

# Monte Carlo Productions for Full Simulation Studies

André Sailer, Brieuc Francois, Lorenzo Valentini

CERN-EP-SFT

3rd ECFA Workshop on Higgs/Electroweak/Top Factory Studies  
October 9–11, 2024  
Paris, France

# Table of Contents



1 Introduction: DIRAC in a Nutshell

2 Interface

- VO Registration
- For Production Managers
- Data Access

3 Usage In 2024

4 Documentation and Support

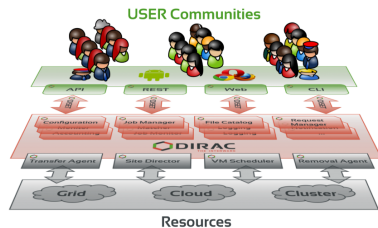
5 Conclusions

# Dirac in a Nutshell



iLCDirac is based on the DIRAC interware originally developed for LHCb

- High level interface between users and distributed resources
- Distributed Workload Management: one interface to execute anywhere: batch farms, grid computing elements, HPCs
- Data Management (file transfers, meta data augmented file catalog)
- Transformation system with high degree of automation
- Web interface for controlling jobs, accounting, system administration
- Supporting Virtual Organisations (VOs): CALICE, FCC, ILC (including CLIC)
  - ▶ CEPC is using a DIRAC instance supported by IHEP, Beijing



[diracgrid.org](http://diracgrid.org)

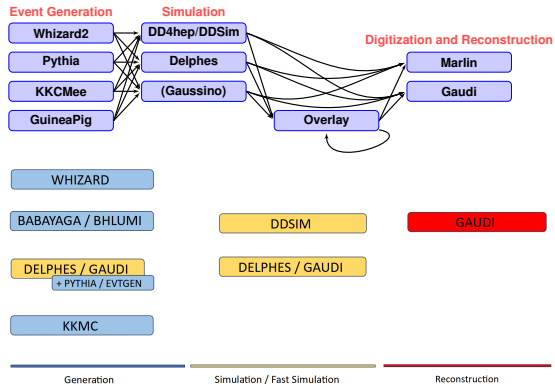
# Job Example: Reconstruction



```
1 from DIRAC.Resources.Catalog.FileCatalog import FileCatalog
2 import DiracILC, UserJob, GaudiApp # incomplete for lack of space
3
4 dIlc = DiracILC()
5
6 inputData = FileCatalog().findFilesByMetadata({'ProdID': 23456, 'DataType': 'SIM'})
7 inputData = inputData['Value'] # assuming success
8
9 job = UserJob()
10 job.setName("SplitTest_%n") # %n will be replaced by the task number
11 job.setFCCConfig('key4hep-latest')
12 job.setOutputData("RecoTest.slcio", OutputPath="RecoTest")
13 job.setSplitInputData(inputData, numberOfFilesPerJob=10)
14
15 gaudi = GaudiApp()
16 gaudi.setVersion('key4hep-latest')
17 gaudi.setSteeringFile('fcc_e4h_reco.py')
18 gaudi.setOutputFile('RecoTest.root')
19
20 job.append(ga); job.submit(dIlc)
```

# Available Applications

- Many applications are available with dedicated **Python interfaces** to configure them
- More MC generators have been added recently to support FCC Production Workflows and User jobs
- Interfaces for Key4hep based applications, such as k4SimDelphes



- VOMS Admins run by DESY (CALICE, ILC) and CERN IT (FCC) have been replaced by IAM services run by CERN IT
  - ▶ CALICE: <https://calice-auth.cern.ch/start-registration>
  - ▶ FCC: <https://fcc-auth.cern.ch/start-registration>
  - ▶ ILC: <https://ilc-auth.cern.ch/start-registration>
- CALICE and ILC VO members have to register again in respective IAM service
- New members have to register to their respective VOs
- Otherwise nothing changes from the user point of view when using iLCDirac!

