GridPP Dirac

Daniela Bauer & Simon Fayer
Imperial College London

DIRAC User Workshop 2016
The GridPP Project

- A collaboration of 19 UK institutes providing Grid and Cloud based computing services to particle physics and other experiments.

- Hosts 43k job slots and 33 PB of storage.
The GridPP DIRAC server
## Supported VOs

<table>
<thead>
<tr>
<th>VOs</th>
<th>VOs</th>
</tr>
</thead>
<tbody>
<tr>
<td>cernatschool.org</td>
<td>vo.landslides.mossaic.org</td>
</tr>
<tr>
<td>comet.j-parc.jp</td>
<td>vo.londongrid.ac.uk</td>
</tr>
<tr>
<td>gridpp</td>
<td>vo.northgrid.ac.uk</td>
</tr>
<tr>
<td>lsst</td>
<td>vo.scotgrid.ac.uk</td>
</tr>
<tr>
<td>lz</td>
<td>vo.southgrid.ac.uk</td>
</tr>
<tr>
<td>na62.vo.gridpp.ac.uk</td>
<td>(skatelescope.eu – soon)</td>
</tr>
<tr>
<td>pheno</td>
<td></td>
</tr>
<tr>
<td>snoplus.snolab.ca</td>
<td></td>
</tr>
<tr>
<td>t2k.org</td>
<td></td>
</tr>
</tbody>
</table>
VOs actually using DIRAC

upcoming:
cernatschool.org
t2k.org
skatelescope.eu
Usage by site

Running jobs by Site

12 Weeks from Week 00 of 2016 to Week 13 of 2016

Max: 3,906, Min: 0.00, Average: 550, Current: 107

We expect the use of cloud and vac[*] based sites to increase

[*]www.gridpp.ac.uk/vac/
- COMET: Coherent $\mu$ to $e$ transition looking for non-SM decays $\mu + \text{Al} \rightarrow e + \text{Al}$ based in Japan
- Currently uses grid for GEANT4 based detector simulation
- Uses experiment specific python scripts to interact with DIRAC API
- Extensive use of the dirac file catalogue
gridpp

- Umbrella VO for UK researchers that do not belong to any other VO
  - e.g. GHOST: Geant Human Oncology Simulation Tool:
    - http://www.comprt.org/research/ghost-project
    - Using dirac tools directly, running Geant4 based simulation
  - Systematic infrastructure testing (similar to the UK nagios tests, including network tests)
  - Using dirac API, extensive file transfers

pheno

- Phenomenology group based at Durham/UK:
  - See https://inspirehep.net/record/1382345 for current work
  - Uses DIRAC via ganga to run home grown Monte-Carlo.
Neutrino experiment based in Canada:
https://www.snolab.ca/science/experiments/snoplus
- Uses ganga as a frontend
- Monte Carlo production (GEANT4 + experiment software)
- Limited user analysis
- Uses LFC
- Introduced the first non-British sites in GridPP dirac
  - Some of these use HTCondor

Iz (Lux Zeplin)

Dark matter experiment based in the USA:
- http://lz.lbl.gov/
- Uses experiment specific python scripts with DIRAC API
- GEANT4 based detector simulations
- output root files stored on Imperial SE used dirac file catalogue
**lsst (Large Synoptic Survey Telescope)**

- Currently being built in Chile: [http://www.lsst.org/](http://www.lsst.org/)
- Current UK contribution: Shape classification of galaxies on data taken by a predecessor (Dark Energy Survey)
- 100 million galaxies, data divided into 30000 files.
- Experiment specific software.
- Classification of a galaxy is an independent measurement, uses 10-20 s of CPU
- All data for a specific galaxy is contained in the same file
- This looks a lot like particle physics.

- No previous experience in grid computing.
- Settled on ganga and dirac file catalogue for job submission and data access.
- 40 days FTE work required to setup and successfully complete workflow.
Feedback - Users

- All of our users are familiar with the concept of batch systems.
- Most (but not all) of our users are familiar with the concept of 'the grid'.
- We have a UK specific dirac support list → low entrance barrier to questions and feedback
- Following pages: Results from a pre-Workshop survey
Feedback – Users/Admins

Feature requests:
- Command line tool that lists all sites available to a VO (User)
- Being able to ban/enable a site for a particular VO only (Admin)
- Support for Condor Submission (User)
- Support for Condor submission without hacks (Admin)
- Retain failure reason if job only succeeds after retrying (User/Admin)
- This helps to find systematic problems at sites (User/Admin/SiteAdmin)
- Ability to use setInputData at VAC/Cloud sites, which have no LocalSE (User)
- Make this a configuration option (Admin) – (or maybe it is one and I don't know it ?)
Better error handling:

- **Exhibit A:**
  User complains that his jobs fails with “Error during execution”, but no output files are available

Pilot log notes:
Maximum output buffer length reached
EXECUTION_RESULT[CPU] after Execution of spObject.systemCall 0.82 0.93 0.19 0.59 60.69
EXECUTION_RESULT[Thread] after Execution of spObject.systemCall
{"Message": "Reached maximum allowed length (10485760 bytes) for called function return value for
'/scratch/condor/dir_20421/[snip]/DIRAC_6rDTGBPilot/470747/runLUXSim_parametric.sh
lz_tpc_pmtresistors_U_an_late_neutron_parametric.mac 9000001' call",
'OK': False, 'Value': (9, '', '')}

- This message should go to the user.
- This could be handled gracefully (truncated output).
Exhibit B (quoting user verbatim):
DEBUG: {'Message': 'Failed to perform getPathPermissions from any catalog', 'OK': False}
DEBUG: dirac.addFile /gridpp/user/n/name.removed/data_100MB /users/nameremoved/gridpp/gridtests/files/data_100MB UKI-LT2-IC-HEP-disk 2016-05-16 08:31:00 UTC Framework
ERROR: FileCatalog._getCatalogs: Failed to get file catalog configuration. Path /Resources/FileCatalogs does not exist or it's not a section

It turned out that one needs to add these lines at the front of your script:

from DIRAC.Core.Base import Script
Script.parseCommandLine()

This is an experienced user who tried to use the documentation provided by:

→ Documentation needs to contain more than trivial cases
→ This way users can be set on the right path from the start
Feedback - Admin

None of these problems are showstoppers, but.....

As an admin I need a reliable way to report issues and an estimate on when (if ever) they will be fixed.

- I tend to report issues on the mailing list first in order to establish whether it's a bug or a configuration issue:
  - this only works if posts get acknowledged within a reasonable time frame
  - some non-trivial (preferably deployed and tested) examples of dirac config files would be helpful to cut down on configuration issues

- So now it's a bug/feature request and a github issue. Yay!
- But there is no scheduled review of issues and I cannot re-open an issue if it is closed:
  https://github.com/DIRACGrid(DIRAC/issues/2063 (August 2014!)
How to make it better?

- Acknowledge all emails to the mailing list
- Monthly review of all outstanding issues (similar to GGUS)
- Estimated time to fix or, if not feasible, “won't fix/can't fix”
- Submitter should be able to re-open issue/don't close without input from submitter
Conclusion

We've been successfully running a multi-VO dirac server in the UK for the past year.

There is room for improvement.

Thank you for listening :-)