MANCHESTER 1824

The University of Manchester

A Lightweight Grid Infrastructure using afs

Roger Barlow, Alessandra Forti,
Andrew McNab
Sabah Salih, and Mike Salt

Particle Physics Group
School of Physics and Astronomy
Manchester University

MANCHESTER 1824

The University of Manchester

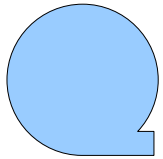
Contents

- Styles of Grid usage
- Our solution: afs and gssklog
- Case studies
- Experience

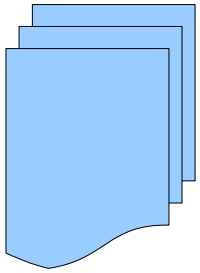
Grid users – ideal vs. real

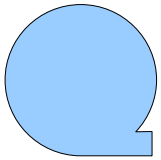
CPU

+Data



+environment





User job: typical needs

To run a program on a remote cluster needs some or all of:

- CPU cycles – compatible architecture and OS
- The data the job is to run on
- The executable(s) and/or scripts
- Control file
- Dynamic libraries (.so, .dll)
- Scripting languages (perl,python)
- Databases
- Calibration files, Look-up tables etc.

  All with the correct version/release number
- Retrieval of various output files results, logs, debug..

Solution 1

Ensure your environment is available at the remote site

- Puts load on remote sysman
- Practical for large user communities, e.g.LHC experiments
- Publish through ldap, or investigated via pilot jobs
- Good solution, but only for very big groups

Solution 2

Put everything in a tarball in the input sandbox

•No load on remote sysman - OK

•'Everything' has to cover everything which might possibly be needed. And you won't know till the job fails.

•'Everything possible' can be enormous. Inefficient and bottleneck

Solution 3

User has tested job in some work directory at their local site

Suppose this local site uses afs as its user file system

cd workdir = cd /afs/mysite.ac.uk/users/myusername/workdir

From within this directory, all files are available

User now moves to production, on a remote cluster.
As afs is global, remote site script can also in principle say
cd /afs/mysite.ac.uk/users/myusername/workdir

And run as before.
In practice...?

Remote afs access

Remote afs access possible – with password

klog -principal myusername -cell mysite.ac.uk -password .....

(Obsolete – use kinit instead)

Sending passwords in jobs is out.

globus provide gssklog

gssklog -principal myusername -cell mysite.ac.uk

Uses proxy certificate sent with the job for (X509) authentification and translates to kerberos, used by afs

Typical testing on local system

```
cd workdir
myprog < control.dat > run1.out
```

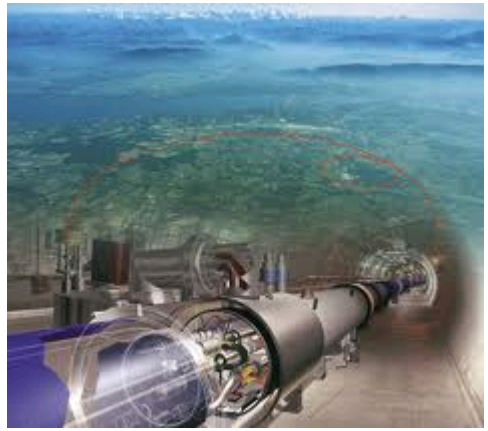
Typical production script sent to remote system

```
gssklog -principal myusername -cell mysite.ac.uk
cd /afs/mysite.ac.uk/users/myusername/workdir
myprog < control.dat > run1.out
```

Example 1: LHC collimation

MERLIN simulation code tracks 100,000 particles for 200 turns, in ~2 hours. Study where particles hit collimators

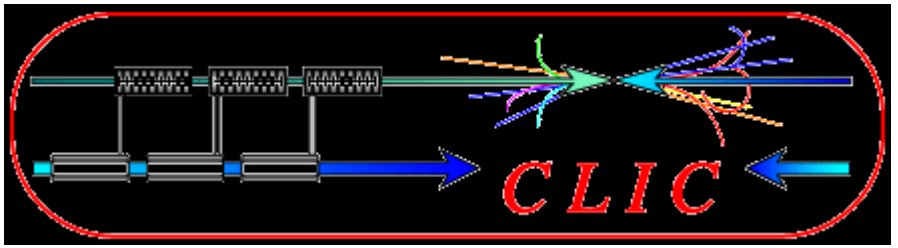
Need ~100 jobs for good statistics. Create subdirectory for the run, and subsubdirectory for each job.
tcl/tk script loops over 100 globus-job-submit calls



```
/usr/bin/gssklog -cell hep.man.ac.uk -principal $name -server afs1.hep.man.ac.uk -port 5750
cd $homedir/$DIR
export ROOTSYS=../../myRoot/root
export LD_LIBRARY_PATH=../../../lib:$ROOTSYS/lib:$LD_LIBRARY_PATH
ln -s ../../LHCB19.tfs LHCB19.tfs
ln -s ../../../Merlin/MerlinExamples/Wakefields/Data Data
../../example411 $iseed >run.out >run.err
```

