

Bookkeeping Tutorial





Bookkeeping content

- Contains records of all "jobs" and all "files" that are produced by production jobs
- o Job:
 - In fact technically a "step" in a workflow
 - ★ E.g. "Gauss step", "Brunel step"...
 - For real RAW data: the "job" is in fact a DAQ run
 - Has input files (except runs and Gauss)
 - Has output files
 - Note that files may not be kept (i.e. have a replica)
 - All files are registered in order to keep the full history
 - Has metadata
 - ★ Location, production number, application, CPUTime, etc...
- o Files:
 - Always output of a "job"
 - Files are defined by an LFN (Logical File Name)
 - Contain metadata
 - ★ Number of events, size, event type, etc...





Bookkeeping purpose

- Provenance database
 - Contains the full history of productions
 - ★ Traceability of datasets
- User dataset search
 - Select a list of files from selection criteria
 - ★ Only files with a replica!
 - Generate Gaudi configuration file
 - Give also access to the job/file tree
 - ★ E.g. investigate history of a file
- Production datasets search
 - Select the dataset to be processed by production jobs
 - ★ Ensures consistency of input files for a production
 - Uses directly the BK API to get the list of files





Bookkeeping partitioning

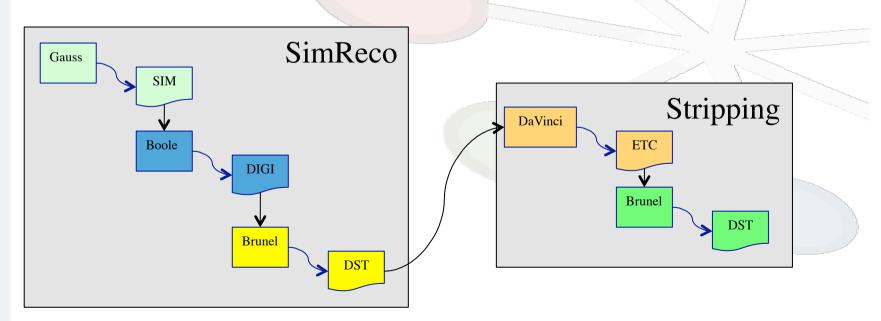
- Configuration Name / version
 - Real data
 - ★ <DAQ partition> / <activity>
 - Simulated data
 - - * <activity>: "2008" / "DC06" / ...
- Conditions
 - Parameters of initial data
 - All subsequent processed data inherit the "conditions"
 - Real data
 - **☆ DAQ conditions**
 - * Beam conditions, energy, magnetic field, detector conditions...
 - Simulated data
 - ★ Simulation conditions
 - * Beam energy, magnetic field, luminosity, generator settings...





Processing pass

- Associated to a level of processing
 - Within a given partition (config name / version + conditions)
 - Corresponds to the whole processing workflow
 - ★ Single workflow for a given processing pass
 - ★ Compatible versions of applications
 - Specifies the processing pass of input data when applicable
 - Sequence of processing
 - Re-processing creates branches







Other query parameters

- Event type
 - File property
 - Real data
 - Simulated data
 - ★ LHCb convention for decay tree
- File type
 - Data content / format
 - ★ Format not yet used





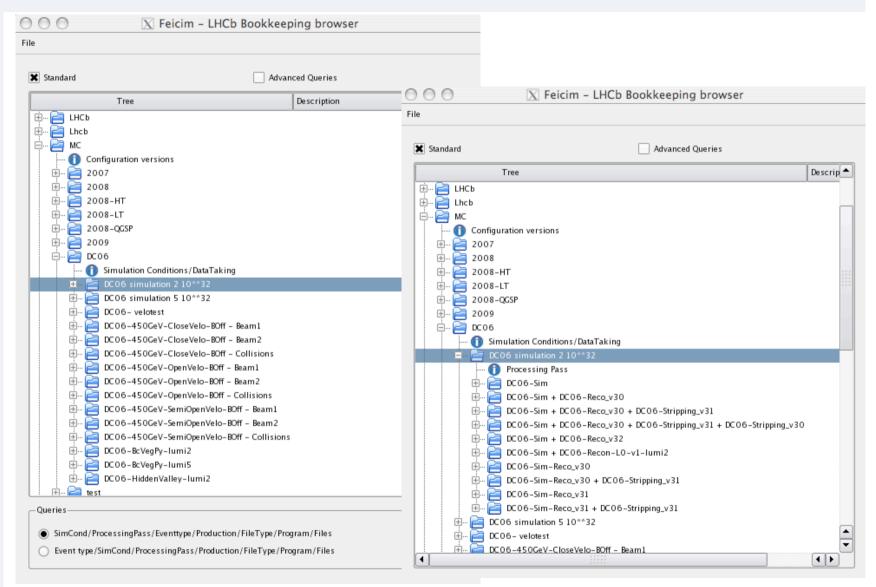
Running the bookkeeping GUI

- Needs a valid Grid certificate
- Needs an X server
- o lhcb-bkk
 - SetupProject Dirac
 - ★ Sets up the environment
 - lf needed: lhcb-proxy-init
 - ☆ Creates a proxy
 - dirac-bookkeeping-gui
- Individual commands can be issued from the prompt!





