A Job Monitoring and Accounting Tool for the LSF Batch System

Subir Sarkar† and Sonia Taneja
Scuola Normale Superiore, Pisa & INFN, Sezione di Pisa, Italy

Abstract. This paper presents a web based job monitoring and group-and-user accounting tool for the LSF Batch System. The user oriented job monitoring displays a simple and compact quasi real-time overview of the batch farm for both local and Grid jobs. For Grid jobs the Distinguished Name (DN) of the Grid users is shown. The overview monitor provides the most up-to-date status of a batch farm at any time. The accounting tool works with the LSF accounting log files. The accounting information is shown for a few pre-defined time periods by default. However, one can also compute the same information for any arbitrary time window. The tool already proved to be an extremely useful means to validate more extensive accounting tools available in the Grid world. Several sites have already been using the present tool and more sites running the LSF batch system have shown interest. We shall discuss the various aspects that make the tool essential for site administrators and end-users alike and outline the current status of development as well as future plans.

1. Introduction
A web based job and group-and-user accounting monitor for the LSF Batch system has been developed for both local and Grid jobs. Both site administrators and end-users will find the monitor useful in spotting problems with jobs at the earliest so that CPU time and hence resource utilisation can be optimised. Moreover, it offers an independent way to validate more extensive monitoring and accounting tools available in the Grid world.

The monitor provides the most up-to-date overview of the state of a batch farm at any time. The user oriented view shows a compact status summary of the local batch farm. For Grid jobs, the Distinguished Name (DN) of the Grid certificate is used to identify users uniquely. The accounting tool solely depends on the LSF accounting log files and does not need any external database support. The accounting monitor primarily shows information for a few pre-defined time intervals. However one can also get the same information for any arbitrary time window using a command line tool in order to prepare monthly reports etc. It might be noted that accounting information for recently completed jobs, e.g. jobs completed in the last 12 hours or so, can be reliably obtained only with such a local tool.

A few sites have been using the present tool and more sites running the LSF batch system have shown interest. We shall discuss the various features that make the monitor indispensable for site administrators and end-users alike and outline the current status of development and plans for the near future.

†Corresponding author, email: subir.sarkar@cern.ch

International Conference on Computing in High Energy and Nuclear Physics (CHEP 2010) IOP Publishing
Journal of Physics: Conference Series 331 (2011) 072064 doi:10.1088/1742-6596/331/7/072064
Published under licence by IOP Publishing Ltd
2. Objective and Monitorable

The monitoring system has been built with the aim to publish at regular intervals the most reliable data on availability of resources, current usage, job flow, usage history, job efficiency, and share utilisation at a batch system. We describe below the various monitorable quantities already available in the present monitor:

- **Overview** - broadly classified as (a) CPU resource availability, usage, and user share, and (b) job status at the batch farm. To suit both site administrators and end-users, in addition to overall job status we also collect information for each (1) group/Virtual Organisation (VO), (2) Computing Element, and (3) user/Distinguished Name (DN). Each individual sub-set shows total, running, pending, and held jobs in the queue, average CPU efficiency, defined as a ratio of total CPU time over total wall-time, number of jobs with very low CPU efficiency and jobflow, an indicator of farm activity defined as the number of submitted, dispatched, and completed jobs in the last hour. For groups we also show the wall-time share, which is a ratio of the total wall-time used by a group over the overall wall-time used at the farm at the time when information is collected.

- **History** - time series plots for (a) resource availability, and occupancy at the farm, and (b) number of running and pending jobs, and CPU efficiency for the whole farm as well as supported groups/VOs.

- **Accounting** - group-and-user accounting shows important quantities like the (a) job share and success rate, (b) CPU usage and efficiency, (c) wall-time share, and (d) average time in seconds to wait in queue etc. for different time intervals for groups/VOs and individual users.

3. Description

The LSF job and accounting monitor implements a simple design shown in Figure 1.

![Diagram of the monitoring framework](image)

**Figure 1.** Schematic view of the monitoring framework.

The information collector consists of three distinct components: (1) a sensor that uses the LSF commands and Grid job definition files to get the relevant farm and job overview data, (2) a simple parser that reads the LSF accounting files to collect the group accounting information for pre-defined time periods, and (3) a parser that correlates LSF accounting data with the user DN to get the user accounting statistics. The sensors must run on services that can access...
the relevant files. The monitoring information is eventually published as web pages and also as XML documents for further processing by applications like central monitors in an experiment [1], widget-based monitoring tools etc.

