

iLCDirac: DIRAC for the linear collider

André Sailer

CERN-EP-LCD

DIRAC User Workshop
London, May 14–17, 2019

Section 1:



1 iLCDirac Use Case

- ILC VO: virtual organisation for linear colliders (ILC and CLIC),
- CALICE VO
- ILCDirac is an extension of the DIRAC system for the ILC VO
 - ▶ Workflow Modules for LC Software
 - ▶ Overlay System for adding beam background files to MC jobs
 - ▶ J. Phys.: Conf. Ser. ILCDirac, a DIRAC extension for the Linear Collider community. Proceedings of CHEP2013. 513 [CLICdp-Conf-2013-003](#)
- Centralized MC Production (Event Generation, Geant4 Simulation, Reconstruction)
- User jobs (Generation, Simulation, Reconstruction, Analyses)

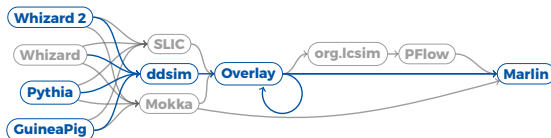
Capacity:

- Using WLCG and OSG resources
 - ▶ Mostly opportunistic, some dedicated
 - ▶ Around 15k to 20k job slots available at best of times

Code: <https://gitlab.cern.ch/CLICdp/ILCDIRAC>

- Define application payload via interfaces
- Chain applications (append one after the other)

```
from DIRAC.Core.Base import Script
Script.parseCommandLine()
import UserJob, Marlin, DiracILC
d = DiracILC()
j = UserJob()
j.setOutputData("recEvents.slcio")
m = Marlin()
m.setVersion("ILCSoft-01-17-09")
m.setSteeringFile("Steering.xml")
m.setInputFile("SimEvents.slcio")
j.append(m)
j.submit(d)
```



Section 2:



2 Current Status

Total of 100 Cores and 200 GB of Ram, SLC6 Virtual Machines

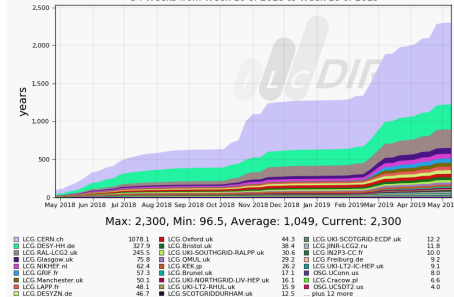
- 2 × 3 Servers running Agents and Services: 8 Cores, 16 GB RAM; Split by DIRAC-System
 - 1 Framework, Transformation, DataManagement, Configuration
 - 2 StorageManagement, WorkloadManagement
 - 3 RequestManagement, Accounting, ResourceStatus
- 3 DIRAC DIP-Storage SEs: 4 Cores, 8 GB Ram, 1 TB Volume
 - ▶ DIP-SE, Log-SE, SB-SE
- Web Server 4 Cores, 8 GB RAM
- DBs hosted on CERN DB on Demand (iLCDirac, ilcacdb (accounting DB), ilcdtest)
- Development, Testing, continuous integration (8 × 1 core), and spare VMs



- DIRAC Version: v6r20pX (Hot FiXes)
- Added additional Optimisation Executor chains on various servers to run Jobs through the Checking state
 - ▶ Lots (few 10k) of incoming jobs choke the system a bit

CPU used by Site

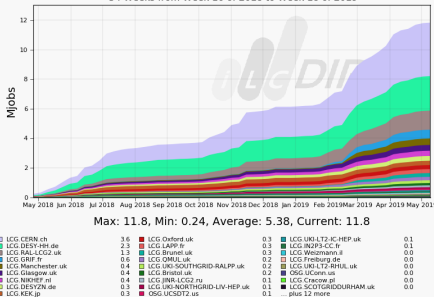
54 Weeks from Week 16 of 2018 to Week 18 of 2019



Generated on 2019-05-15 15:03:03 UTC

Cumulative Jobs by Site

54 Weeks from Week 16 of 2018 to Week 18 of 2019



Generated on 2019-05-15 15:03:03 UTC

2.3k CPU Years, 11.8M Jobs

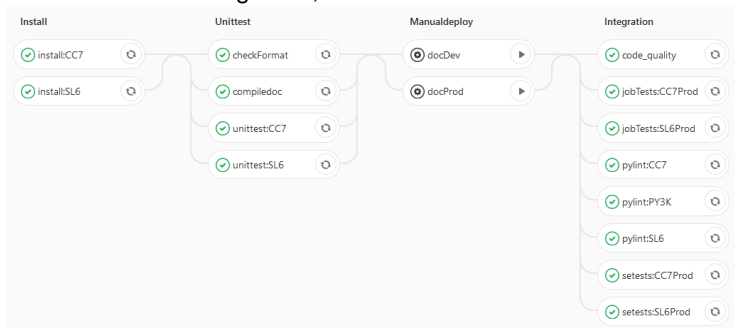
Section 3:



3 Developments in iLCDirac

PR+Nightly:

Nightly pipelines with: Rel-v29r0+rel-v6r20, Rel-v29r1+rel-v6r21,
master+integration, DIRAC_USE_M2CRYPTO=True



Tag:



Calibration System

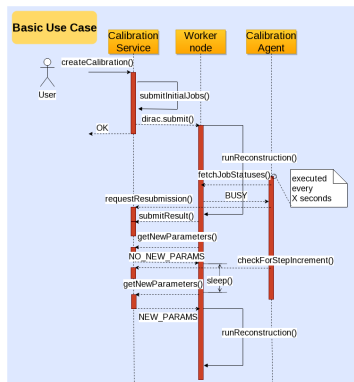


- Have to run an iterative procedure to obtain reconstruction calibration constants

- ▶ Split tasks into small chunks that run in parallel
- ▶ individual chunk takes only a few minutes
- ▶ for the next step, have to know output of all chunks
- ▶ repeat until constants stable

- Calibration System

- ▶ Service that collects results and distributes input parameters for next iteration
- ▶ Workload Module that runs reconstruction, reports results, obtains new parameter set and repeats
- ▶ Agent to monitor health of jobs, submits new ones when necessary



Developments by J. Ebbing, O. Viazlo

- We have many small files for which archiving to tape is inefficient
- Developing an RMS workflow to Archive files
 - 1 new `ArchiveFiles` operation: Copy all files to local disk, Create tarball from files, Upload tarball to a `StorageElement`
 - 2 (`ReplicateAndRegister`: Replicate Tarball from disk SE to tape SE)
 - 3 new `CheckMigration` operation: Wait for migration to tape
 - 4 Optional: Register “archived” replicas on separate SE to keep metadata
 - 5 Optional: Remove original replicas (or files)

- **DataRecoveryAgent(v2.0):** fix inconsistencies in transformation file and job status
 - ▶ Check each job: outputfile(s) exist(s), inputfile exists, what are the status(es), if other jobs treated the inputfile, and act accordingly
 - ▶ Has to know expected output files for job (Uses ProductionOutputData JDL parameter (iLCDirac specific?))
- **FileStatusTransformationAgent:** fix inconsistencies in Replication Requests (H. Zafar)
- **MonitorAgents Agent:** initially to monitor agents; now also services, executors; report and optionally restart in case of problems (H. Zafar)

Instance	Treatment	Log File Age (Minutes)
FileCatalog2	Please restart it manually	Not Relevant

Section 4:



4 Wish List

Wish List



- Fix CI errors after merge more quickly
- v7
- More Documentation
- Anything that reduces Ops workload