STFC Daresbury Laboratory
Grid-SAFE Case Study

"Grid-SAFE can provide us with a mechanism for tracking and evaluating user activity of a range of HPC machines as well as providing users a mechanism for querying their own usage."

Tavia Stone,
STFC Daresbury Laboratory

Introduction

STFC Daresbury Laboratory has a long history of operating and providing HPC services for UK academics, including the previous national HPC service (HPCx). They currently have a number of different HPC systems, from a cluster provided for the NW-GRID (consisting of 128 cores) to a 1024 node BlueGene/L to a GPGPU cluster and others. The machines are provided for a range of users from many different communities and consortium.

Challenges

With a range of HPC systems there are a number of different usage logs, in a variety of formats. These are generally produced by the specific batch systems of the machines and stored locally. The systems initially selected for reporting/accounting upon use the SGE and LSF/Slurm batch systems.

Currently, reports are hand generated bi-monthly for each of the systems operated at STFC Daresbury Laboratory. This is a time consuming process which is not well suited to changes in the reporting metrics or format (i.e. it is difficult to change what is reported on). This process also requires manual intervention to report usage across all HPC systems, and has no mechanism for investigating specific project, group, or user usage.

To reduce the effort associated with the current reporting procedures, reduce the scope for errors, and provide more functionality in reporting upon the usage of services they would like a solution which enables dynamic reporting across a range of metrics (including reporting for individual machines, users, or projects) for any requested timescale.

Solutions

The Grid-SAFE software, developed at EPCC with JISC funding, handles accounting, reporting, and usage monitoring for advanced computing facilities. It is a software framework, which comprise a number of modularised components which can be assembled to provide end-to-end HPC or Grid service management.

Grid-SAFE provides a SGE parser to process and store the usage data generated by one of the HPC systems, and a tool for obtaining data from the LSF/Slurm log format. Usage data is processed and categorised by user. It is also categorised by the machine it came from, ensuring usage can be reported on a per machine basis if required, or over all machines if appropriate. Group data is extracted from the usage logs using the group id field recorded in those logs.

Processing and storing the data through Grid-SAFE enables resource managers to generate reports over a variety of timescales on demand meaning reporting is no longer restricted to bi-monthly data which has been manually generated. Different timescales can be reported over, and different metrics assessed (such as usage of different groups), simply by selecting the required parameters and re-running the report generation.

The Grid-SAFE solution has been locally installed and configured by system administrators at Daresbury Laboratory and is being evaluated along side their existing systems to ensure it can produce the reports required in a reproducible and stable manner.

Technical Achievements

- **Multiple resources or machines** are supported; Reporting is enables from a number of different HPC machines using different batch systems and log formats. This includes both the SGE and LSF/Slurm batch system for these particular HPC machines.
- **Configurable policies** are available; this provides a mechanism for resource managers to customise how data is processed and stored in the Grid-SAFE system. In this instance, policies are used to generate groups for reporting across based on the usage data.
- **Multiple roles** system allows one or more roles to be assigned to the user; User and Administrator are used here to ensure that users can only access their data and Administrators can access all data. Further roles can be defined if required, for example to restrict access to particular groups for project managers or PIs.
- **Dynamic reports** can be defined by the service providers; Reports can be changed or added by the Grid-SAFE managers. This allows a high degree of tailoring of the reporting functionality for different sites or situations. We tailored reports so that specific reports per user, per group, and per machine, were created. We also provide the functionality to add local text and logos to reports to enable them to be used internally or externally with no changes required.
- **Restricted access to reports** prohibits unauthorised users for viewing a report; for this instance the specific reports which generate group, machine, and user usage are only viewable by the Administrator.
- **Restricted data within a report** prohibits unauthorised users for generating reports own data which they are not authorised to view; in this instance an individual users report covers only their usage.

Further information

Grid-SAFE: http://www.epcc.ed.ac.uk/projects/grid-safe
STFC Daresbury Laboratory: http://www.stfc.ac.uk/About/Struc/Locs/DL/facs.aspx