WMSMonitor advancements in the EMI era

D. Cesini, D. Dongiovanni, E. Fattibene

INFN-CNAF - Ist. Naz. di Fisica Nucleare Viale Berti Pichat, 6/2 40126 Bologna (Italy)

E-mail: wms-support@cnaf.infn.it

Abstract. In production Grid infrastructures deploying EMI (European Middleware Initiative) middleware release, the Workload Management System (WMS) is the service responsible for the distribution of user tasks to the remote computing resources. Monitoring the reliability of this service, the job lifecycle and the workflow pattern generated by different user communities is an important and challenging activity. Initially designed to monitor and manage a distributed cluster of gLite WMS/LB (Logging and Bookkeeping) services, WMSMonitor has proved to be a useful and flexible tool for a variety of user categories. In fact, after asynchronously extracting information from all monitored instances, WMSMonitor re-aggregates it by different keys (WMS instance, Virtual Organization, User, etc.) providing insight both on services status and on their usage to service administrators, developers, advanced Grid users and performance testers. The positive feedback on WMSMonitor utilization from various production Grid sites pushed us to improve the tool to enhance its flexibility and scalability exploiting a new architecture. Moreover the tool has been made compliant to recent evolutions in the monitored services. We therefore present the new version of WMSMonitor which can monitor EMI WMS/LB services and shows an improved user interface allowing better report capabilities. Among main novelties, we mention the collection of Job Submission Service (JSS) error type statistics and the adoption of ActiveMQ messaging system which now allows multiple data consumers to exploit collected information. Finally, it is worth to mention that the implemented architecture and the exploitation of a messaging layer commonly adopted in EMI Grid applications make WMSMonitor a flexible tool that can be easily extended to monitor other Grid services.

1. Introduction

In the framework of the EMI Grid middleware [1], the WMS is the service responsible for the job scheduling and dispatching on the remote computing resources. It is coupled with the LB service which is commissioned to store information about the events that characterize the job lifecycle. These services have been developed as part of the gLite middleware stack and continued to evolve within the EMI middleware.

On the basis of the Italian WMS/LB cluster administration experience, WMSMonitor [2] was developed in order to make easier the service configuration, run and problem debugging. It was designed to provide WMS/LB status and usage information that could be useful to different kind of users, such as service administrators and developers, advanced Grid users and performance testers. In the last years this tool has been used to simplify the WMS/LB clusters administration in gLite based
production Grid infrastructures. In order to continue to supply its functionalities, WMSMonitor has recently adapted to the updated characteristics of WMS/LB services.

In this paper we present the new version of WMSMonitor. The paper is organized as follows: in Section 2 recent evolutions of the WMS/LB services are mentioned; in Section 3 we present the new WMSMonitor architecture in details; Section 4 gives an overview of the graphical web interface, with particular attention to the functionalities added to the previous version; Section 5 concludes the paper.

2. Monitoring EMI WMS/LB services

EMI is a close collaboration of four major European technology providers: ARC [3], gLite [4], UNICORE [5] and dCache [6]. Its main objective is to deliver a consolidated set of components for deployment in the European Grid Initiative (EGI) [7] (as part of the Unified Middleware Distribution, UMD[8]), PRACE [9] and other Distributed Computing Infrastructures (DCIs). The harmonized set of EMI components thus enables the interoperability and integration between Grids. EMI aims at creating an effective environment that satisfies the requirements of the scientific communities relying on it.

The EMI WMS is a Grid service responsible for the distribution of user tasks (referred in this paper as “jobs”) to the remote computing resources. The WMS supports this process by tracking events related to the tasks during their lifecycle, by adopting failure recovery procedures in case of problems and by performing resource discovery. This last step relies on the availability of information from a Grid Information Service publishing the set of available remote resources from which the list of candidates is selected in order to match the requirements specified by the users.

The LB service is tightly coupled to the WMS and tracks jobs managed by the WMS. It gathers events from Grid components providing the user with detailed information about the status of jobs during their whole lifecycle. The WMS can handle various job types such as: single batch jobs, MPI jobs, Direct Acyclic Graph jobs (DAG – to handle job sets with complex internal dependencies), job collections (handling groups of independent jobs) and finally parametric jobs (i.e. group of tasks depending on a user-defined parameter).

Both WMS and LB entered the new EMI middleware stack implying several improvements as the adoption of EPEL[10] distribution standards and development of new features. The description of the two EMI services is beyond the scope of this document, and for a review of the new features introduced in the EMI era the reader can refer to [13][14].

3. WMSMonitor new architecture and features

3.1. WMSmonitor overview

WMSMonitor is a tool to monitor a pool of distributed WMS/LB instances, the EMI services responsible for the job submission to Grid resources. It allows early detection of failures affecting the services and supports administrators in fault prevention.