# Registration



## Registration Workflow

- 1 Click the link(s) related to your VO(s)
- 2 Fill the form as shown on the right
  - ▶ Do not worry about username, I will probably change it
  - ▶ Do write a short note, so we know who you are
- 3 Confirm your registration via link from email
- 4 Wait for administrator approval
- 5 After approval, you need to set a password for your account
- 6 And finally add your x509 certificate to the account
- 7 Soon after, you will be auto-registered in iLcDirac

To proceed with the registration please fill in your personal information below.

**Given name**

**Family name**

**Email**

**Username**

**Notes**

Providing a clear explanation on the motivation behind this request will likely speed up the approval process

By submitting this registration request, you declare that you agree with the terms of this organization [Acceptable Usage Policy](#) and [Privacy policy](#).

# Registration



## Registration Workflow

- 1 Click the link(s) related to your VO(s)
- 2 **Fill the form as shown on the right**
  - ▶ Do not worry about username, I will probably change it
  - ▶ Do write a short note, so we know who you are
- 3 Confirm your registration via link from email
- 4 Wait for administrator approval
- 5 After approval, you need to set a password for your account
- 6 And finally add your x509 certificate to the account
- 7 Soon after, you will be auto-registered in iLcDirac

To proceed with the registration please fill in your personal information below.

**Given name**

**Family name**

**Email**

**Username**

**Notes**

Providing a clear explanation on the motivation behind this request will likely speed up the approval process

By submitting this registration request, you declare that you agree with the terms of this organization [Acceptable Usage Policy](#) and [Privacy policy](#).

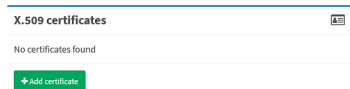
Register

Reset Form



## Registration Workflow

- 1 Click the link(s) related to your VO(s)
- 2 Fill the form as shown on the right
  - ▶ Do not worry about username, I will probably change it
  - ▶ Do write a short note, so we know who you are
- 3 Confirm your registration via link from email
- 4 Wait for administrator approval
- 5 After approval, you need to set a password for your account
- 6 **And finally add your x509 certificate to the account**
- 7 Soon after, you will be auto-registered in iLCDirac



- Large Monte Carlo samples require large amount of CPU hours
- Efficient and automatic management of job handling is essential
  - ▶ Easy submission of workflows
  - ▶ Automatic handling of job (re)submission, consistency checks
  - ▶ Monitoring and ergonomic error diagnostics

# Transformation Definition



Over the years, different scripts developed

- ILD: ILDProductionChain.py
- CLIC: dirac-clic-create-transformation
- FCC: dirac-fcc-create-transformation

Transformation definition also includes the definition of the output path, filenames, and meta data information to be stored

# Configuration example: FCC: Whizard2 and Delphes



- Production Managers define productions with config file, multiple transformation can be defined in one go
- Mostly vary Event type and energy, generator steering file
- Application parameters can be changed as well
- Store config files in git for later re-use debugging

```
1 [whizard2]
2 Version = 2.8.3
3 [delphesapp]
4 ExecutableName = DelphesSTDHEP_EDM4HEP
5 DetectorCard = card_IDEA.tcl
6 OutputCard = edm4hep_IDEA.tcl
7 Version = key4hep_230408
8 [Production Parameters]
9 machine = ee
10 softwareVersion = key4hep_230408
11 generatorApplication = whizard2
12 generatorSteeringFile = ee_ZH_240.sin
13 processingAfterGen = delphesapp
14 configVersion = key4hep-devel-2
15 configPackage = fccConfig
16 eventsPerJobs = 1000
17 campaign = winter2023
18 energies = 240
19 processes = ZH
20 detectorModel = idea
21 outputSE = CERN-DST-EOS
22 ProdTypes = Gen
```

# Output Meta Data



Output path:

```
■ /fcc/ee/<season-year>/<energy>/<evt-type>/<detector>/<datatype>/<ProdID>
```

