

CSF4: A WSRF Compliant Meta-Scheduler

Wei Xiaohui¹, Ding Zhaohui¹, Yuan Shutao², Hou Chang¹, LI Huizhen¹
 (1: The College of Computer Science & Technology, Jilin University, China
 2: Platform Computing, Markham, Canada)

Abstract: *In a grid computing environment, heterogeneous resources are distributed in multiple VOs which may have different local policies. Hence, scheduling and resource management are challenging tasks in this context. Since a meta-scheduler is able to provide a virtualized resource access interface to end users, and enforce global policies for both resource providers and consumers as well, it plays more and more important roles in computational grids. In this paper, the design and implementation of Community Scheduler Framework 4.0(CSF4) will be discussed. CSF4 is a WSRF compliant community meta-scheduler, and released as an execution management service of GT4. Using CSF4, the users can work with different local job schedulers, which may belong to different domains. Moreover, the Pre-WS-GRAM protocol is also supported by CSF4, although it is based on WSRF specifications. Therefore, CSF4 can be deployed in a GT4 and GT2 mixed grid environment, which is very convenient for the users who have legacy applications.*

Keywords: grid computing, meta-scheduler, WSRF, GT4, CSF4

1. Introduction

The resources in grids are located in multiple sites and owned by different VOs. Each site may have its local policies which are enforced by local job schedulers. And the protocols used by these local schedulers, such as LSF^[1], PBS^[2], SGE^[3], Condor^[4] etc, are also different. Therefore, scheduling and resource management in grids are much more challenging than in traditional clusters. A meta-scheduler is able to provide a virtualized resource access interface so that the end users can make use of heterogeneous grid resources via standard protocols such as GRAM. Further more, as global policies for both resource providers and consumers can be enforced, the performance of the system and the quality of services provided are improved. Hence, meta-schedulers increasingly play key roles in grid resource management.

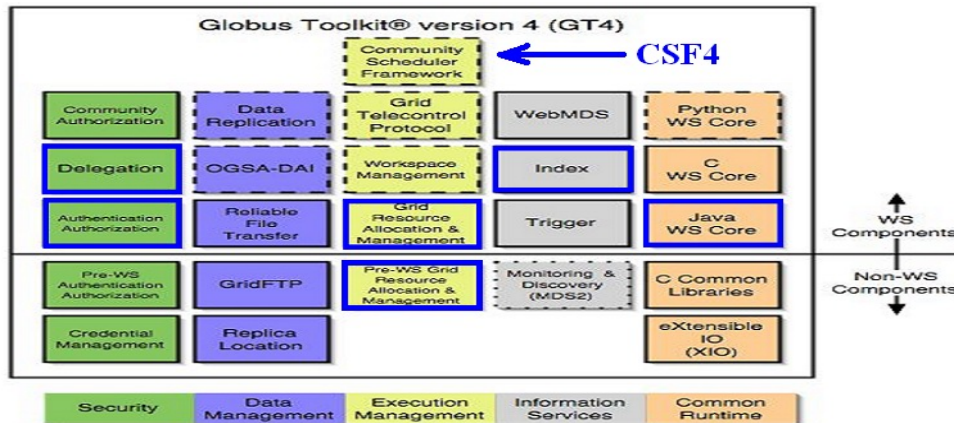


Figure 1 CSF4 in Globus Toolkit

CSF4, Community Scheduler Framework 4.0, is an OGSA based open source

meta-scheduler, which is built on top of GT4 Java WS Core. It is the first WSRF^[5] compliant meta-scheduler, and released as an execution management service component of Globus Toolkit 4^[6], see figure 1. A number of web services are provided, like job service, queue service, reservation service and resource management service and so on. Using CSF4, the grid users can work with different local job schedulers, for instance, LSF, Condor, SGE, and PBS etc via standard GRAM protocols. Since both WS-GRAM and Pre-WS-GRAM protocols are supported, CSF4 can be used in GT4 and GT2 mixed environments. Moreover, the users are also able to reserve resources in advance through CSF4 in a LSF cluster, which is not supported by GRAM yet.

