

GridPP DIRAC:
Supporting non-LHC VOs on LHC centric
resources

Daniela Bauer & Simon Fayer
Imperial College London

CHEP 2016

Overview

- Introduction to GridPP and DIRAC
- User communities
- Experiment specific approaches within this framework
- What have we learnt ?

The DIRAC Project

- The DIRAC framework for distributed computing was originally developed for LHCb
- In 2014 the DIRAC consortium was founded to make the software available to a wider audience
- There are currently at least 11 DIRAC instances in use worldwide
- Users can access these instances via scripts (“DIRAC tools”) from their desktop, via a python API, using [ganga\[*\]](#) or via a Web interface
- Homepage: diracgrid.org

* <https://ganga.web.cern.ch/ganga/>
More on this topic in Rob Currie's
talk on Oct 13th

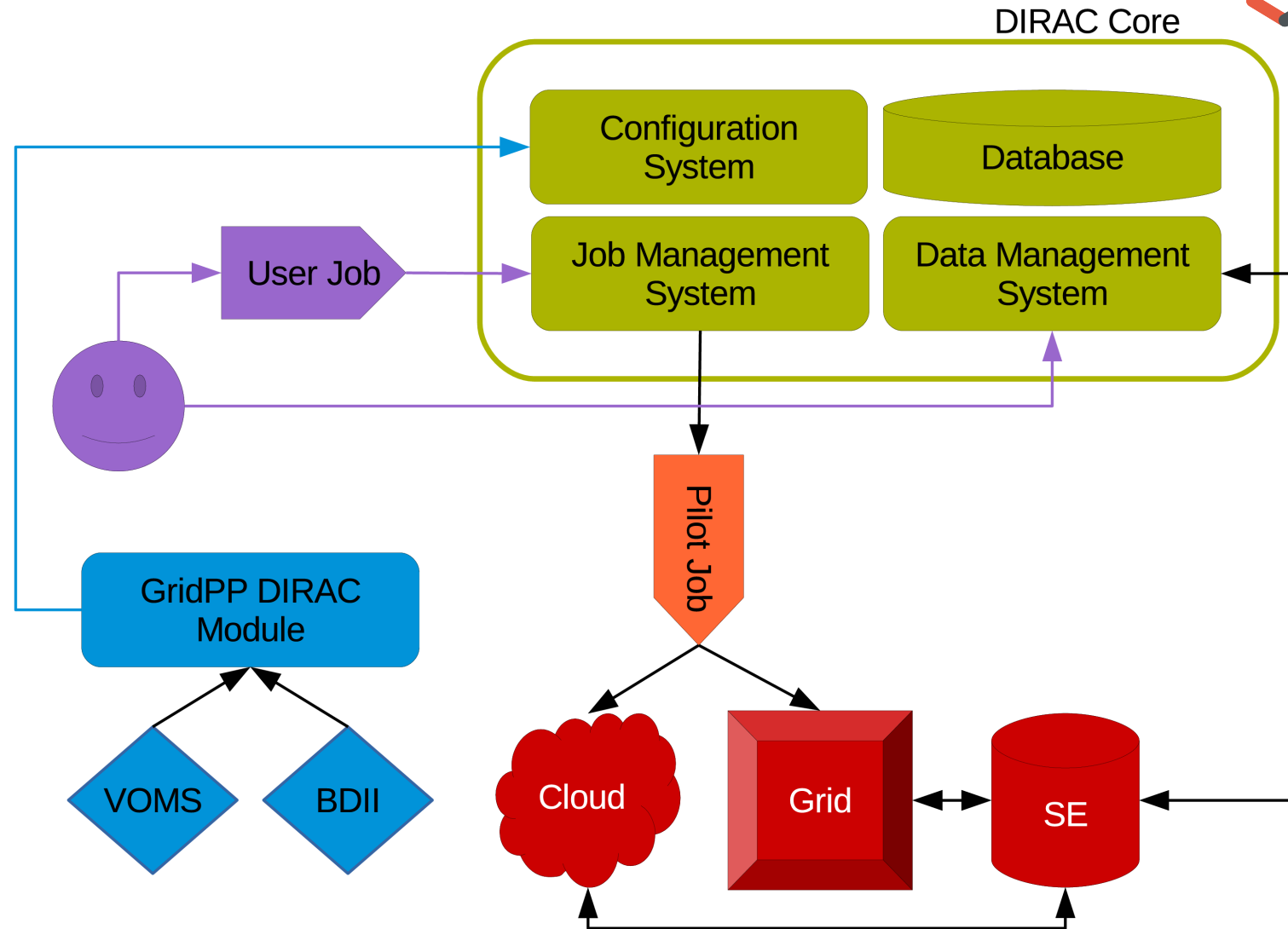


The GridPP Project



- 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
- www.gridpp.ac.uk

GridPP DIRAC



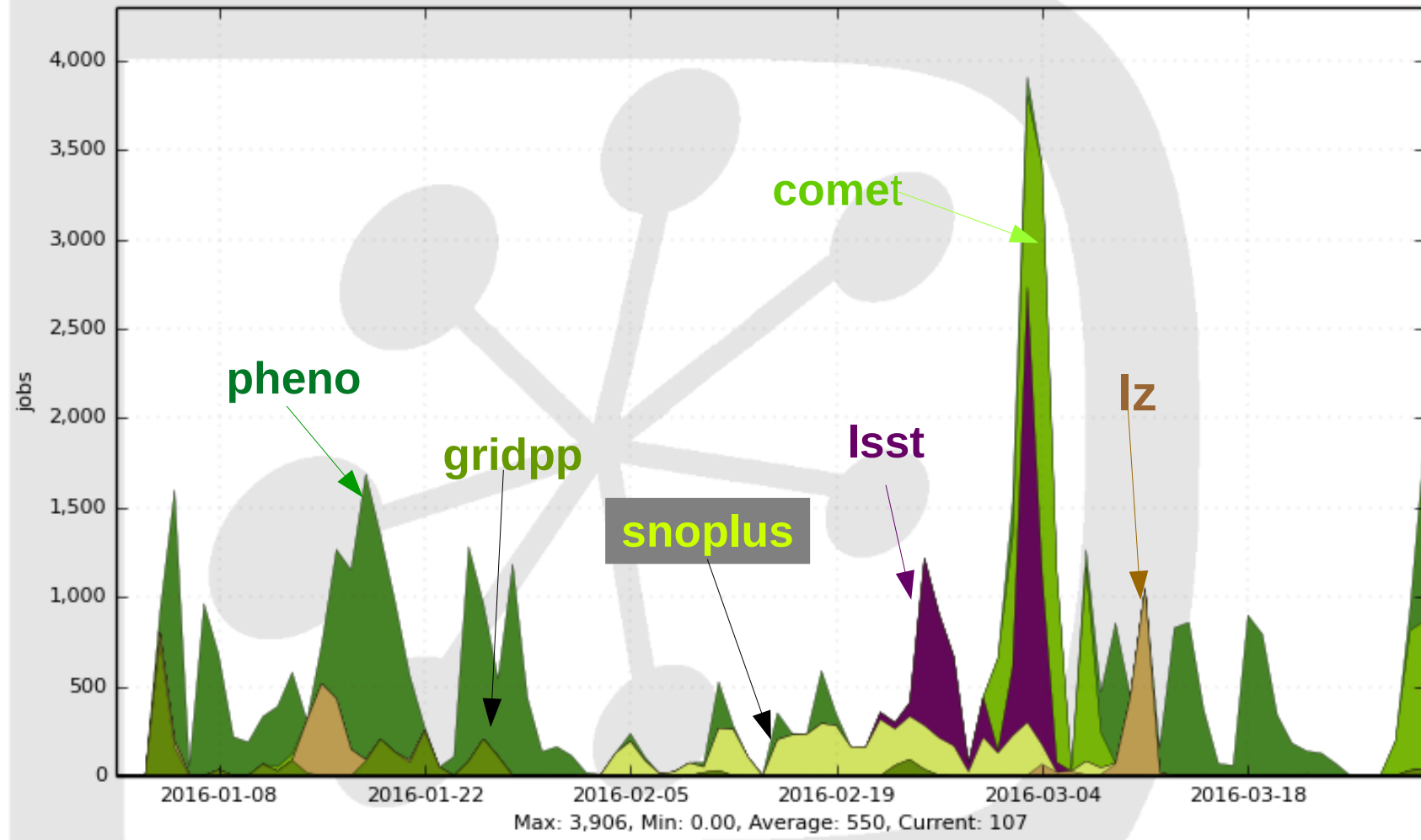
User base

- Small (< 100 users) VOs, usually with UK involvement
- These VOs have no dedicated computing support
- All of our users are familiar with the concept of a batch system
- Most (but not all) user are familiar with the concept of the grid
- Currently 16 supported VOs, 6 of them using GridPP DIRAC actively
- Three of these VOs were new to the grid/distributed computing

