

## ECDF Case Study

*“Before GridSAFE was available, hours were spent cutting and pasting information extracted from Sun Grid Engine log files. GridSAFE extracts the same information quickly and reliably, so reports can be produced in minutes. An important feature is that users and group managers can extract their own information on demand.”*

**Jean Ritchie, ECDF Service Director**



### **Introduction**

The Edinburgh Compute and Data Facilities (ECDF) came into existence with funding from SRIF3 in October 2007. It offers a high-performance cluster, called Eddie, to researchers from, or associated with, the University of Edinburgh. Eddie, a 1456 processors Intel Xeon machine, was launched as a service on August 1st 2008. As of May 2010 there are 678 registered users, and the machine averages 1.9 million jobs per month. The service is now deemed so successful that the hardware upgrade due in the summer, 2010 will be Full Economic Cost (fEC) funded.

### **Challenges**

The charging model for the Eddie service is complex. The batch system, Sun Grid Engine (SGE), supports a fair-share resource allocation policy that shares resources equally, and automatically, among users. Any user within an University of Edinburgh school can use this ‘baseline’ service. ECDF also offers a ‘guaranteed’ service to support research groups, or projects, who wish to reserve time. The charging rates for ‘baseline’ and ‘guaranteed’ are calculated such that all the running costs, including the electricity, to run the service are recovered.

The process of calculating the reports for a schools’ baseline usage involves running hand rolled scripts which parse the batch system logs - the results of which are pasted into excel spreadsheets and then undergo a series of error-prone transformations to clean the data. Finally a set of complex accounting formula is applied to the data to produce charging reports. A similar process has to be undertaken for a projects’ ‘guaranteed’ usage reports. As these reports are time consuming to evaluate they are not generated often which meant the school and project managers had no visibility of their usage in between reports.

## ***Solutions***

ECDF adopted Grid-SAFE about a year after the launch of the Eddie. Grid-SAFE developed at EPCC with JISC funding, handles accounting, reporting and usage monitoring for advanced computing facilities. It is a software framework, which is comprised of a number of modularised components which can be assembled to provide end-to-end HPC or Grid service management. Its dynamic reporting component allows not only the school and project managers using ECDF up-to-date visibility of their usage, but it also allows the individual users to view their job details. It allows the ECDF service providers to generate required service reports as and when required. The SGE parser used for Eddie is one of many batch parser supported by Grid-SAFE along with PBS, Load Leveller and OGF Usage Record parsers. Grid-SAFE also offers a highly configurable organisation and authorisation components which can support the school, project groups, users and roles required by ECDF. Grid -SAFE supports daily aggregate records which allows ECDF service reports to be generated on-the-fly, despite the fact they may be processing data form over 24 million records,

## ***Technical Achievements***

- **External authentication services** are supported; for ECDF this was the University of Edinburgh Cosign-based EASE authentication service.
- **Configurable organisation structure** allows the relationships between users and groups of users to be defined using ‘classifiers’; for ECDF this covered School, Project and Project Member relationships along with the School and User relationships.
- **Multiple roles** system allows one or more roles to be assigned to the uses; for ECDF, this included the ‘Support Staff’, ‘School Administrator’ and ‘Project PI’ roles.
- **Auxiliary information** can be stored for users or other organisational groups over and above what is available from the usage records; for ECDF, the Unix password file is uploaded to extract the users full names, also the admin forms can be used to add and remove Schools and Projects.
- **Dynamic reports** could be defined by the service providers; for ECDF, the service team defined a variety of reports for different users from the Project Members to the Service Director.
- **Restricted access to reports** prohibits unauthorised users for viewing a report; For ECDF, only the ‘PI’ for a Project could access that Project’s report.
- **Restricted data within a report** prohibits unauthorised users for generating reports own data which they are not authorised to view; for ECDF, a user report cover only their usage.
- **Scalable reports** are possible by the use of daily aggregate records; for ECDF at present this reduces the present number of records from 55 million to 35 thousand.

## ***Further information***

ECDF: <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/research-computing/ecdf/>

GridSAFE : <http://www.epcc.ed.ac.uk/projects/grid-safe>