Service-Level-Agreements (SLA) based Management of Clusters and Data Centres

Staff: Rajkumar Buyya, Srikumar Venugopal
Students: Chee Shin Yeo, Suraj Pandey
Grid Computing and Distributed Systems (GRIDS) Laboratory
Australian Research Council, Department of Education, Science and Training
Web: http://www.gridbus.org/libra/

Clustering involves connecting two or more computers together to take advantage of combined computational power and resources. Hence, a cluster works as an integrated collection of resources that can provide a single system image to users. Clustering is a popular strategy for resource-intensive applications because it transparently spreads the processing of different jobs throughout the cluster. It is used for applications such as weather forecasting, oil explorations, drug discovery, investment risk analysis, and e-commerce data centres.

Computational economy refers to the inclusion of user-specified QoS (Quality-of-Service) parameters with jobs so that resource management is based on a user-centric approach rather than on a system-centric approach. This essentially means that user QoS expectations such as deadline and budget play an important role in determining the priority of a job by the scheduler, than system policies like ordering jobs according to the basis of submission time. Current systems have no or limited support for resource management mechanisms to enable negotiation of differing QoS levels for different users and resource allocation to meet their competing demands. The main purpose of our project is to:

- Develop a framework for negotiating and establishing contract (SLA) between service providers and users
- Develop a resource management system and optimal scheduling algorithms that support SLA-based allocation of resources to meet user QoS requirements

Key References: [1] Jahanzeb Sherwani, Nosheen Ali, Nausheen Lotia, Zahra Hayat, and Rajkumar Buyya, Libra: A Computational Economy-based Job Scheduling System for Clusters, *Software: Practice and Experience*, Volume 34, Issue 6, Pages 573-590, May 2004.

[2] Chee Shin Yeo and Rajkumar Buyya, Service Level Agreement based Allocation of Cluster Resources: Handling Penalty to Enhance Utility, *Proceedings of the 7th IEEE International Conference on Cluster Computing (Cluster 2005),* September 2005, Boston, MA.
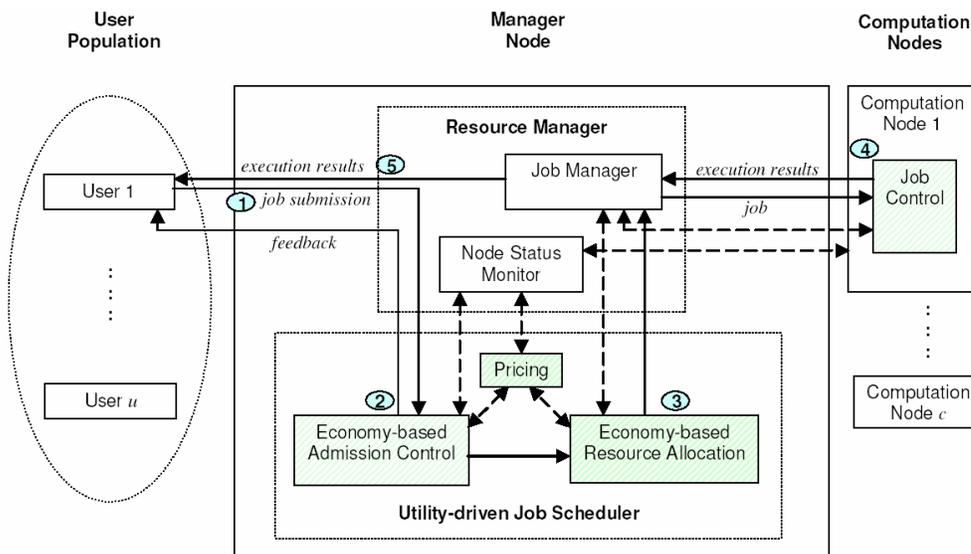

Fig. 1: SLA oriented allocation of Cluster and Data Centre resources.