In this paper, we will discuss the design and implementation of CSF4. The rest of the paper is organized as follows: Section 2 describes the architecture of CSF4. In section 3, the services of CSF4 are presented. In section 4, we introduce how to deploy CSF4 in a grid community briefly. The related works are discussed in section 5. At last, we draw the conclusion and recount future plan.

2. CSF4 Architecture

According to ^[7], a grid system can be divided into five layers. From bottom up, they are fabric layer, connectivity layer, resource layer, collective layer and application layer. CSF4 is located in collective layer and takes charge of global resource management and job scheduling. It acts as an intermediary between a user community and local resources by providing a single point of task management and global policy enforcement.

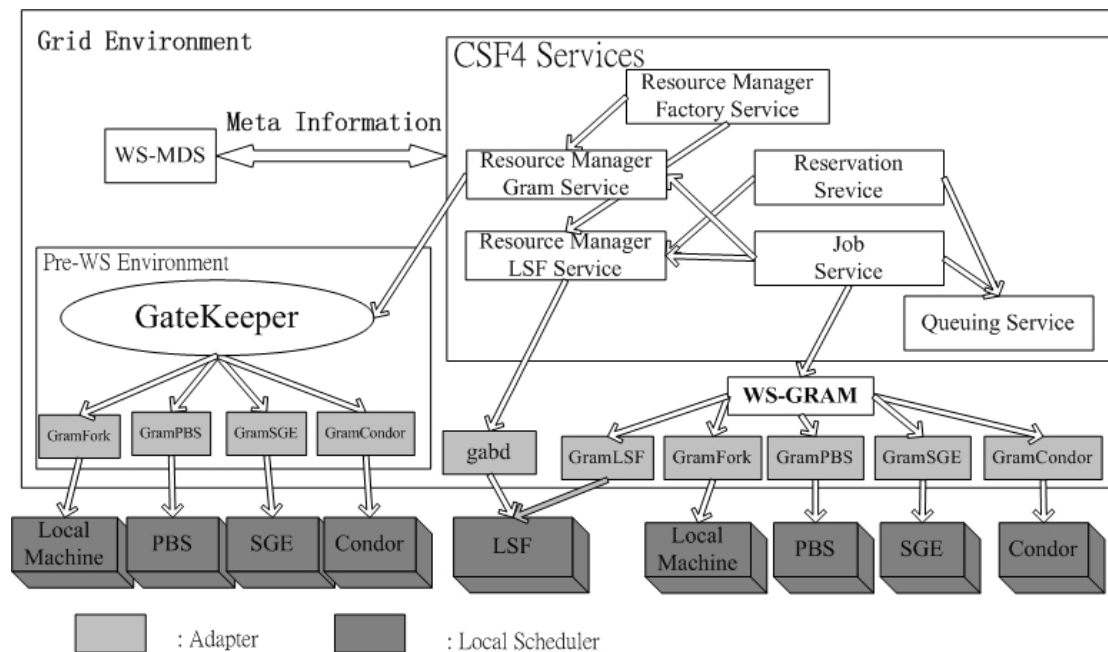


Figure 2 CSF4 Architecture

CSF4 consists of a bunch of web services, which are Job Service, Reservation Service, Queuing Service, and Resource Manager Factory Service etc. See Figure 2. Job Service and Reservation Service provide the virtualized interface for end users to submit jobs and reserve resources. In concept, Queuing Service is the container holding the jobs, reservation requests and the scheduling policies to be applied. Multiple queues can be configured in CSF, and different queues may have different scheduling policies. At submission time, the user should

specify a queue for the job. Otherwise, it will be put into the default queue.