VOs using DIRAC

Running jobs by UserGroup

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



pheno_user	47.4%	lz_user	6.4%	vo.northgrid.ac.uk_user	0.0%
comet.j-parc.jp_user	17.7%	gridpp_user	5.3%	na62.vo.gridpp.ac.uk_user	0.0%
lsst_user	12.8%	snoplus.snolab.ca_user	0.0%		
snoplus.snolab.ca_production	10.3%	cernatschool.org_user	0.0%		

LSST (Large Synoptic Survey Telescope)

- Currently being build in Chile: <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.

comet.j-parc.jp

- COMET: Coherent μ to e transition looking for non-SM decays
 $\mu + Al \rightarrow e + Al$ based in Japan
 - <http://comet.kek.jp/Introduction.html>
 - GEANT4 based detector simulation
 - Uses experiment specific Python scripts to interact with DIRAC API
 - Extensive use of the DIRAC file catalogue

LZ (Lux Zeplin)

- Dark matter experiment based in the USA:
 - <http://lz.lbl.gov/>
 - GEANT4 based detector simulations
 - Output root files stored on Imperial SE used DIRAC file catalogue
 - So far: Used experiment specific Python scripts with DIRAC API
 - Now: Custom job submission system, includes database to keep track of production data

LZ user interface

Alexander Richards, Imperial College London

New Request Form

Basic Information

Requester Name:

Sim Lead:

Source type:

Detector components:

Description:

Application Setup

Select Application Type: **Select Application Version:**

Request Macro Information

Select Tag:

Select Macros:

Number of Jobs/Files:

Beam On Events:

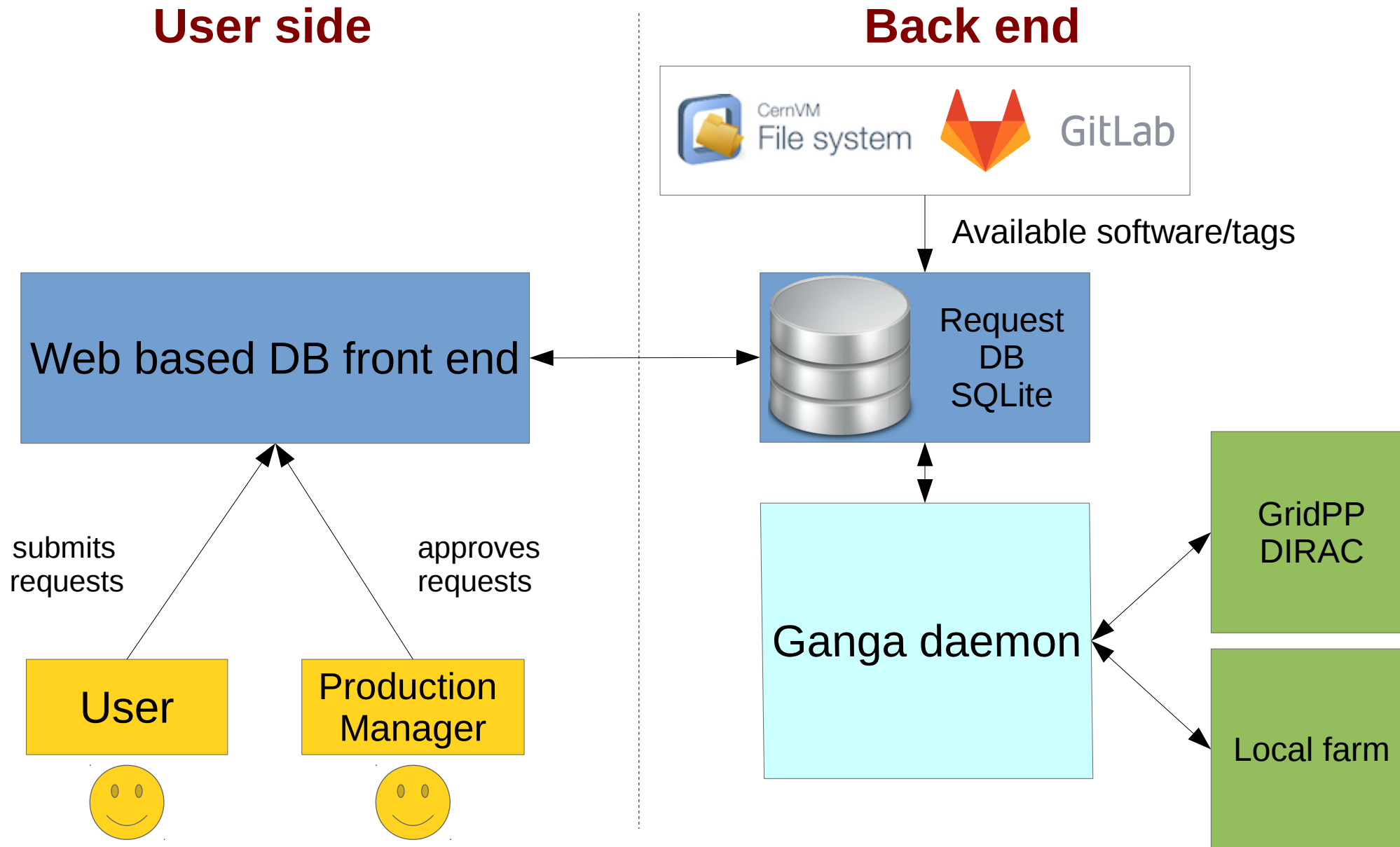
Initial Seed:

Selected Macros:

Allows selection by available software/tags → reduces user error
4 weeks FTE to implement

The LZ production system

Alexander Richards, Imperial College London



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.

snoplus.sanolab.ca

- Neutrino experiment based in Canada:
<https://www.sanolab.ca/science/experiments/snoplus>
 - Established experiment, converting from WMS based submission
 - 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

Observations from User Feedback

- DIRAC was conceived as a single VO framework and not all features are truly multi-VO (e.g. listing available sites)
 - Solving this is primarily a funding issue
- Slow implementation of feature requests
 - Requests are generally both sensible and viable:
e.g. support for HTCondorCEs
 - Manpower issue
- Error messages tend to be unhelpful
 - Not helped by the fact that the underlying grid tools aren't necessarily reporting helpful errors either
 - Lost cause ?

Observations/Conclusions

- To get from a fairly straight forward single test job submission to a production setup in most cases requires several weeks worth of (outside VO) expert effort or a computing affine person in a VO
 - This is reasonable. It's a complicated system doing non-trivial work and we should treat it as such.
 - Focus on making the expert effort available/accessible.
 - Expectation management.
 - Accept that solutions for individual VOs might not be portable.
 - But: Try and steer VOs early to one or two proven approaches, even if it's badly documented.

Conclusion

We've been successfully running a multi-VO DIRAC server in the UK for the 18 months.

