

DIRAC

(Distributed Infrastructure with Remote Agent Control)

Duncan Walker

IPPP

October 25, 2018

- ▶ DIRAC is a middleware/interface for the grid, which distributes jobs across all of the UK GridPP sites.
- ▶ Initially designed for LHCb, but now used much more widely
- ▶ In particular, it chooses where to submit based on what's free across the UK, rather than relying on the user to monitor resources.
- ▶ Gives access in one unified place to all available UK resources (even with a nice web interface!)
- ▶ Need to set up your jobs to run in a generic sandbox, without direct access to Durham filesystem [→ LFN storage, CVMFS, (GFAL)].

Site Listings

Lancs

Liverpool

Manchester

Sheffield

Durham

Glasgow

Birmingham

Cambridge

Oxford

Rutherford Labs

UCL

...

Striaightforward to exclude certain sites
if they're having trouble or you're
submitting directly at the same time

Total core count considerably larger
than Durham alone!

Important Bits and Bobs

- ▶ Jobs have lower priority than direct ARC submission
- ▶ 2 day job time limit
- ▶ Single core jobs only (not particularly usable for Mathematica runs with huge memory needs etc.)
- ▶ As far as I'm aware, CPU only
- ▶ Certificates last for an entire year - don't need to update daily
- ▶ You have to provide most libraries yourself (or ask for them to be put on CVMFS)
 - ▶ Python 2.4 is the newest version you can assume for the job sandboxes, so no python 3 :(
 - ▶ LHAPDF 6.1.6 works, LHAPDF 6.2.1 doesn't (path issues with PDF set storage)

Input/Output File Storage

- ▶ In my experience, the main issue in getting set up with DIRAC
- ▶ Jobs run in temporary sandboxes, without direct access to Durham filesystem
- ▶ Alternatives are needed to get software in, results out:
 - ▶ CVMFS for input executables/libraries that don't change often
 - ▶ LFN using `lfc*`, `lcg*` commands for input **and** output
[DEPRECATED, BUT WORKS EVERYWHERE]
 - ▶ GFAL using `gfal*` commands for input **and** output
[ARC ONLY, DIRAC HAS PYTHON ISSUES CURRENTLY]
- ▶ LFN+CVMFS is what I use. Slow and lossy, but generally ok.
- ▶ In general, some degree of experimentation and patience is needed

Installation

Installing Dirac is quite easy!

This information comes directly from

https://www.gridpp.ac.uk/wiki/Quick_Guide_to_Dirac

```
DIRAC_FOLDER=~/.dirac_ui"
DIRAC_VERSION="-r v6r20p5 -i 27 -g v14r1"

mkdir $DIRAC_FOLDER
cd $DIRAC_FOLDER
wget -np -O dirac-install https://raw.githubusercontent.com/DIRACGrid/DIRAC/integration/Core/scripts/
    ↳ dirac-install.py
chmod u+x dirac-install
./dirac-install $DIRAC_VERSION
source $DIRAC_FOLDER/bashrc
dirac-configure -F -S GridPP -C dips://dirac01.grid.hep.ph.ic.ac.uk:9135/Configuration/Server -I
dirac-proxy-init -g pheno_user -M
```

Syntax is much like for ARC submission

- ▶ Getting Started Guide

<https://dirac.readthedocs.io/...>

- ▶ Full Manual

<https://media.readthedocs.org/pdf/dirac/latest/dirac.pdf>

Nice features:

- ▶ Parametric job submission
 - ▶ Allows submission of up to 1k jobs with a varying parameter (e.g. MC integration seed, parameter for scanning etc.) in one go, rather than as 1k separate jobs → Much faster than ARC!
- ▶ All source is python based (can go and tinker if you don't like something...)
 - ▶ 180+ scripts for monitoring, submission etc. included in installation already.
 - ▶ Set up as modules if you want to have direct access to API through python imports

Example .jdl file:

```
JobName = "gridjob";
Executable = "nnlorun.py";
StdOutput = "StdOut";
StdError = "StdErr";
InputSandbox = {"nnlorun.py"};
OutputSandbox = {"StdOut","StdErr"};
Arguments = " --lhpdf_local lhpdf --runname top_secret --warmup_folder warmup --lhpdf_grid
    ↪ input/lhpdf.tar.gz --debug 0 --executable NNLOJET --output_folder output --input_folder
    ↪ input --runcard N4LO.run --lfn_dir /grid/pheno/dwalker --Production --seed %s";
Parameters = 500;
ParameterStart = 265001;
ParameterStep = 1;
ParameterFactor = 1;
```

```
dirac-wms-job-submit <filename>.jdl
```

