



Training Outreach and Education

<http://www.nesc.ac.uk/training>



<http://www.ngs.ac.uk>

Data services on the NGS

PPARC

<http://www.pparc.ac.uk/>

eGEE

<http://www.eu-egee.org/>



Policy for re-use

- This presentation can be re-used for academic purposes.
 - However if you do so then please let training-support@nesc.ac.uk know. We need to gather statistics of re-use: no. of events, number of people trained.
- Thank you!!



Data services

- The Storage Resource Broker
- OGSA-DAI: data access and integration
- GridFTP: a protocol for large file transfer
- And also: ORACLE – available for data storage

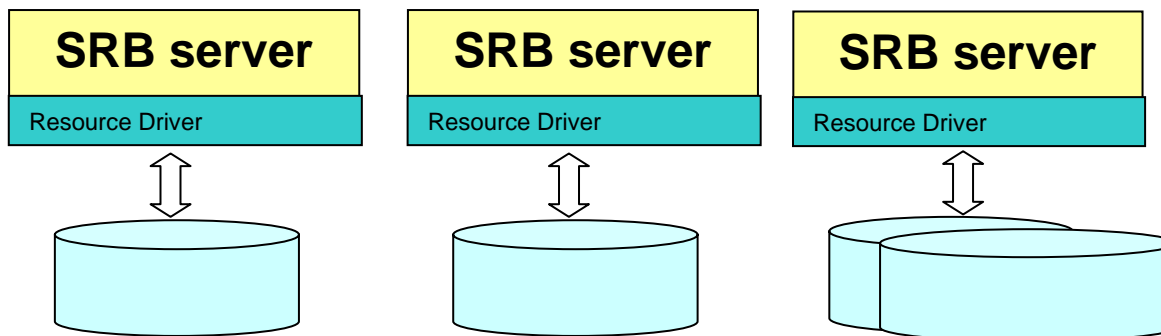
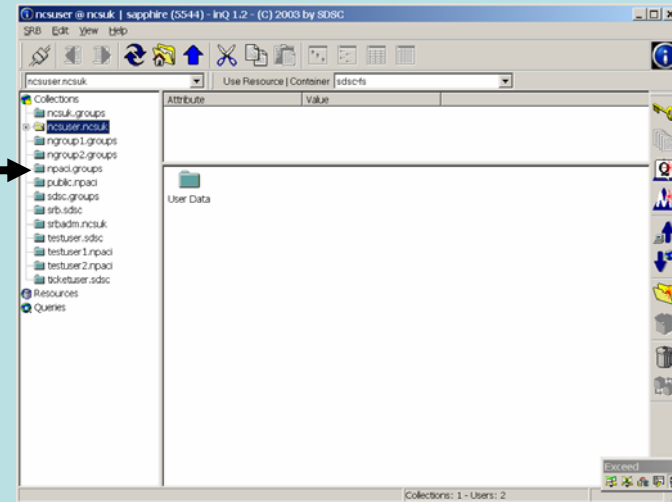
- **Simple data files on grid-specific storage**
- **Middleware supporting**
 - **Replica files**
 - to be close to where you want computation
 - For resilience
 - **Logical filenames**
 - **Catalogue:** maps logical name to physical storage device/file
 - **Virtual filesystems,** POSIX-like I/O
 - Services provided: storage, transfer, catalogue that maps logical filenames to replicas.
- **Solutions include**
 - **gLite data service**
 - **Globus: Data Replication Service**
 - **Storage Resource Broker**

- **Other data! e.g.**
 - Structured data: RDBMS, XML databases,...
 - Files on project's filesystems
 - Data that may already have other user communities not using a Grid
- Require extendable middleware tools to support
 - Computation near to data
 - Controlled exposure of data *without replication*
- Basis for integration and federation
- **OGSA –DAI**
 - In Globus 4
 - Not (yet...) in gLite

Storage Resource Broker

User sees a virtual filesystem:

- Command line (S-Commands)
- MS Windows (InQ) →
- Web based (MySRB).
- Java (JARGON)
- Web Services (MATRIX)



**Filesystems in
different
administrative
domains**



Oracle and the NGS

- The NGS core nodes, from the outset, have been partitioned into compute and data clusters.
- As the NGS matures the requirement for data hosting will grow
- Oracle database: for both users and services offered by the NGS.

**The NGS Oracle service
is ready to host data
for your project!**

- The RAL and Manchester sites are designated as the data clusters with each site having the ability to dedicate up to eight nodes for use by Oracle.
- Additional application needed after joining the NGS



A view on grids and data

- Most common current use case for grids in the UK:
 - “I want more CPU time and I want it now”
 - Answer should often be: “use a Condor pool”
- Occasionally
 - “I want CPU time close to where my data are”
- Research is about
 - Exploring data... not just simulations....
- Data services will drive the uptake of grids??