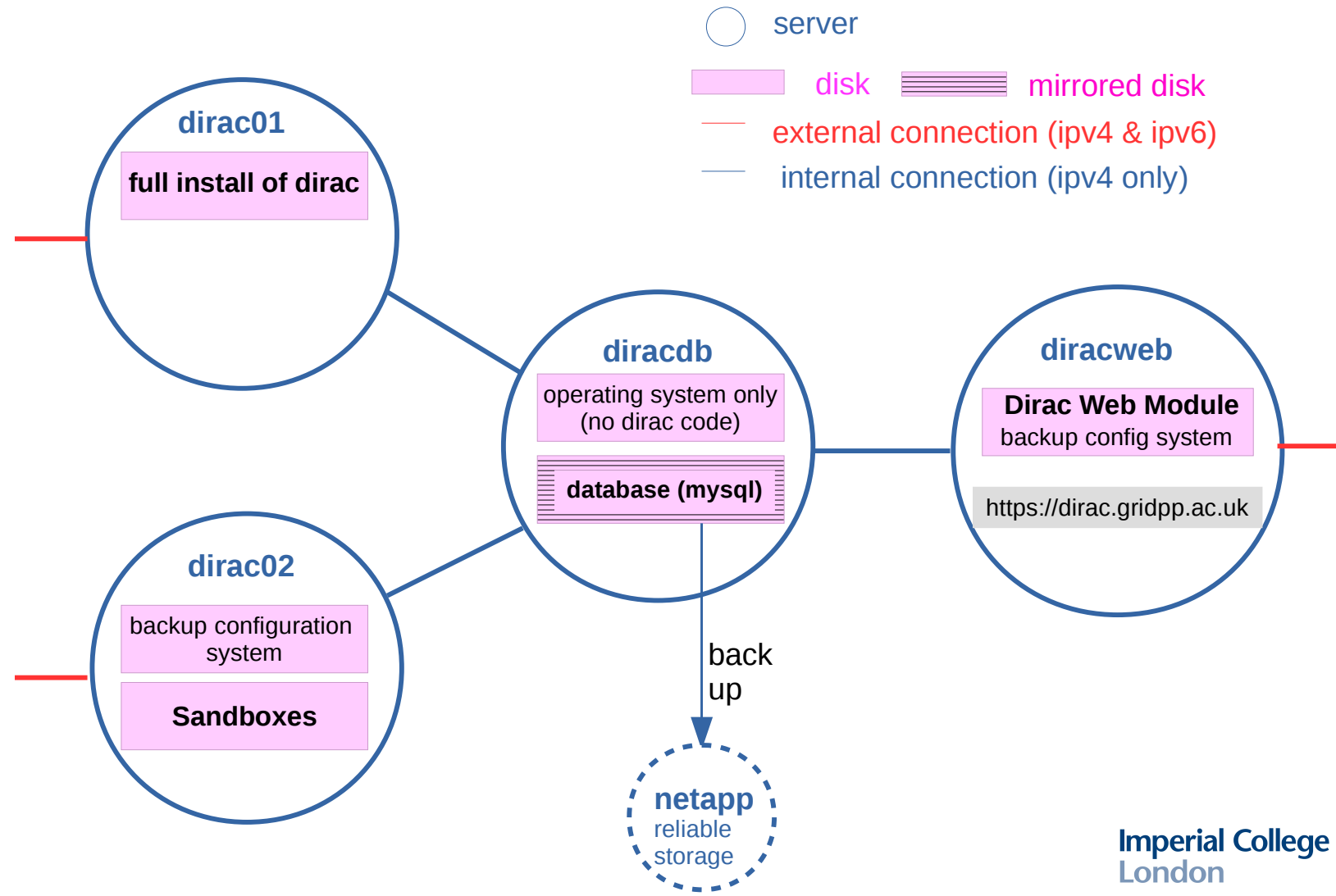
This has proven useful to allow small VOs access to LHC-centric resources.

There is room for improvement.

Thank you for listening :-)

Questions ?