Web Interface

Menu

Desktops/Applications

- Tools
- Applications
 - Public State Manager
 - Job Monitor
 - File Monitor
 - Accounting
 - Configuration Manager
 - Registry Manager
 - File Catalog
 - System Administration
 - Activity Monitor
 - Transformation Monitor
 - Request Monitor
 - File Summary
 - Resource Summary
 - Site Summary
 - Proxy Manager
 - Component History
 - Job Summary
- Help
- DIRAC
 - My Desktops
 - Shared

Settings

Job Monitor (Untitled 1)

Selectors

Site:

Status:

Minor Status:

Application Status:

Owner:

Owner Group:

Job Group:

Job Type:

Time Span:

From:

To:

Job ID(s):

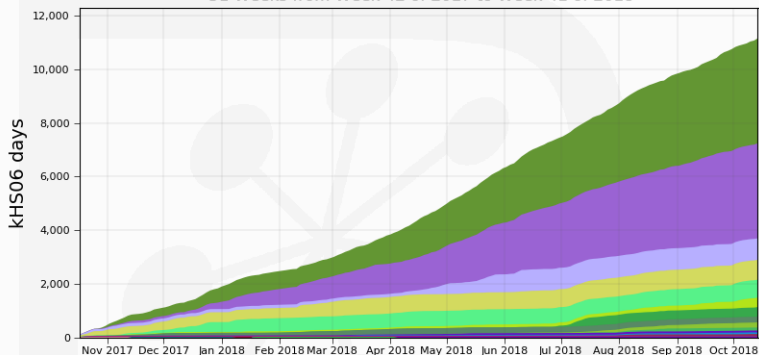
Items per page: 1000 | Page 1 of 2 | Updated: 2010-10-16 15:20:00 (UTC+01:00) | Displaying topics 1 - 1000 of 1527

JobID	Status	Minor Status	Application Status	Site	JobName	LastUpdate(JT/C)	LastSpool(JT/C)	SubmissionTime(JT/C)	Owner
13256188	Running		Unknown	LCG.UK1-L72...	grid06	2010-10-16 15:52:05	2010-10-16 15:52:05	2010-10-16 10:36:43	duncan.walker
13256164	Running		Application	LCG.UK1-L72...	grid06	2010-10-16 15:52:44	2010-10-16 15:52:57	2010-10-16 10:36:42	duncan.walker
13256162	Running		Application	LCG.UK1-L72...	grid06	2010-10-16 15:49:57	2010-10-16 15:50:09	2010-10-16 10:36:42	duncan.walker
13256147	Running		Application	LCG.UK1-L72...	grid06	2010-10-16 15:49:03	2010-10-16 15:49:19	2010-10-16 10:36:41	duncan.walker
13256141	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:05	2010-10-16 15:50:19	2010-10-16 10:36:41	duncan.walker
13256136	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:03	2010-10-16 15:50:16	2010-10-16 10:36:41	duncan.walker
13256130	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:49:35	2010-10-16 15:49:48	2010-10-16 10:36:41	duncan.walker
13256129	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:18	2010-10-16 15:50:31	2010-10-16 10:36:41	duncan.walker
13256119	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:41	2010-10-16 15:50:54	2010-10-16 10:36:41	duncan.walker
13256111	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:49:30	2010-10-16 15:49:42	2010-10-16 10:36:40	duncan.walker
13256109	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:27	2010-10-16 15:50:39	2010-10-16 10:36:40	duncan.walker
13256106	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:48:31	2010-10-16 15:48:45	2010-10-16 10:36:40	duncan.walker
13256105	Running		Application	VAC.UK1-SOUT...	grid06	2010-10-16 15:38:15	2010-10-16 15:38:29	2010-10-16 10:36:40	duncan.walker
13256098	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:48:34	2010-10-16 15:48:47	2010-10-16 10:36:40	duncan.walker
13256096	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:49:35	2010-10-16 15:49:48	2010-10-16 10:36:40	duncan.walker
13256091	Running		Application	Unknown	VAC.UK1-SOUT...	2010-10-16 15:38:05	2010-10-16 15:38:19	2010-10-16 10:36:40	duncan.walker
13256090	Running		Application	Unknown	VAC.UK1-SOUT...	2010-10-16 15:37:54	2010-10-16 15:38:08	2010-10-16 10:36:40	duncan.walker
13256081	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:08	2010-10-16 15:50:21	2010-10-16 10:36:40	duncan.walker
13256067	Running		Application	Unknown	VAC.UK1-SOUT...	2010-10-16 14:35:36	2010-10-16 14:35:58	2010-10-16 10:36:39	duncan.walker
13256063	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:48:23	2010-10-16 15:48:36	2010-10-16 10:36:39	duncan.walker
13256061	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:49:50	2010-10-16 15:50:03	2010-10-16 10:36:39	duncan.walker
13256059	Running		Application	Unknown	VAC.UK1-SOUT...	2010-10-16 15:52:03	2010-10-16 15:52:17	2010-10-16 10:36:39	duncan.walker
13256058	Running		Application	Unknown	VAC.UK1-SOUT...	2010-10-16 15:37:42	2010-10-16 15:37:56	2010-10-16 10:36:39	duncan.walker
13256056	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:52	2010-10-16 15:51:05	2010-10-16 10:36:39	duncan.walker
13256054	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:48:30	2010-10-16 15:48:43	2010-10-16 10:36:39	duncan.walker
13256051	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:49:06	2010-10-16 15:49:19	2010-10-16 10:36:39	duncan.walker
13256049	Running		Application	Unknown	VAC.UK1-SOUT...	2010-10-16 15:39:29	2010-10-16 15:39:44	2010-10-16 10:36:39	duncan.walker
13256048	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:49:12	2010-10-16 15:49:24	2010-10-16 10:36:39	duncan.walker
13256042	Running		Application	Unknown	VAC.UK1-SCOT...	2010-10-16 16:00:58	2010-10-16 16:01:13	2010-10-16 10:36:38	duncan.walker
13256039	Running		Application	Unknown	LCG.UK1-L72...	2010-10-16 15:50:57	2010-10-16 15:51:10	2010-10-16 10:36:38	duncan.walker
13256032	Running		Application	Unknown	LCG.UK1-NOR...	2010-10-16 15:42:03	2010-10-16 15:42:16	2010-10-16 10:36:38	duncan.walker
13256030	Running		Application	Unknown	LCG.UK1-NOR...	2010-10-16 15:22:00	2010-10-16 15:22:16	2010-10-16 10:36:38	duncan.walker