Metadata:

```
/fcc/ee/winter2023/240gev/: {'Energy': '240'}  
/fcc/ee/winter2023/240gev/ZH/: {'EvtType': 'ZH'}  
/fcc/ee/winter2023/240gev/ZH/idea: {'DetectorType': 'idea'}  
/fcc/ee/winter2023/240gev/ZH/idea/delphes: {'Datatype': 'delphes'}  
/fcc/ee/winter2023/240gev/ZH/idea/delphes/00012345:  
  {'ProdID': 12345, 'NumberOfEvents': 1000}  
/fcc/ee/winter2023/240gev/ZH/stdhep: {'Datatype': 'stdhep'}  
/fcc/ee/winter2023/240gev/ZH/stdhep/00012345:  
  {'ProdID': 12345, 'NumberOfEvents': 1000}
```

Registered non searchable metadata:

```
/fcc/ee/winter2023/240gev/ZH/idea/delphes/00012345 =  
  {'SWPackages': 'whizard2.2.8.3;delphesapp.key4hep_230408'}  
/fcc/ee/winter2023/240gev/ZH/stdhep/00012345 =  
  {'SWPackages': 'whizard2.2.8.3;delphesapp.key4hep_230408'}
```

- To provide information of available samples and parameters a constantly running agent is providing a json file that is consumed and published on an [FCC website](#)

```
{  
  "last_file_update": "2024-08-29 11:15:32.927707+00:00",  
  "transformations": {  
    "16622": {  
      "Status": "Active",  
      "Version": 0,  
      "cross-section": 6.77319253587844,  
      "cross-section-error": 0.0001743518712179634,  
      "efficiency": 0.04,  
      "total-number-of-events": 498000,  
      "number-of-events-per-file": 1000,  
      "production-manager": "/DC=ch/DC=cern/OU=Organic Units/OU=Users/...",  
      "path": "/fcc/ee/test_spring2024/240gev/mumuH-mH-lower/CLD_o2_v05/rec/00016622"  
    }  
  }  
}
```

# Extending Transformations: More, more, more!

## TransformationMonitor WebApp



The screenshot shows the TransformationMonitor WebApp interface. At the top, there are buttons for 'Start', 'Stop', 'Flush', 'Complete', and 'Clean'. Below these is a table with columns: ID, Status, AgentType, Type, Name, Files, Processed (%), Created, Total Created, and Submitted. The table contains several rows of data, including jobs with IDs 401242, 401241, 401189, 401136, 401135, and 401134. A context menu is open over the table, showing options like 'Show Jobs', 'Logging Info', 'Workflow xml', 'File Status', 'File Retries', 'InputData Query', 'Additional Params', and 'Show Details'. The 'Actions' sub-menu is also open, showing 'Start', 'Stop', 'Extend', 'Flush', 'Complete', and 'Clean'. A tooltip points to the 'Extend' option with the text 'Click to extend the selected transformation(s)'.

ID	Status	AgentType	Type	Name	Files	Processed (%)	Created	Total Created	Submitted
401242	Active	Automatic	MCGeneration	tautau_91.2 ee... 0		0	0 (-20)	20	0
401241	Active	Automatic	MCGeneration	tautau_91.2 ee... 0		0	0	20	0
401189	Active	Automatic	MCGeneration	bb_4		0	0	30	0
401136	Active	Automatic	MCRReconstruct...	bb_4		95.0	0	740	0
401135	Active	Automatic	MCSimulation	bb_4		100.0	0	286	0
401134	Active	Automatic	MCGeneration	bb_4		0	0	280	0

Or

- `dirac-ilc-add-tasks-to-prod --total [prodID] 500`
- DIRAC can automatically extend transformations when jobs fail

- Centrally produced samples are uniquely identified by *ProductionID*
- With the ProductionID we can find out everything about how the connected files were produced
  - ▶ `dirac-ilc-get-info -p 12345`
  - ▶ `dirac-ilc-get-prod-log -P 12345`
- We can find the related files
  - ▶ `dirac-dms-find-lfns Path=/ ProdID=12345 Datatype=delphes`
- We can also use meta data to find files when we do not know the ProductionID
  - ▶ `dirac-dms-find-lfns Path=/ EvtType=ZH Energy=240 Datatype=delphes DetectorType=idea`



