



Contribution ID: 42

Type: **Poster**

Data awareness in gqsub

Monday 12 April 2010 17:33 (3 minutes)

gqsub is a user (command line) interface for submitting and monitoring Grid jobs that conforms to the IEEE standard for qsub (and friends).

Recent development has focused around data and data awareness; both in terms of data local to the submission machine, and data elsewhere in a Storage Element.

Detailed analysis

The gLite and Arc computing infrastructures are both distinguished by their data awareness - that is, the location of data matters to job scheduling.

Firstly, specifying the request. For staging in of data, this is not too tricky; by allowing the files to be specified as a url, it's straightforward to handle this. The staging out of data, to push it to a Storage Element at the end of the job, is more complex, as it requires two distinct bits of information. At the time of writing, there are a few alternatives in development, pending user trials before final selection.

The second part is to handle the requests. In the cases where the computing infrastructure supports the desired behaviour, it's a matter of compiling the request into the appropriate form. Where it's not, then the requisite code needs to be supplied, typically via the use of a wrapper script.

One interesting case of this is local use. A design goal of gqsub was to have job scripts that could be submitted either locally or on the Grid without modification. Data access changes this, as most local systems are not aware of storage elements. At the time of writing, no clear solution is available.

Conclusions and Future Work

This work takes gqsub from being able to treat the Grid as a cluster extension to being able to leverage the full power of the the Grid model. As noted, there are a couple of outstanding points to be resolved, which is fully expected by February. Once complete, this will enable the Grid to come to any cluster.

As a further work, the opposite is planned - taking a cluster to the Grid. By capturing a sub directory tree, from the users machine, and staging that remotely, the setup of complex environments can be simplified.

Impact

The data awareness increases the scope of suitable application scales, from using the Grid as a cluster replacement to being able to manage tera scale applications. It has lost none of the simple user interface in the progress, meaning it is able to scale with the users needs, from the simplest use case, out to multi-institutional data-analysis projects.

gqsub has always stood out from other attempts to provide a 'simple' way to use the Grid, because it offers an interface that is *already* known to many e-Scientists, rather than expecting them to learn a new one. For users that have previously used a cluster of some sort, the time spent on learning gqsub is trivial - less, and often significantly less, time than it takes to sort out the X.509 certificates.

It is designed to facilitate having a cluster head node also serve as a submission system for Grid jobs, so that it is natural and straightforward for a user to switch between the two.

Keywords

gLite, data management, HCI

URL for further information

<http://www.scotgrid.ac.uk/gqsub/>

Justification for delivering demo and/or technical requirements (for demos)

(see comments)

A demo would need just an internet connection, and be run from any connected system (i.e. my laptop). An external monitor would be handy, to increase visibility.

Authors: Prof. DOYLE, Anthony (University of Glasgow); Dr PURDIE, Stuart (University of Glasgow)

Presenter: Dr PURDIE, Stuart (University of Glasgow)

Session Classification: Poster session

Track Classification: End-user environments, scientific gateways and portal technologies