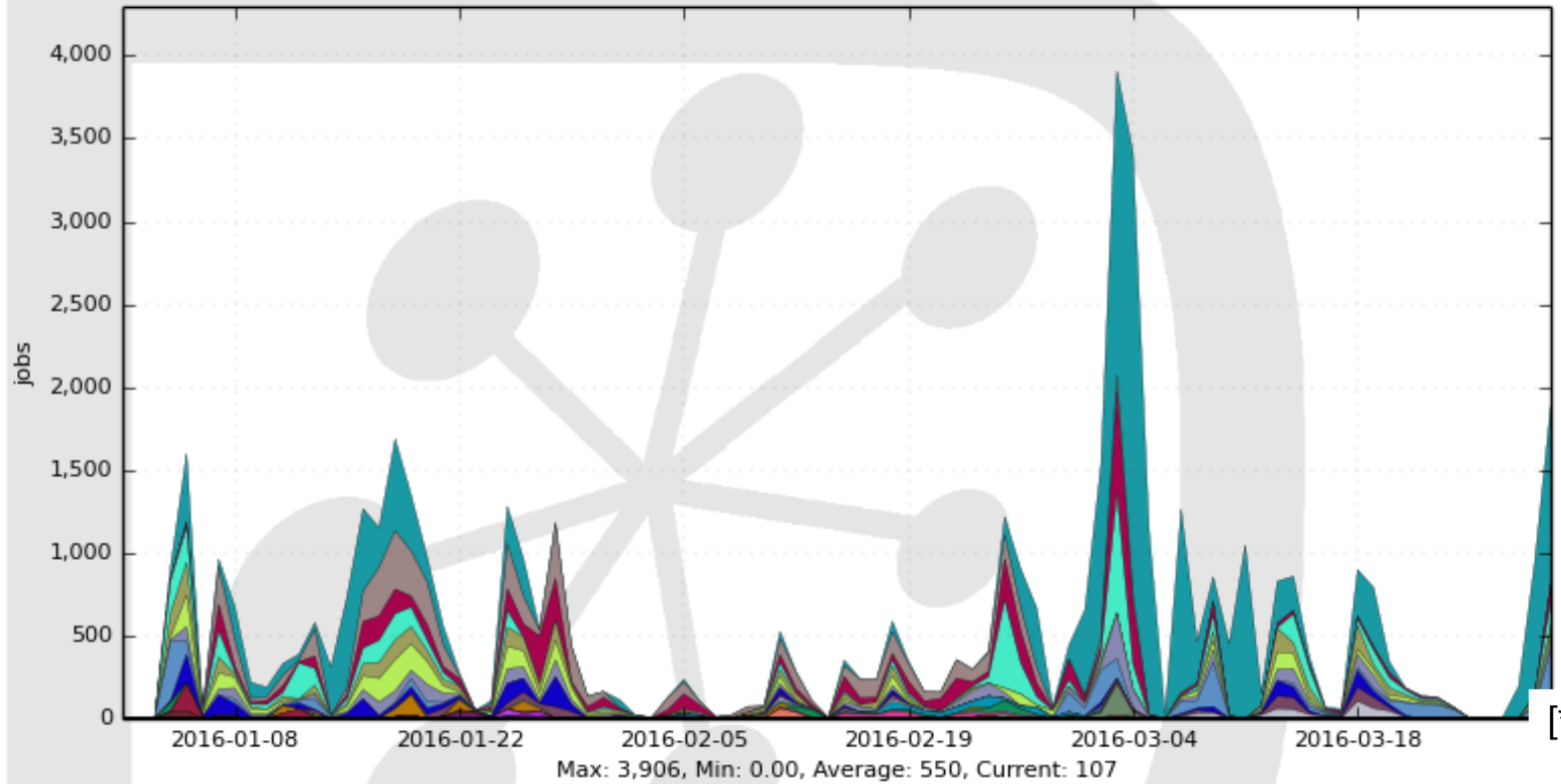
The GridPP DIRAC server



Usage by site

Running jobs by Site

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



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

[*]www.gridpp.ac.uk/vac/

LCG.UKI-LT2-IC-HEP.uk	38.0%	LCG.EFDA-JET.uk	1.7%	LCG.UKI-SOUTHGRID-CAM-HEP.uk	0.4%
LCG.RAL-LCG2.uk	10.7%	LCG.UKI-LT2-Brunel.uk	1.3%	VAC.UKI-NORTHGRID-MAN-HEP.uk	0.2%
LCG.UKI-NORTHGRID-LIV-HEP.uk	9.9%	LCG.NCG-INGRID-PT.pt	1.0%	VAC.UKI-LT2-UCL-HEP.uk	0.1%
LCG.UKI-NORTHGRID-MAN-HEP.uk	9.1%	LCG.UKI-SCOTGRID-GLASGOW.uk	0.9%	CLOUD.UKI-LT2-IC-HEP.uk	0.0%
LCG.UKI-LT2-RHUL.uk	5.2%	LCG.UKI-NORTHGRID-LANCS-HEP.uk	0.8%	VAC.UKI-SOUTHGRID-OX-HEP.uk	0.0%
LCG.UKI-SOUTHGRID-RALPP.uk	5.2%	LCG.CA-ALBERTA-WESTGRID-T2.ca	0.6%	LCG.UKI-SOUTHGRID-BHAM-HEP.uk	0.0%
LCG.UKI-SOUTHGRID-OX-HEP.uk	5.2%	LCG.UKI-SCOTGRID-ECDF.uk	0.6%	LCG.UKI-SOUTHGRID-BRIS-HEP.uk	0.0%
LCG.UKI-LT2-QMUL.uk	4.6%	LCG.EFDA-JET.xx	0.5%	CLOUD.CERN-PROD.ch	0.0%
LCG.UKI-SCOTGRID-DURHAM.uk	3.5%	LCG.UKI-SOUTHGRID-SUSX.uk	0.5%	... plus 2 more	