WMSMonitor collects two main categories of metrics for the WMS/LB cluster: i) data on each instance hardware utilization (disk, RAM memory, cpu) and service functionality (daemons status, internal components queue status) ; ii) usage statistics aggregated per WMS or VO or User over configurable time intervals. WMSMonitor also provides statistics about WMS internal dynamic, reporting about rates of processed Grid jobs across internal components. This is useful to service administrators to discover possible causes of faults in specific components. A full description of WMSMonitor metrics is also provided in [2].

Several administration facilities are provided to help the administration of big clusters of WMS/LB services. For example it provides an automated cluster status analyzer which periodically sends notifications to the NAGIOS alarm system [11]. Another feature widely exploited at INFN-CNAF is the DNS based cluster loadbalancing system: WMSMonitor calculates a WMS service global status metric for each monitored instance and ranks instances according to that metric. This rank is then exploited by a loadbalancing arbiter which dynamically adds to a DNS alias the subset of the best
instances removing those overloaded or not functioning. Different DNS aliases are supplied to different Virtual Organizations. For a full description of the loadbalancing module refer to [12].

3.2. WMSMonitor 3.0 improvements
In this section we will focus on WMSMonitor 3.0 novel features. First main improvement is the completely renewed tool architecture represented in figure 1. As shown in the figure, sensors execution is no longer triggered by the central server and it is now autonomously started by cron jobs on the clients. Another important difference is the new implementation of the messaging layer, now based on the ActiveMQ service [15]. The independent sensors execution on clients together with the adoption of independent messaging module allows WMSMonitor 3.0 to have an asynchronous data collection strategy, improving both scalability and robustness of the monitoring process to faults. Another advantage of the ActiveMQ messaging service adoption is the possibility of having multiple consumers for a given data producer (WMS or LB instance).

Other changes in the new WMSMonitor version resulted from compliance to EMI evolutions of the WMS and LB services. In particular now both Condor [16] and ICE [17] job submission services are monitored implying modification to the web interface (see Section 4). Also the collection of Grid jobs rates information from LBs has changed to be compliant with the new LB database implementation. Notice that this last major change makes 3.0 version not backward compatible with previous WMSMonitor server versions, implying that the new EMI versions of WMS and LB services must be monitored by the new WMSMonitor 3.0 server versions.

Another important feature introduced in the 3.0 version is the collection of metrics regarding job submission service error type statistics as shown in Section 4.

![Figure 1 - WMSMonitor 3.0 architecture](image)
4. The presentation layer

The WMSMonitor 3.0 web interface is PHP-based and implements OpenFlashChart2 plot library.

The monitoring data are organized following two keys of investigation: WMS and VO. For each key, the web navigation is designed to permit the user to analyze the services status and utilization from a general to a detailed view.

The main page of the WMS navigation key aims at providing the general status of the monitored services at a glance, highlighting the critical situations. In addition to the latest measurement results, intuitive icons summarize the general status of the instance (for each metric, thresholds have been defined to identify warning or critical situations) and the status of the daemons. The new version of WMSMonitor contains this kind of general view also for the monitored LB instances. Moreover it’s possible to filter the WMS and LB instances by the hosting Grid site, the service usage (test, production, etc.) and supported VO (in the case that the instance is dedicated to a particular VO).

Figure 2 shows a screenshot of the LB summary.

| LB                  | DATE       | CPU LOAD | DISK LB | VARI LB | MYSQL | LB CON | LB Daemon | SL Daemon | NTDP Daemon | GENERAL Daemon |
|---------------------|------------|----------|---------|---------|-------|--------|-----------|-----------|-------------|----------------|
| cg07.dic.uv.es      | 2012-06-21 | 0.1000   | 74      | 74      | 0     | 0      | 0         | 0         | 0           | ✔              |
| b-sern.cern.ch      | 2012-06-21 | 0.0000   | 9       | 9       | 0     | 0      | 0         | 0         | 0           | ✔              |
| b013.cern.ch       | 2012-06-24 | 0.0000   | 13      | 13      | 0     | 0      | 0         | 0         | 0           | ✔              |
| ngr-b-b1           | 2012-06-21 | 0.0100   | 39      | 44      | 0     | 0      | 0         | 0         | 0           | ✔              |
| wms-c4.cern.ch      | 2012-04-24 | 0.0000   | 87      | 87      | 0     | 0      | 0         | 0         | 0           | ✔              |
| wmsu01.ics.polsl    | 2012-06-21 | 2.3500   | 22      | 76      | 0     | 0      | 0         | 0         | 0           | ✔              |
| wmsu02.ics.polsl    | 2012-04-27 | 0.3700   | 23      | 1       | 67    | 0      | 0         | 0         | 0           | ✔              |
| wmsu03.ics.polsl    | 2012-06-21 | 0.0800   | 5       | 3       | 0     | 0      | 0         | 0         | 0           | ✔              |
| wms04.ics.polsl     | 2012-06-21 | 0.4600   | 48      | 48      | 6     | 0      | 0         | 0         | 0           | ✔              |