The collector side sensors are written in Perl. Several Perl modules are required, namely (a) RRDtool to save and plot historical information about job overview, (b) A templating engine to generate HTML, (c) GD and ImageMagick to generate accounting charts, and (d) XML tools etc., that must be installed. The sensors run as cron jobs. The web interface uses two different JavaScript libraries, namely Ext JS [2] for the job overview and group accounting monitor and jQuery [3] for the user accounting monitor, in order to build a rich user interface with tabbed view, table with advanced features like sorting, searching etc. and Ajax calls. The monitor supports only standard features and all the major browsers are supported.

4. Web Interface

In the following subsections we shall briefly discuss the various parts of the monitoring interface. Note that presently there are two different web pages (a) job overview and group accounting monitor, and (b) user accounting monitor.

4.1. Farm Overview and Usage History

Figure 2 shows (a) resource availability and overall and group/VO job status at a farm in several tables and (b) evolution of the same information with time, in different time bins. The overview

![Figure 2. Farm overview: The tables show resource availability and job status. The plots show historical information about resource availability and usage. One can select different pre-defined time bins for both the plots and a group from a drop-down list for the job status plot.](image_url)
what to expect from the farm in the next hours and provides rough accounting information for the groups/VOs.

4.2. User Oriented View

Figure 3 shows the status of jobs belonging to individual users. Users are uniquely identified by the Distinguished Name (DN) of the Grid Certificate that was used to submit Grid jobs. This view makes the monitor useful for end-users. Grid job submission files are parsed in order to map a local job to a remote user.

4.3. Group Accounting

Figure 4 shows the accounting information for each group that finished a minimum number of jobs at the farm, using a number of tabs for different time periods. In addition to the table that presents the various numbers, important quantities are also presented as pie- and bar-charts for the majors groups for each period.

4.4. User Accounting

Figure 5 shows the accounting information in tabular form for each individual user who finished a minimum number of jobs at the farm. A number of tabbed panels are used to present information for different time intervals. The table supports search, sort, and pagination capabilities. A Grid user belonging to several groups/VOs is accounted for correctly. Note that only the Grid jobs are considered in user accounting.

5. Site Monitors

A number of sites running different version of the LSF batch systems have deployed the monitoring system, as shown in table 1:

**Figure 3.** User oriented view: The Distinguished Name (DN) of the Grid Certificate identifies the Grid users uniquely.
Figure 4. Group accounting monitor: The table shows, for each time period, accounting information for each group while the charts compare the important parameters for the major groups in that period.

Figure 5. User accounting monitor: The Distinguished Name (DN) of the Grid Certificate identifies the Grid users uniquely.

6. Deployment

The monitoring system is in active development. New features are added regularly while the existing ones get refined. The job overview and group accounting software is distributed as a tarball that can be installed easily and configured extensively. We briefly describe below the
Table 1. The monitor has been deployed at a few sites. The monitoring instances are not yet synchronised to the latest development of the web interface.

deployment procedure:
  > wget http://sarkar.web.cern.ch/sarkar/dist/lsfmon_v1.8.1.tgz
  > tar xzvf lsfmon_v1.8.1.tgz -C /opt
  > ln -s lsfmon_v1.8.1 lsfmon
  > cd /opt/lsfmon/install
A configuration file app.cfg included in the distribution has to be suitably modified according to the site setup. The following script
  > ./setup.sh
will generate the necessary files (scripts, cron jobs, application level configuration file etc.). The application level configuration file has a lot of options that might be changed to affect both the content and look-and-feel of the monitor. Further details can be found in reference [5]. The user accounting software will soon be available as a separate package with documentation.

7. Conclusion
The present tool compliments existing global monitors effectively and helps both site administrators and users to spot and eventually fix problems early and reliably. Accounting information for recently completed jobs can only be obtained with such a local tool. The overview monitor inspired development of similar tools for Condor and PBS batch systems [4] and more importantly an XML based uniform local job monitoring framework in the CMS experiment [1].

The monitor now has basic support for parallel jobs. It has been noted that more work is needed for better support of local job accounting and to adapt the monitor to changes in complex farm organisation. As we mentioned earlier, the user accounting monitor is an independent application at present. We plan to merge the user accounting monitor with the rest in near future.

Acknowledgment
The work has been partially funded under contract 2008MHENNA_002 of Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale (Italy). We acknowledge contribution from Massimo Biasotto on LSF 7 accounting file parser and Claudio Strizzolo on HTML validation and font optimisation respectively. We are thankful to Tommaso Boccali, Alberto Ciampa, Enrico Mazzoni, and Satish Patel for helpful suggestions and support.

References
[1] C. Grandi et. al., CMS Distributed Computing Integration in the LHC sustained operations era, PO-WED-061 (Contribution to CHEP 2010, Taipei)
[2] http://www.sencha.com
[3] http://jquery.com
[4] https://twiki.cern.ch/twiki/bin/viewauth/CMS/AnalysisOpsT2Monitoring
[5] http://cern.ch/sarkar/doc/lsfmon.html