- ▶ Can monitor all of your jobs in one place
 - ▶ Group by site, status, name etc.
- ▶ Provides a quick preview of stdout if you want to check your submission(s)
- ▶ Able to directly download output files as tar.gz if you need to debug rare failures
- ▶ Can produce output stats to monitor your usage and/or environmental impact

Web Interface - Stats

Normalized CPU used by User
51 Weeks from Week 42 of 2017 to Week 41 of 2018



Max: 11,158, Min: 1.25, Average: 5,036, Current: 1.25

duncan.walker	3918.5	james.perry	111.2	matthew.hogan	1.0
robotgridclient3	3544.6	dimitrije.maletic	69.3	maksim.khabarov	0.8
lzproduser.grid.hep.ph.ic.ac.uk	803.8	behzad.hosseini	61.6	arka.santra.csic.es	0.8
elena.korolkova	742.3	sophie.king	49.6	tom.whyntie	0.2
juan.m.cruzmartinez	668.8	ewen.lawson.gillies	43.1	soliduser.grid.hep.ph.ic.ac.uk	0.1
jan.niehues	353.8	robotgridclient2	12.5	brais.lopez.paredes	0.0
james.whitehead	352.3	steve.lloyd	7.8	daniela.bauer	0.0
dflowprodproc	213.2	rohini.joshi	2.9	rebecca.hayley.pickles	0.0
ameir.shaa.bin.akber.ali	198.3	vladisavljevic.tomislav	1.5	... plus 43 more	

Generated on 2018-10-16 17:40:50 UTC

- ▶ Unified set of grid scripts with easy access to submission backends, programs and file storage systems with a local SQL database of jobs.
- ▶ Source: <https://github.com/scarlehoff/pyHepGrid>
- ▶ In collaboration with J.C. Martinez. Looking for alpha testers!

scarlehoff / pyHepGrid

Unwatch 3 Star 0 Fork 0

Code Issues Pull requests Projects Wiki Insights

Public version of the pyhepgrid project (early alpha)

5 commits 1 branch 0 releases 1 contributor GPL-3.0

Branch: public New pull request Create new file Upload files Find file Clone or download

scarlehoff Update README.md Latest commit de654a3 on 14 Sep

examples	public branch sync. with b99564e on Wed 12 Sep 14:22:28 CEST 2018	a month ago
pyhepgrid	public branch sync. with b99564e on Wed 12 Sep 14:22:28 CEST 2018	a month ago
.gitignore	initial commit, migrated from mercurial	a month ago
LICENSE	initial commit, migrated from mercurial	a month ago
README.md	Update README.md	a month ago
main.py	public branch sync. with b99564e on Wed 12 Sep 14:22:28 CEST 2018	a month ago

README.md

pyHepGrid

Public repository for the pyHepGrid project

What is pyHepGrid?