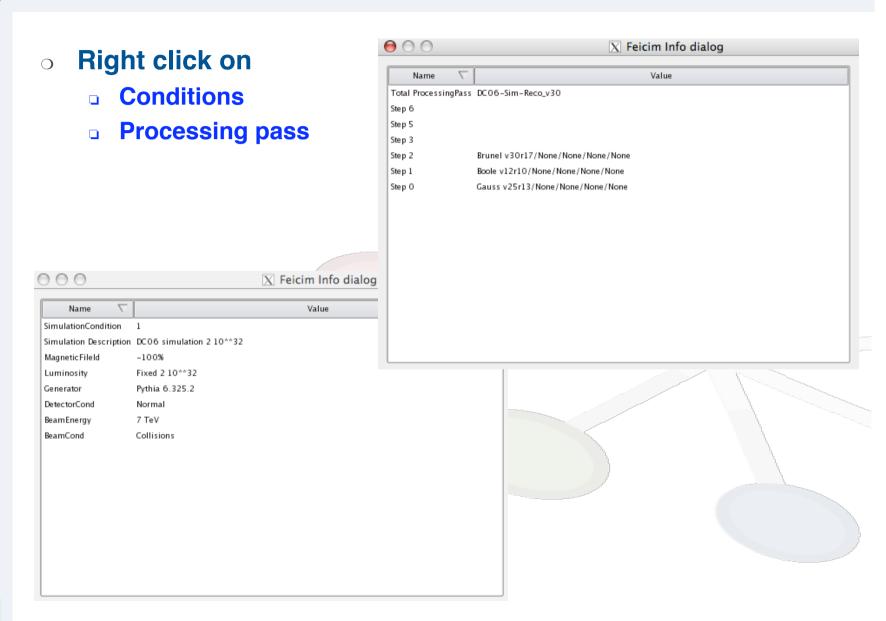
The query tree







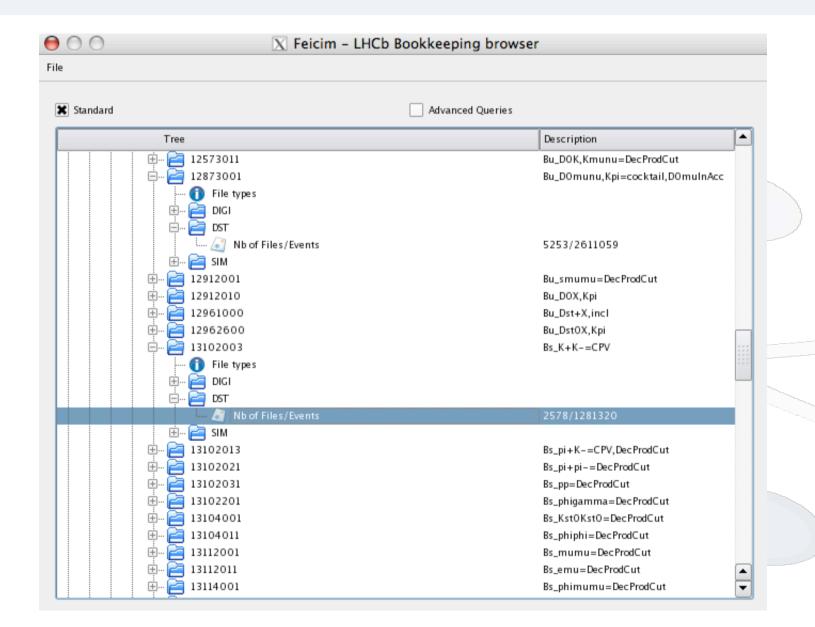
More info







Event type and file type

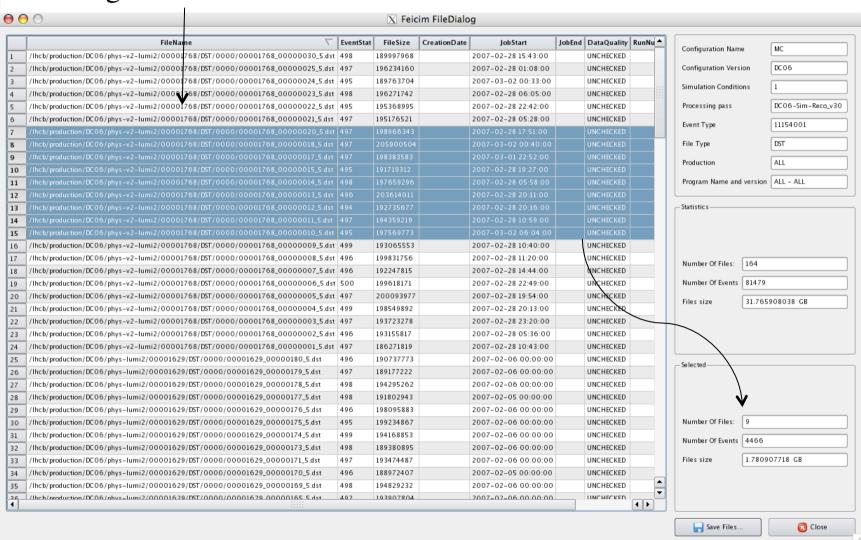






Dataset selection

Logical File name







Saving configuration (a.k.a. options) file

- Python configuration (default)
 - Still possible to create .opts (discouraged!)
 - .txt file for just a list of LFNs

All files or selected files (if any) X Feicim Save file(s) dialog 🔽 🖯 🖯 🞧 👺 🛅 🖽 /afs/cern.ch/user/p/phicharp Look in: _ORPHANDIR_.705.178890 amyGangaJobs Computer 🚧 ns_imap agent_runsv alient2 🚄 nsmail physics cmtuser Desktop plots 롣 private gangadir inis 🚰 🚄 public 🚄 k0s scripts LFCtest HCbGrid newOptions.py CPUCalib.py pool_xml_catalog.py DIRAC 🕋 test Dirac Example toto.py Emacs xxxxx.py [[xplus209] ~ > more myOptions.pv a logs #-- GAUDI jobOptions generated on Wed Dec 10 15:28:19 2008 ail mail #-- Contains event types : myDLLs 10012000 - 4 files - 10429 events - 5.05 GBytes myOptions.py File name: from Gaudi.Configuration import * Files of type: Python option(*.py) Cancel EventSelector().Input = [DATAFILE='LFN:/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001592_2.dst' TYP='POOL_ROOTTREE' OPT='READ'" DATAFILE='LFN:/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001601_2.dst' TYP='POOL_ROOTTREE' OPT='READ'" DATAFILE='LFN:/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001608_2.dst' TYP='POOL_ROOTTREE' OPT='READ'" DATAFILE='LFN:/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001615_2.dst' TYP='POOL_ROOTTREE' OPT='READ'"] [lxplus209] ~ >





Dealing with PFNs or XML catalogs

- Using ganga + DIRAC
 - Bookkeeping integrated in ganga:
 - ☆ dataset = browseBK()