Resource manager adapters are responsible for aggregating information of distributed resources from local resource managers. Such information will be put into Global Index Service for scheduling jobs and reserving resources. The GRAM protocol is used in the communications between CSF4 and local job schedulers. Since LSF, PBS, and Condor use different protocols for local job submission and resource management, the GRAM adapters for specific local schedulers are included in GT4 packages, such as Gram-Fork, Gram-PBS and Gram-Condor etc. However, there is no such an adapter for SGE shipped with GT4. The one we used for testing is developed by London e-Science Centre, Gridwise Technologies and MCNC.

However, it is not good enough to support WS-GRAM protocol only. For example, some users still need GT2 environments to run their legacy applications. Hence, the Resource Manager Factory Service and its several instance services are designed to make CSF4 more extensible. Currently, two instance services are available for Resource Manager Factory Service. One is Resource Manager Gram Service which is designed to support Pre-WS-GRAM protocol. The other one, Resource Manager Lsf service, is used to support resource reservation that is not supported by GRAM yet. More details will be discussed in the next section.

3. CSF4 Services

CSF4 consists of a number of web services hosted in GT4 container such as, Job Service, Reservation Service, Queuing Service, and Resource Manager Services etc. See Figure 3. Among them, Job Service, Reservation Service and Queuing Service provide the interface for end users to submit job, reserve resource and enforce scheduling policies. Resource Manager Services are designed to support the alternative protocols other than WS GRAM.

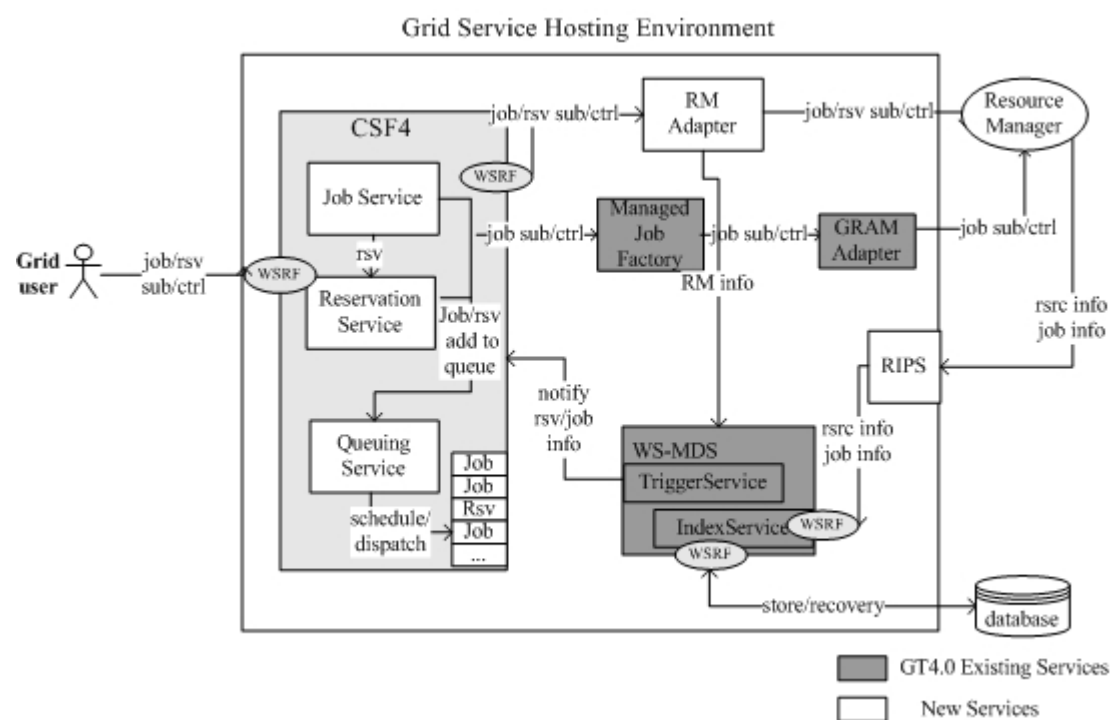


Figure 3 CSF4 scheduling functionalities

- **Job Service**

Job Service provides the interfaces for end users to fully control a job. The users are able to create job instances, submit jobs to a queue, modify a job's description and monitor job status etc. Once created, the job's EPR will be returned to the user for further operations. A new created job does not belong to any queue until a submission operation is issued. CSF jobs are described in RSL^[8]. And a number of new tags are introduced by CSF, for example, <ReservationID>, <ClusterName>, <QueueName> and so on. With using new tags, the users can specify a job's queue name, reserved resource ID, or execution cluster name. However, as the new tags are not recognized by GRAM, such jobs have to be dispatched via Resource Manger Services.