**Figure 2:** LB summary view

For each monitored WMS, detailed pages are provided with textual and graphic information. The possibility to save each plot as an image as requested by users has been implemented. A new section shows the statistics about the JSS error occurred in the desired period of time, as shown in figure 3.
Figure 3: JSS error statistics page
5. Conclusion

In this paper we presented the new advancements in WMSMonitor, a tool that monitors and eases the management of clusters of EMI WMS/LB services. The architecture of the tool has been revised to enable independent sensors execution (not triggered by a central server) and to exploit the messaging system based on ActiveMQ for the data transport layer. The needed adjustments to make the tool compliant with the new EMI WMS and LB services have been implemented. Moreover the web interface was enriched to encompass the new data sets collected by the sensors and to meet user requirements. New views were added (i.e. JSS error statistics and LB summary pages) and the user experience was improved with new filters and buttons.

The adoption of the ActiveMQ messaging system in this new version of WMSMonitor, facilitates the monitoring of geographically distributed instances and allows multiple consumers to use the data produced by the sensors. In addition, being ActiveMQ a new standard for messaging layer for several Grid services and applications, the flexible architecture of the new WMSMonitor tool can be easily extended to monitor other Grid services. It is worth mentioning that the WMSMonitor server at INFN-CNAF in Italy is already collecting data from WMS/LB clusters spread across Europe besides those coming from national clusters.

6. References

[1] European Middleware Initiative (EMI) http://www.eu-emi.eu/
[2] Cesini D., Dongiovanni D., Fattibene E., Ferrari T.; WMSMonitor: a Monitoring Tool for Workload and Job Lifecycle in Grids; 2008; In Proceedings of the 2008 9th IEEE/ACM International Conference on Grid Computing; IEEE Computer Society, Washington, DC, USA, 209-216 doi: 10.1109/GRID.2008.4662801
[3] Appleton O. et al; The next-generation ARC middleware; Annals of telecommunications, December 2010, Volume 65, Issue 11-12, pp 771-776; doi: 10.1007/s12243-010-0210-2
[4] Hemmer F., et al.; Middleware for the Next Generation Grid Infrastructure; In Proceedings of the Conference on Computing in High-Energy Physics 2004, Intelaken, Switzerland, 2004
[5] Streit A. et al; UNICORE 6-Recent and Future Advancements; Annals of Telecommunications, 2010, Volume 65, Numbers 11-12, Pages 757-762; doi: 10.1007/s12243-010-0195-x
[6] Fuhrmann P. et al; dCache, agile adoption of storage technologies; to appear in Proceedings of GRID’2012, The 13th ACM/IEEE International Conference on Grid Computing, 20-23 September 2012, Beijing, China
[7] Andreozzi S. et al; EGI: an Open e-Infrastructure Ecosystem for the Digital European Research Area; To appear in Proceedings of e-Challenges e-2012 Conference, Lisbon, Portugal, 17-19 October 2012
[8] UMD release http://repository.egi.eu/category/umd_releases/
[9] PRACE http://prace-project.eu
[10] EPEL http://fedoraproject.org/wiki/EPEL
[11] Barth, Wolfgang; 2008; Nagios: System And Network Monitoring, 2nd edition (No Starch Press); ISBN 1-59327-179-4
[12] Cesini D., Dongiovanni D. Fattibene E., Ferrari T.; On gLite WMS/LB Monitoring and Management through WMSMon; 2009; Journal of Physics: Conf. Ser. 219 062040; doi: 10.1088/1742-6596/219/6/062040
[13] Andreetto P. et al.; Evolution of Grid meta-scheduling towards the EMI era; 2012; In Proceedings of Conference on Computing in High-Energy Physics 2010, Taipei (Taiwan)
[14] Sustr Z. et al.; Logging and Bookkeeping – What Can it Do for You?; In Proceedings of the European Grid Infrastructure User Forum, 2011, Vilnius (Lietuva)
[15] ActiveMQ messaging service http://activemq.apache.org/
[16] Tannenbaum T., Wright D., Miller K., Livny M.; Condor - A Distributed Job Scheduler, in Thomas Sterling, editor, Beowulf Cluster Computing with Linux, The MIT Press, 2002. ISBN:0-262-69274-0

[17] Aiftimiei C. et al.; Design and Implementation of the gLite CREAM Job Management Service; Future Generation Computer Systems, Volume 26, Issue 4, April 2010, pp. 654-667, doi: 10.1016/j.future.2009.12.006