



# iLCDirac and the Grid

André Sailer

CERN-PH-LCD

AWLC

Software Meeting

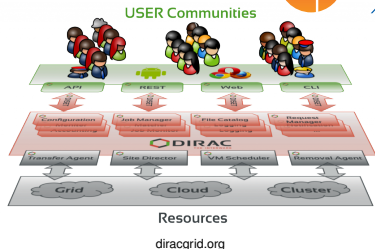
KEK, Japan, April 2015

# iLCDDirac in a Nutshell



iLCDDirac is based on the DIRAC interware originally developed for LHCb

- Dirac (Distributed Infrastructure with Remote Agent Control): High level interface between users and distributed resources
- iLCDDirac: Additional functionality to provide simple interface for the users to the LC Software (Whizard, Marlin, Mokka, org.lcsim, SLIC, ROOT)



```
from DIRAC.Core.Base import Script
Script.parseCommandLine()
import UserJob
import Marlin
import DiracILC
d = DiracILC()
j = UserJob()
j.setOutputSandbox("recEvents.slcio")
m = Marlin()
m.setVersion("0116")
m.setSteeringFile("Steering.xml")
m.setInputFile("SimEvents.slcio")
j.append(m)
j.submit(d)
```

# More Dirac Features:

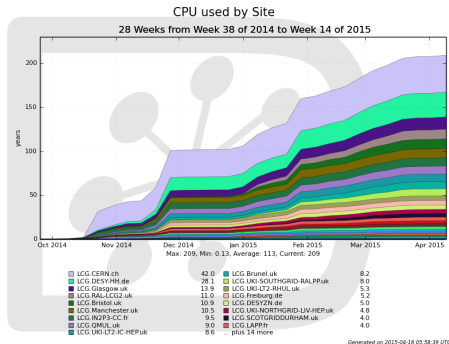


- Dirac File Catalog: Metadata and replica catalog
  - Command line interface with completion, metadata search, datasets
- Webinterface for monitoring and controlling of: User Jobs, Productions, Sites, Pilots, Dirac systems
  - Interesting developments in Belle2 Dirac for automatising monitoring of pilot failures
- ...

# CPU Consumption



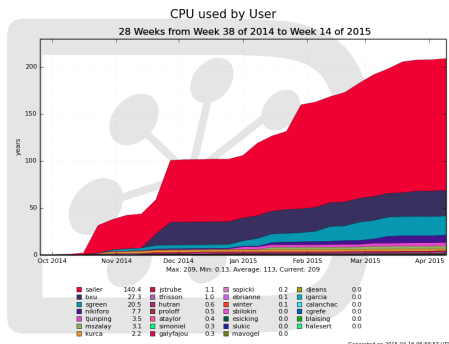
- CPU consumption in the last 6 months
- Not a lot of activity
  - ▶ Some smaller samples produced for CLICdp studies
  - ▶ Some user analyses



# CPU Consumption



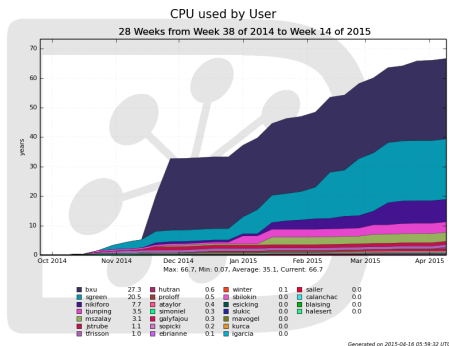
- CPU consumption in the last 6 months
- Not a lot of activity
  - ▶ Some smaller samples produced for CLICdp studies
  - ▶ Some user analyses



# CPU Consumption



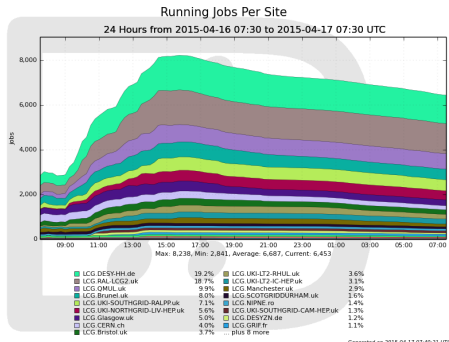
- CPU consumption in the last 6 months
- Not a lot of activity
  - ▶ Some smaller samples produced for CLICdp studies
  - ▶ Some user analyses



# Capacity



- 8k jobs in parallel maximum
- No more waiting jobs in the system at this moment
- Nicely distributed across many sites



Running jobs last Friday

# Production/Transformation System



Production: Centralised MC Production to generate, simulate and reconstruct millions of events

Transformation: Transform one set of files into another

- Generation, Simulation, Reconstruction, Replication, Splitting, Merging

CLICdp(SiD) transformation flow:

- 1 Generation (whizard): Define process, energy, number of events per file, ...
- 2 Simulation: Define detector model, input production ID, ...
- 3 Reconstruction: Define input production ID, overlay, ...