- **Reservation Service**

Reservation Service allows the users to reserve the resources for their jobs in advance so that the availability of the resources can be guaranteed. It is a very useful feature for large-scale and critical jobs. The user's reservation requests will be put into a queue first, and then be scheduled to the local scheduler by Queue Service. However, since GRAM does not support resource reservation functionality, the reservation requests are forwarded to local schedulers via specific Resource Manager Service instance, for example, Resource Manager Lsf Service.

Both the jobs and reservation requests are hosted in GT4 container as RPs (Resource Property), and their EPRs will be returned to the user for future operations. In the mean time, those EPRs are also saved in WS-MDS as well. It will provide the below advantages. First, the recovery mechanism of GT4 Index Service will make the jobs and reservations persistence after reboot. Second, according to customized conditions and RPs status, Trigger Service is able to notify the clients once their jobs or reservations status changed.

- **Queuing Service**

In CSF, a queue represents a set of specified scheduling policies and the associated jobs and reservation requests. The queue's scheduling policies are specified by the admin through configuration. After a queue service instance created, the user can submit jobs and reservations to the queue. Then the queue will start to schedule the jobs and reservations periodically. Currently, only FCFS and Throttle polices are provided by CSF4. CSF4's design philosophy is to provide an extensible framework and enable the end users to develop tailored scheduling polices themselves. So CSF4 implements a scheduler plugin prototype to dynamic load scheduling modules, however, it is not quite matured.

- **Resource Manager Services**

Resource Manger Services are not used by end users directly. They are designed to support alternative protocols other than WS GRAM. Resource Manager Services consist of one factory service, *Resource Manager Factory Service*, and two instance services, *Resource Manager Lsf Service* and *Resource Manager Gram Service*.

Resource Manager Lsf Service is an instance service designed to support non-GRAM protocol between CSF4 and LSF. Some advanced features, such as resource reservation and extended RSL tags, are supported via this service. However, a special component, *gabd*, need

be deployed as an adapter between the CSF4 and LSF clusters. Although *Resource Manager Lsf Service* is for LSF only, similar instance services can be designed for SGE, and PBS as well by following the same idea.

Although Globus Toolkit 4 has been released for a while, however, it is not reasonable to expect that all users can switch to GT4 platform immediately, especially for those who have legacy applications. Hence, CSF4 provides the *Resource Manager Gram Service* to support GRAM2 protocol via Java Commodity Grid Kit^[9] library. The service recognizes the job descriptions in both RSL format and non-RSL format, and is able to work with both WS-GRAM and Pre-WS GRAM protocols. Therefore, CSF4 can be deployed in a GT4 and GT2 mixed grid environment.