# How to Access the Samples



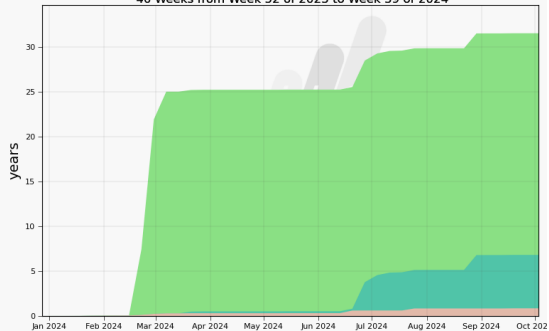
- Access can always be done via (ilc)dirac tools: `dirac-dms...`
  - ▶ `dirac-dms-find-lfns`, `dirac-dms-get-file`, `dirac-dms-filecatalog-cli`, [WebApp](#)
  - ▶ Isolation of file access and storage specifics: KEK changed access protocol in complete transparency to users
- Depending on site configuration direct local read-access is also possible!
  - ▶ CERN: `/eos/experiment/fcc/prod` ; `/eos/experiment/clicdp/grid/ilc`
  - ▶ DESY (NAF): `/pnfs/desy.de/ilc/prod/ilc/`
- You can probably infer local access via the `dirac-dms-lfn-replicas` command

# CPU Use in 2024



### CPU used by UserGroup

40 Weeks from Week 52 of 2023 to Week 39 of 2024



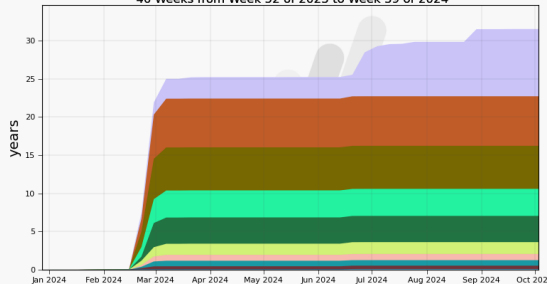
Max: 31.5, Min: 0.00, Average: 21.7, Current: 31.5

ilc\_prod 24.7 fcc\_prod 6.0 ilc\_user 0.8 fcc\_user 0.0

Generated on 2024-10-08 08:12:53 UTC

### CPU used by Site

40 Weeks from Week 52 of 2023 to Week 39 of 2024



Max: 31.5, Min: 0.00, Average: 21.7, Current: 31.5

LCG.CERN.ch	8.8	LCG.Freiburg.de	0.8	Multiple	0.0
LCG.KEK.jp	6.5	LCG.UKI-LT2-IC-HEP.uk	0.7	LCG.RAL-LCG2.uk	0.0
LCG.Manchester.uk	5.6	LCG.IHEP.cn	0.4	LCG.CNAF.it	0.0
LCG.DESY-HH.de	3.5	LCG.JINR-LCG2.ru	0.1	ANY	0.0
LCG.IN2P3-CC.fr	3.4	LCG.GRIF.fr	0.0	LCG.UKI-SOUTHGRID-RALPP.uk	0.0
LCG.DESYZIN.de	1.6	LCG.Tau.it	0.0	LCG.UKI-SCOTGRID-ECDF.uk	0.0

Generated on 2024-10-08 08:13:20 UTC

## ■ In case of fire:

- 1 Consult documentation: <https://ilcdircac-doc.web.cern.ch/>
- 2 Before submitting a ticket, see:  
<https://ilcdircac-doc.web.cern.ch/DOC/Files/UserGuide/support.html>
- 3 Submit a ticket to the issue tracker <https://its.cern.ch/jira/browse/ILCDIRAC>
  - ★ See also “Report a Problem” buttons in web portal and documentation (it is hopefully in the foreground!)
- 4 Email: [ilcdircac-support@cern.ch](mailto:ilcdircac-support@cern.ch)



# Conclusions



- iLCDirac has been used for many years successfully by ILC and CLIC communities
- Adapting to changes in Grid infrastructure: VOMSAdmin → IAM, Tokens
- Developed new features for FCC and other users
- Now able to create samples for ECFA H/E/T studies and ready for users

[ilcdircac.cern.ch](https://ilcdircac.cern.ch)

[fccdirac.cern.ch](https://fccdirac.cern.ch)

# Acknowledgements



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 871072

Thank you for your attention!