely Professional University, Phagwara, India, PH-7508493397. E-mail: malkitsingh06@gmail.com*

### 2.6 Distributed Monitoring System

The challenges faced by high performance distributed system are scalable monitoring of system state. Network switches, links, computational nodes and storage devices can be complex in large-scale systems. To handle these challenges Ganglia distributed Monitoring system was built. It provides high performance computing system such as clusters and grids. It is a scalable distributed monitoring system. The key design challenges for distributed monitoring system like Scalability, Robustness, Extensibility, Manageability, Portability and Overhead. To minimize static within clusters Ganglia uses a multicast based listen/announce protocol. Researchers measured Ganglia's scalability as a function of cluster size and the number of clusters being federated [10].

### 2.7 Grid Data Management

To contribute various resources like computation, storage, data and applications, Grid is a formed collection of nodes in network. Application and data source can be fairly independent. Peer to peer (P2P) techniques are useful for Grid Data Management which focus on scaling, Dynamicity, autonomy and decentralized control.

To deal with semantically rich data (e.g. XML documents, relational tables, etc) grid and P2P data management are used. Issues of grid and P2P are:

(i) To scale up to high number of nodes data management techniques are required.

(ii) It is difficult for global schema management and access control due to lack of centralized control [11].

### 2.8 Sensor Grid

To meet the computational requirements of applications a compute grid provides distributed computational resources. On the other hand, to provide access to large amounts of storage resources and distributed data, Data Grid is used. In wireless Sensor Networks, to sharing of sensor resources sensor grids extend the grid computing paradigm. In the design of sensor grid the issues and challenges faced that is Grid APIs for Sensors, Network Connectivity and Protocols, Scalability, Power Management, Scheduling, Security, Availability, and Quality of Service. In this study, researchers used Sensor grid testBed tool to improve the issues and sensor grid architecture design [12].

### 2.9 ASKALON

Portability and interoperability of Software tools are critical issues in the grid. In this study researchers used the ASKALON tool set for Cluster and Grid computing. ASKALON mixes four interoperable tools like SCALEA, ZENTURIO, AKSUM and PerformanceProphet. It is designed as a set of distributed grid service-based architecture. Using advanced user portals each tool can be accessed and manipulated. Each tool will be extended with new functionalities [13].

## 3 CONCLUSION

In this paper, we discuss on the issues of integrating cluster and grid computing. There are so many challenging issues that are faced like scalability, scheduling, security, Quality of service etc. To overcome these issues, tool sets are available like ASKALON toolset is used for cluster and grid computing.

## REFERENCES

[1] Smith, M., Engel, M., Friese, T., & Freisleben, B. (2006, May). Security issues in on-demand grid and cluster computing. In *Cluster Computing and the Grid, 2006. CCGRID 06. Sixth IEEE International Symposium on* (Vol. 2, pp. 14-pp). IEEE.

[2] Foster, Ian, Carl Kesselman, Jeffrey M. Nick, and Steven Tuecke. "Grid services for distributed system integration." *Computer* 35, no. 6 (2002): 37-46.

[3] Foster, I., & Kesselman, C. (Eds.). (2003). *The grid 2: Blueprint for a new computing infrastructure*. Morgan Kaufmann.

[4] Thain, Douglas, Christopher Moretti, and Jeffrey Hemmes. "Chirp: a practical global filesystem for cluster and Grid computing." *Journal of Grid Computing* 7, no. 1 (2009): 51-72.

[5] Phan, T., Huang, L., & Dulan, C. (2002, September). Challenge: integrating mobile wireless devices into the computational grid. In *Proceedings of the 8th annual international conference on Mobile computing and networking* (pp. 271-278). ACM.

[6] Troger, P., Rajic, H., Haas, A., & Domagalski, P. (2007, May). Standardization of an api for distributed resource management systems. In *Cluster Computing and the Grid, 2007. CCGRID 2007. Seventh IEEE International Symposium on* (pp. 619-626). IEEE.

[7] Barmouta, A., & Buyya, R. (2003, April). Gridbank: A grid accounting services architecture (gasa) for distributed systems sharing and integration. In *Parallel and Distributed Processing Symposium, 2003. Proceedings. International* (pp. 8-pp). IEEE.

[8] Lee, C., & Talia, D. (2003). Grid programming models: Current tools, issues and directions. *Grid Computing: Making the Global Infrastructure a Reality*, *21*, 555-578.

[9] Xhafa, F., & Abraham, A. (2010). Computational models and heuristic methods for Grid scheduling problems. *Future Generation Computer Systems*, *26*(4), 608-621.

[10] Massie, M. L., Chun, B. N., & Culler, D. E. (2004). The ganglia distributed monitoring system: design, implementation, and experience. *Parallel Computing*, *30*(7), 817-840.

[11] Pacitti, Esther, Patrick Valduriez, and Marta Mattoso. "Grid data management: Open problems and new issues." *Journal of Grid Computing* 5, no. 3 (2007): 273-281.

[12] Lim, H. B., Teo, Y. M., Mukherjee, P., Wong, W. F., & See, S. (2005, November). Sensor grid: Integration of wireless sensor networks and the grid. In *Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on* (pp. 91-99). IEEE.

[13] Fahringer, Thomas, Alexandru Jugravu, Sabri Pllana, Radu Prodan, Clovis Seragiotto, and Hong-Linh Truong. "ASKALON: a tool set for cluster and Grid computing." *Concurrency and Computation: Practice and Experience* 17, no. 2-4 (2005): 143-169.

**Malkit Singh** received his B.Tech degree in Information Technology from Lovely Professional University, Phagwara, Punjab, INDIA, in 2011, now he is doing M.Tech (CSE) from Lovely Professional University, Phagwara, Punjab, INDIA. His research interests include Grid computing and Software engineering.