Example 2: CLIC backgrounds

To analyse backgrounds from backscattered photons at the proposed CLIC accelerator, BDSIM / GEANT4 needs to run for ~7 days, processing 3,000,000 particles.

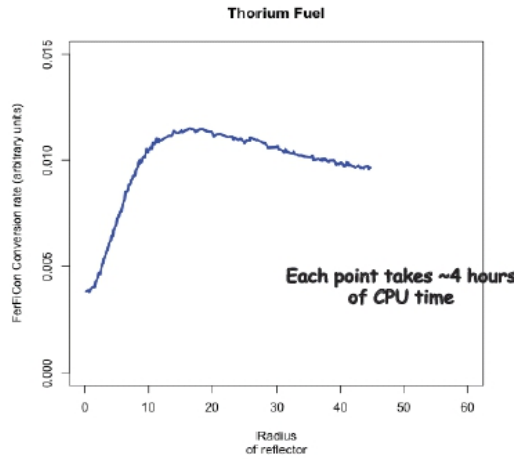



```
myproxy-init -n -d
voms-proxy-init -voms vo.northgrid.ac.uk
glite-wms-job-delegate-proxy -d mick
glite-wms-job-submit -d mick -o jobid –config autowmsconf
        -r ce02.tier2.hep.manchester.ac.uk:2119/jobmanager-lcgpbs-long bdsim.jdl
```

```
/usr/bin/gssklog -cell hep.man.ac.uk -principal mike -server afs1.hep.man.ac.uk -port 5750
export BDSIM_BASE_DIR=/afs/hep.man.ac.uk/u/mick/programs/BDSIM
....
cp /afs/hep.man.ac.uk/u/mick/tointdump/coll112_geomlist .
...
bdsim –batch –file=run_cohmin_2708.gmad –output=root –outfile=output_2708 >log_2708.txt
/usr/bin/gssklog -cell hep.man.ac.uk -principal mick -server afs1.hep.man.ac.uk -port 5750
cp log2708.txt /afs/hep.man.ac.uk/u/mike/tointdump
```

Example 3: Thorium Conversion

The MCNP program was used to simulate the coversion of ^{232}Th to ^{233}U by spallation neutrons hitting a Thorium fuel rod surrounded by a lead reflector. The radius of this was optimised by scanning 50 cm in 1 mm steps, each point running 100000 neutron cascades.



Subdirectory structure as before.
Each result is a single number written to a text file
One script combines and plots, another checks and resubmits

```
/usr/bin/gssklog -cell hep.man.ac.uk -principal $name -server afs1.hep.man.ac.uk -port 5750
cd $homedir/$DIR
RAD =`echo 0.1*$iseed | bc`
export RAD
cat >temp ..EOF
..
1 CZ 0.5
2 CZ $RAD
..
EOF
export DATAPATH=/afs/hep.man.ac.uk/g/accelerators/sw/MCNP/MCNP_DATA
/afs/hep/man.ac.uk/g/accelerators/sw/MCNP/MCNP/bin/Linux/mcnp5_i386 n=temp
```

What has to be done

1) Local site running afs server
In general use for user convenience: they do not have to keep track of physical disks

2) Remote site running afs client
This is widespread – strong user demand as they want jobs to read files from /afs/cern.ch etc

3) Local site running gssklog server
This is work for the local sysman, but not onerous. One simple daemon, runs smoothly.

4) Entry in local site mapfile to map certificate DN to afs username
This need only be done once per user

Common objection

"But afs file i/o is slow..."

afs can be slow for edit-type operations. Not relevant here.

afs is not optimal for high throughput read/write. But it is not being used for that in this example. Just one-off reads and occasional writes.

(If large results files are to be written then they can be written to /tmp and copied at the end.)

Conclusions

X509-enabled access to afs using gssklog provides a tool for grid use for a community of users with a variety of requirements

Associated management overheads are light

Not a complete solution, but a tool users can use in writing their own submission scripts

It is being used to foster and encourage use of the computing centres accessible through the grid