Contribution ID: **67**   Type: **demo presentation**

Managing Mass Storage with Grid Jobs

Wednesday 9 May 2007 19:30 (20 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

GridPP is the UK Grid for particle physics, and covers the UK
participation in the
Large Hadron Collider Computing Grid (LCG) and also supports
other scientific
communities, including the EGEE biomedical community. Users are
located not just in
the UK, but across the world: each community has a need to
transfer data to or from
collaborator sites in the UK, and to analyse the data with grid
jobs and to store the
output. The work presented here is also relevant to the UK
National Grid Service, NGS

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

We aim to demonstrate interoperability between different storage
systems by showing
data being accessed or transferred by Grid jobs on behalf of the
user, or by higher
level transfer services that schedule and manage data replication
between the sites.
Common use cases include regular transfers of data from one site
to another,
replication of "popular" files to improve access time, and
fetching data to local
disk before the job runs for efficient access. This demo thus
does not just depend
on the SEs, but also on the job submission system and high level
transfer services.
We also depend on the networks for the transfer, and aim to show
different transfer
rates achieved, depending on a site's network bandwidth and how
busy the SE is. We
aim to demonstrate this with a live demo, showing the details of
the file transfers
as they happen on the Grid.

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)

Providing a standard interface to storage is essential but it must also have
sufficient functionality to support the different use cases. Traditionally, getting
SEs to interoperate has taken lots of effort, particularly with specifications being
loose or leaving things optional. A dependable interoperating core is essential, and
this requires lots of testing. Clients needing advanced functionality must be
prepared to cope with systems that don't provide it.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

One thing all these communities have in common is the need to store and manage
scientific data. Standardised interfaces are essential to ensure that the data can
be accessed regardless of the type of storage system: from a single disk running on a
PC, over large dedicated disk servers, to the largest tape robots. From the Grid's
view, a storage system is a Storage Element, or SE. Each SE provides control
interfaces, data transfer interfaces, and information systems, which in turn are used
by client applications or higher level Grid services to manage data consistently. The
storage group in GridPP supports different storage solutions to enable UK Grid
communities to participate in global Grid collaborations with SEs that interoperate
with other SEs across the world. Variations in sites' infrastructure are essentially
hidden from the grid clients, and SEs fit together to form a global data grid. The
NGS is evaluating these SEs as well.

Author: Dr JENSEN, Jens (CLRC-RAL)

Co-authors: Mr GARRETT, Barney (Edinburgh); Dr KANT, David (CLRC-RAL); Dr MOONT, Gidon (Imperial College); Dr COWAN, Greig (Edinburgh); Dr HODGES, Matthew (CLRC-RAL)

Presenters: Dr MOONT, Gidon (Imperial College); Dr JENSEN, Jens (CLRC-RAL)

Session Classification: Poster and Demo Session