4. Deploy CSF4 In A Grid Community

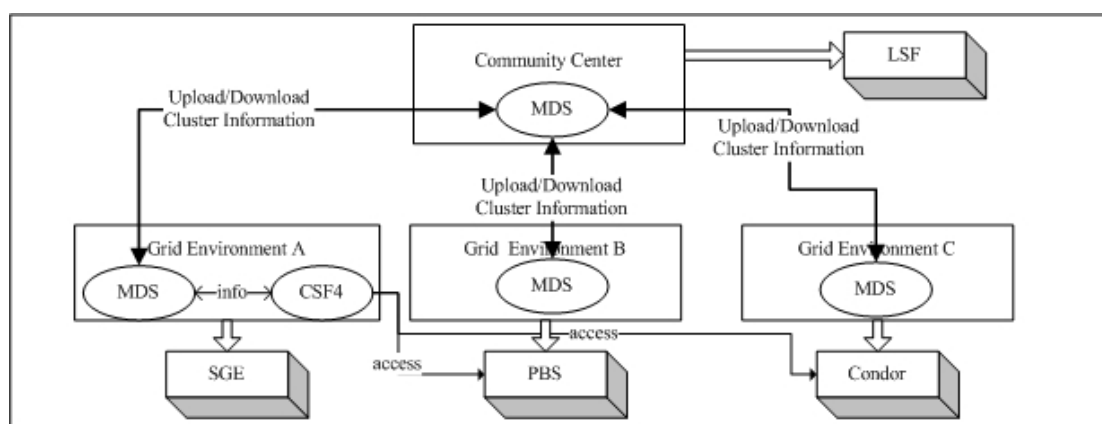


Figure 4 Resource Sharing in A Grid Community

CSF4 also supports resource sharing in a grid community environment. For example, in Figure 4, there are four independent grid systems. Each sub-grid system has CSF4 and local resource managers installed. These sites can be deployed as a community via configuration. And one of the sites will be deployed as the community center.

After that, all the sites in the community will upload their resource information from local MDS to the community center's MDS. And the center MDS will also publish the aggregated resource information to the sites in the community. Hence, each site's MDS will have the global resource information. As long as the users in the community are mutual trust at each site, the jobs can be dispatched to anywhere in the community. For instance, the jobs submitted from site A can be forwarded by to Site B or Site C's clusters to run. Therefore, CSF4 can be used as a grid community scheduler.

5. Related Works

CSF4's previous version, CSF3^[10], is a OGS^[11] compliant meta-scheduler contributed by Platform. CSF4 re-implements all the functionalities of CSF3 based on WSRF specification with some enhancements. For example, CSF4 provides the backward compatibility to GT2. Gridway^[12] is another meta-scheduler that can work with both Pre-WS and WS GRAM services simultaneously. GridWay is a light-weight meta-scheduler that performs job execution management and resource brokering. However, GridWay works on top of GT4,

while CSF4 is part of GT4 and deployed as a bunch of web services inside GT4 container.

MARS^[13] is an open-source meta-scheduling framework that can be integrated into existing campus infrastructure to provide task scheduling and resource management capabilities. MARS supports both Globus Pre-WS GRAM and PBS protocols.

Condor-G^[14] is a grid-level task manager to integrate Condor with the Globus middleware. It allows the jobs to be executed on remote GRAM-enabled computers, and provides a number of features including job monitoring, logging, notification, policy enforcement, fault tolerance, job migration and credential management. But it only supports Pre-WS GRAM.

Silver^[15] is an advance reservation based meta-scheduler. It provides single point access to multiple independently managed local clusters, like PBS, Torque and SGE. Users or systems can submit workload to Silver and have Silver determine where and when to run the job. Silver can inter-operate with GT2 services, like gatekeeper and GridFTP services, to do job migration, data staging, and credential management.

6. Conclusion

This paper describes CSF4, a WSRF compliant community meta-scheduler framework meta-scheduler. It provides job submission, job management, resource advance reservation, and policy enforcement. Moreover, CSF4 supports both WS-GRAM and Pre-WS-GRAM so that it can be used in a GT4 and GT2 mixed grid environment. We have deployed CSF4 on PRAGMA's^[16] grid test bed and successfully integrate it with Gfarm^[17], SGE and LSF to support data-intensive bioinformatics applications.

In the future, we are going to introduce more advance scheduling policies to CSF4, such as data aware scheduling, and co-scheduling etc. However, in the real world, each user has a different requirement. No matter how many scheduling polices are provided, no scheduler can meet all users' needs. Therefore, it is also a priority to us to enhance CSF's scheduler plugin prototype so that the end users can develop tailored scheduling policies in convenient.

7. Acknowledgements

This work is supported by Jilin University and Platform under Grant 419070200053, 420010302338, and 3B6056721421.

