IL NUOVO CIMENTO Vol. 32 C, N. 2 Marzo-Aprile 2009
DOI 10.1393/ncc/i2009-10412-x

Colloquia: CSFI 2008

The gLite Workload Management System

M. Cecchi(1), F. Capannini(1), A. Dorigo(1), A. Ghiselli(1), F. Giacomini(1), A. Maraschini(2), M. Marzolla(1), S. Monforte(1), F. Pacini(2), L. Petronzio(2) and M. Sgaravatto(1)

(1) *INFN, Sezione di Bologna - Viale Berti Pichat 6/2, Bologna, Italy*
(2) *Elsag-Datamat s.p.a. - Via Laurentina 760, Rome, Italy*

(ricevuto il 4 Agosto 2009; pubblicato online il 6 Ottobre 2009)

Summary. — The gLite Workload Management System represents a key entry point to high-end services available on a Grid. Being designed as part of the European Grid within the six years long EU-funded EGEE project, the WMS is meant to provide reliable and efficient distribution and management of end-user requests. The WMS basically translates user requirements and preferences into specific operations and decisions—dictated by the general status of all other Grid services—while taking responsibility to bring requests to successful completion. The WMS implements an "early binding" approach to meta-scheduling as a neat Grid-aware solution, able to optimise resource access and to satisfy requests for computation together with data. Several added value features are provided for job submission, different job types are supported from simple batch to a variety of compounds, all described in this paper. Conceived to be a scalable service, the WMS has been deployed in different physical layouts according to specific needs for performance, thanks to the modularity of its design. In this paper we show what has been achieved to provide adequate workload and management components, suitable to be deployed in a production-quality Grid while covering the design and development of the gLite WMS.

PACS 89.20.Ff – Computer science and technology.

1. – Introduction

Resource management and scheduling of distributed, data-driven applications in production Grid environments are challenging problems. The interested domains include workload management, resource discovery, brokering, accounting, authorization and authentication, resource access, reliability and dependability. Although significant results were achieved in the past few years, the development and the proper deployment of generic, robust, reliable and standard components involving such huge scales and factors as the ones a production Grid has to deal with, has brought out non-trivial issues requiring joint efforts with a strong degree of cooperation to be attained.

2. – The gLite WMS in a nutshell

The gLite WMS represents a key entry point to high-end services available on a Grid. It has been designed with some fundamental principles in mind: first of all aiming at providing a dependable and reliable service, where primary importance is given to never losing track of jobs to be processed and always providing a prompt, responsive quality of service, yet keeping up with huge and even growing factors of scale. It is designed as part of a Service Oriented Architecture (SOA) complying with Web-Service Interoperability (WS-I) specifications and strives to implement recommendations on web service foundations made by the Open Grid Forum (OGF).

Fundamental to any Grid environment is the ability to discover, allocate and monitor the use of resources. The term "workload management" is commonly used to describe all those aspects that involve discovering the resources and selecting the most suitable ones, arranging for submission, monitoring and information gathering. In this respect, the WMS has to deal with a heterogeneous computing environment that in general encompasses different architectures and loss of centralized control, all this in the presence of potential faults due to the distributed and diverse nature of the Grid environment, computers, networks and storage devices.

3. – Functionality

The WMS does support different types of jobs:

- Single batch jobs;
- DAGs: jobs with dependencies expressed as a direct acyclic graph (DAG);
- Collections: sets of jobs without dependencies grouped together and identified by a single handler;
- MPI: based on message passing interface—a widely-used library to allow for parallel programming within a single cluster (intra-cluster);
- Interactive: establishing a synchronous two way communication with the user on a socket stream;
- Parametric: allowing multiple jobs to be defined by a single description with attributes varying with a parameter.

The characteristics of a job are defined using a flexible and expressive formalism called Job Description Language (JDL) which basically consists of a list of key/value pairs that represent the various characteristics of a job (input files, arguments, executable, etc.) as well as its requirements, constraints and preferences (physical and virtual memory, CPU, operating system, etc.).

Here is a summary of the more relevant functionalities implemented in the gLite WMS:

- Resubmission: shallow or deep;
- Stochastic ranking;
- Bulk-submission and bulk-matchmaking;

- Parallel match-matchmaking;
- Proxy renewal;
- Support for MPI jobs even if the file system is not shared between CE and Worker Nodes (WN);
- Support for execution of all DAG nodes within a single CE—chosen by either user or by the WMS match-maker;
- Support for file peeking to access files during job execution;
- Load limiting mechanism to prevent system congestion based on machine's vital parameters;
- Automatic sandbox files archiving/compression and sharing between jobs;
- Match-making with data;
- Gang-matching.

4. – Interoperability

Given the typically large number of different parties involved in a Grid infrastructure, interoperability plays a key role to facilitate establishing and coordinating agreements and interactions between all the involved entities. In this respect, the WMS, especially by virtue of his central, mediating role, has to deal with a wide variety of people, services, protocols and more, ranging from users—belonging to different VOs—to other services of the EGEE/gLite infrastructure and to other Grids as well.

For what concerns users, to be able to allow interaction adhering to the SOA model, a Simple Object Access Protocol (SOAP) Web Service has been implemented, its interface being described through a Web Service Description Language (WSDL) specification written in accordance to the WS-I profile, which defines a set of Web Services specifications to promote interoperability. This newly introduced Web Service based implementation replaced a legacy network interface based on a proprietary protocol. It manages user authentication/authorization and operation requests.

Here is a summarized view of the functionality provided in the areas of integration with other services and interoperability:

- Backwards compatibility with LCG-2;
- Automatic renewal of credentials;
- GridFTP and HTTPS to handle secure file transfer for the sandbox;
- Service Discovery for obtaining new serivice endpoints to be contacted;
- Support of different mechanisms to populate the ISM from several sources (BDII, R-GMA, CeMon);
- Support for submission and monitoring for the LCG, gLite and CREAM CEs;
- Support for Data management interfaces (DLI and StorageIndex);

– Support for JSDL;

– Support for Grid Site delegation 2.0;

– Interoperability with the american Open Science Grid (OSG), Nordugrid and UNICORE;

– Integration with D-GAS—a Grid accounting system;

– User prologue/epilogue scripts accompanying the job, more custom scripts allowed to be hooked for use by resource and VO administrators.

5. – Conclusions

The gLite WMS is designed and implemented to provide a dependable, robust and reliable service adopting open standards to promote interoperability among Grid services and allowing easier compliance with emerging protocols. This is done through a fully-fledged set of added-value features built on top of the Job Submission service and with the flexibility of a Service Oriented Architecture.

Development continues by supporting enhancements requested by the experiments and keeping up with compliancy to formal and *de facto* standards. We will also continue facing the challenge of reaching even higher levels of performance, scalability and reliability to find us prepared to meet the growing demand of the EGEE infrastructure.