After transformation chain is defined set number of jobs for generation, the rest is done by the system

- Generator jobs create files, files are picked up by the simulation, and so on
- If sim or reco job fails, a new job is created, almost no interventions necessary
- I wish there were more documentation about the Dirac Transformation System though
  - ▶ I stumbled across the `dirac-transformation-cli` to add, reset or remove files from transformations by looking at tab-completion list...
  - ▶ Nothing for the O(10) Agents/Services involved

Will hear from Tino about the ILD Production specifics



There are some problems looming on the horizon:

- There are assumptions about file types and paths in the code
  - ▶ Whizard output files are assumed to be `stdhep`, will not be true for Whizard2
  - ▶ The way outputfiles are automatically named is a lot of spaghetti code (I think this was the biggest issue for ILD)
  - ▶ Need to think of a more generic and clean way in the near future
- Overloading of CERN Storage by just a few thousand parallel jobs
  - ▶ Need to make more use of other site local SEs at least to temporarily store files and then copy them to their final destination later on
  - ▶ Obviously this will be automatic via the Request Management System



## ■ Using dirac v6r11p31

- ▶ Started adapting to v6r11 with v6r11p14, quite a few bug-fixes necessary
  - ★ Very fast response from the developers in most cases
- ▶ Using ARC resources (More and more sites moving to this CE Type)
- ▶ Fixes in the file catalog permissions (e.g., `diracAdmin==root`)

## Software installation:

- CVMFS support for installations on CVMFS
  - ▶ source `init_ilcsoft.sh` from CVMFS
  - ▶ Use the same definition of applications in the ConfigurationSystem
  - ▶ Some special variables, e.g., for Mokka DB Slice
- CVMFS used as shared area for pre-installed tarballs (Not completely New)
  - ▶ If there is no CVMFS fall's back to downloading tarballs
  - ▶ If tarball not on CVMFS, falls back to downloading tarballs, still allows fast turn around for validating and debugging on the grid with large sample sizes
  - ▶ Can mix applications from CVMFS and working directory (NEW)
- Biggest trouble: Mokka and the DB
  - ▶ Re-discovered max. socket length of 117 characters, need to place it in /tmp
  - ▶ Now writing Mokka Log files to `workingDir/mysqllogs` so they can be added to `outputsandbox` for easier debugging
- When software is not cached, the jobs can time-out
  - ▶ If we could get a Stratum 1 in the US or Asia. . .



- LCG-Sites working
- Not working: OSG sites claiming to support ILC VO
  - ▶ Currently trying to get PNNL working, might be example to use other OSG sites with Globus or HTCondor Computing Elements using direct submission
  - ▶ Relying on WMS Resource Brokers is painful
- Cracow working on starting to support ILC-VO



- Virtual Machine based setup is running fine so far
- No overloads of the iLCDirac servers so far
- Only problems when CERN is in trouble (power cut, network trouble)



- Interface for Whizard 2
- Interface for ddsim.py
- Moving completely to DiracFileCatalog, dropping LFC
- Moving to next Dirac release v6r12
- Cleanup of application interfaces
- Improve documentation



■ If there are problems (ordered list):

- 1 FAQ: <https://twiki.cern.ch/twiki/bin/view/CLIC/DiracForUsers>
- 2 Submit a ticket to the issue tracker  
<https://its.cern.ch/jira/browse/ILCDIRAC> or top middle of  
[ilcdirac.cern.ch](http://ilcdirac.cern.ch), also has search function. (see the TWIKI for access)
- 3 [forum.linearcollider.org](http://forum.linearcollider.org)
- 4 Email: [ilcdirac-support@cern.ch](mailto:ilcdirac-support@cern.ch)

Registration: [ilcdirac-register@cern.ch](mailto:ilcdirac-register@cern.ch)

- ILCDirac is offering an interface for running the LC Applications on the grid
- Adopted by all detector concepts for centralised productions
- Easy to use for individual simulation, reconstruction or analyses
- No major developments planned, but small improvements foreseen
- Possibility to increase the available capacity to OSG resources