References

- [1] Songnian Zhou, Xiaohu Zheng, Jingwen Wang *et al.* Utopia: a Load Sharing Facility for Large, Heterogeneous Distributed Computer Systems[J]. SOFTWARE—PRACTICE AND EXPERIENCE, Dec 1993: 23(12), 1305–1336 .
- [2] James, P. J, Portable Batch System: Exterernal Reference Specification Altair PBS Pro 5.3[M]. <http://www.mta.ca/torch/pdf/pbspro54/pbsproers.pdf>, March 2003.
- [3] Sun Microsystems, Inc. Sun Grid Engine 5.3 Administration and User's Guide[M]. <http://gridengine.sunsource.net/project/gridengine-download/SGE53AdminUserDoc.pdf>, April, 2002.
- [4] Jim Basney and Miron Livny, "Managing Network Resources in Condor"[C]. Proceedings of the Ninth IEEE Symposium on High Performance Distributed Computing (HPDC9), Pittsburgh, Pennsylvania, August 2000, pp 298-299.
- [5] Marty Humphrey, Glenn Wasson, Jarek Gawor, Joe Bester, Sam Lang, Ian Foster, Stephen Pickles, Mark Mc Keown, Keith Jackson, Joshua Boverhof, Matt Rodriguez, Sam Meder,

- “State and Events for Web Services: A Comparison of Five WS-Resource Framework and WS-Notification Implementations” 14th IEEE International Symposium on High Performance Distributed Computing (HPDC -14), Research Triangle Park, NC, 24-27 July 2005
- [6] I. Foster. “Globus Toolkit Version 4: Software for Service-Oriented Systems.” IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp 2-13, 2005
- [7] I. Foster, C. Kesselman, S. Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International J. Supercomputer Applications*, 15(3), 2001.
- [8] K. Czajkowski, I. Foster, N. Karonis, C. Kesselman, S. Martin, W. Smith, S. Tuecke, A Resource Management Architecture for Metacomputing Systems. *Proc. IPPS/SPDP '98 Workshop on Job Scheduling Strategies for Parallel Processing*, pg. 62-82, 1998.
- [9] Commodity Grid Kits - Middleware for Building Grid Computing Environments, Gregor von Laszewski, Jarek Gawor, Sriram Krishnan, and Keith Jackson. Chapter in *Grid Computing: Making the Global Infrastructure a Reality*, pages 639–656. Communications Networking and Distributed Systems. Wiley, 2003
- [10] Platform Computing Co. Open source metascheduling for Virtual Organizations with the Community Scheduler Framework (CSF)[WP]. http://www.cs.virginia.edu/~grimshaw/CS851-2004/Platform/CSF_architecture.pdf , 2004.
- [11] I. Foster, ANL, J. Frey, IBM, S. Graham, IBM, C. Kesselman, USC/ISI, T. Maquire, IBM, T. Sandholm, ANL, D. Snelling, Fujitsu Labs, P. Vanderbilt, NASA “Open Grid Services Infrastructure(OGSI) version 1.0” April 5, 2003
- [12] http://www.globus.org/grid_software/computation/gridway.php
- [13] A. Bose, B. Wickman and C. Wood, MARS: A Metascheduler for Distributed Resources in Campus Grids, 5th IEEE/ACM International Symposium on Grid Computing (Grid2004), October 8, 2004, Pittsburgh.
- [14] J. Frey, T. Tannenbaum, I. Foster, M. Livny, S. Tuecke, Condor-G: A Computation Management Agent for Multi-Institutional Grids. *Cluster Computing*, 5(3):237-246, 2002.
- [15] <http://www.supercluster.org/projects/silver/>.
- [16] http://pragma-goc.rocksclusters.org/pragma-doc/pragma9/wilfred-li_igap-gfarm-csf.ppt
- [17] Osamu Tatebe, Youhei Morita, Satoshi Matsuoka *et al.* Grid Datafarm Architecture for Petascale Data Intensive Computing [C]. *Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid*, pp.102-110, 2002.