Bristol University
Grid-SAFE Case Study

"We have a number of HPC systems which produce usage data. We would like to be able to analyse both user and project usage across our machines and also enable users to track such usage. The Grid-SAFE software can provide a solution which satisfies these requirements."

Callum Wright
ACRC, University of Bristol

Introduction

The Advanced Computing Research Centre (ACRC) at Bristol University has two main HPC systems, BlueCrystal Phase 1 and BlueCrystal Phase 2, where Phase 1 (consisting of 384 cores) is intended for serial and small scale parallel jobs, and Phase 2 (consisting of 3328 cores) is used for larger parallel jobs.

Challenges

The two HPC systems use the PBS batch system for the submission of user jobs. PBS records usage "events" in a log file. Currently, the managers/operators of the HPC systems manually produce a monthly report of the usage of the systems by processing the relevant PBS logs with scripts. This report simply states the number of jobs submitted, the average and maximum wait time of the jobs, the total number of processor hours used and a break down of usage over a number of basic groups. They use six groups which are specified at the time the job is submitted dependent on what institute the user belongs to.

In order to more efficiently manage and report upon the usage of the HPC systems they would like to be able to report on the usage of individual users and of specific projects. Projects are defined in a separate file which maps individual usernames to projects. The user jobs do not record the project associated with the job, simply the group of the user. There are also other details held in files, such as a description of the individual project, which are available for processing.

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 comprises a number of modularised components which can be assembled to provide end-to-end HPC or Grid service management.

Grid-SAFE provides a PBS parser to process and store the usage data generated by the two HPC systems at Bristol. Usage data is parsed, the user for each job extracted, and the usage data stored with a link to the user. This ensures that reports can be generated analysing individual usage. We also configured the Grid-SAFE service to allow data from multiple machines to be uploaded and categorised

by machine, ensuring usage can be reported on a per machine basis if required, or over all machines if appropriate. Furthermore, we provided more customised text-based parsers for this service as they require the ability to upload project information (i.e. the project that each user belongs to) as well as accounting/usage information.

Processing and storing the data through Grid-SAFE enables resource managers to generate reports over a variety of timescales on demand, so reporting is no longer restricted to the manually produced monthly report. Different timescales can be reported over, and different metrics assessed (such as usage by user) when required with no additional effort from the resource managers/administrators. There is also the potential to provide this service for resources users (to enable them to analyse their own usage) a little initial effort (i.e. configuring usernames and passwords for users using a range of available authentication mechanisms).

The system managers are currently evaluating the deployed Grid-SAFE with the intention of using it as a production service within 2010.

Technical Achievements

- **Multiple resources or machines** are supported; Grid-SAFE is designed to allow data from a variety of HPC machines to be collected and reported upon in a single place. This enables managers to generate reporting information (and undertake accounting operations) across all their HPC facilities using the same service and also provides users with a single reporting service for all their jobs. Bristol collates data from 2 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 link users to projects.
- **Multiple roles** system allows one or more role(s) to be assigned to a user; User and Administrator are used here to ensure that users can only access their data and Administrators can access all data.
- **Auxiliary information** can be stored for users or other organisational groups over and above what is available from the usage records; specifically, project details (including the description of a project and the link between a specific user and a project) are stored.
- **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 of different sites or situations. We tailored reports so that specific reports per user, per project, and per machine were created as required. We also provide the functionality to add local text and logos to reports if required.
- **Restricted access and data in reports** prohibits unauthorised users from viewing a report; for Bristol the specific reports which generate project, group, machine, and selectable user usage are only viewable by the Administrator.

Further information

Grid-SAFE: http://www.epcc.ed.ac.uk/projects/grid-safe
Bristol ACRC: http://www.acrc.bris.ac.uk/acrc