 - LFN handling is then automatic...
- If you really need XML catalog or PFNs, use genXMLCatalog
 - Ensures files are available on the specified site
 - Gets the PFN from the Storage Element
 - Not constructed "by hand"

```
[Ixplus209] ~ > genXMLCatalog --help
Usage: genXMLCatalog 
Options:
    -s, --site <site>: site name (default=CERN)
    -d, --depth <depth>: depths for ancestors in BK (default=1)
    -f, -p, --catalog <atalog-name>: XML file catalog name (default=./pool_xml_catalog.xml)
    -n, --newoptions <anfig-file>: generate a new config file (no catalog is created)
    -o, --options <anfig-file>: python config file to be parsed (for backward compatibility)
    <a href="mailto:sonfig-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-file>">config-fil
```





Dealing with XML catalog and PFNs

```
~ > genXMLCatalog --site CERN --catalog myCatalog myOptions.py
POOL XML catalog myCatalog.xml successfully created
Add myCatalog.py to your list of options file
 txplus2091 ~ > more myCa
                 myCatalog.xml
                                   myCatalog.xml.temp
myCatalog.py
flxplus2891 ~ > more myCatalog.*
.....
myCatalog.py
......
FileCatalog().Catalogs = [ 'xmlcatalog_file:myCatalog.xml' ]
......
nyCatalog.xml
......
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!-- Edited By PoolXMLCatalog.py -->
<!DOCTYPE POOLFILECATALOG SYSTEM "InMemory">
₽00LFILECATALOG>
 -File ID="5E4AAA90-835A-DD11-AA07-0019B9F6E9F1">
     32069/DST/0000/00002069_00001615_2.dst"/>
    </physical>
    <logical>
     <lfn name="/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001615_2.dst"/>
    </loaical>
  </File>
[Lxplus209] ~ > genXMLCatalog --site CERN --newoptions myPFNOptions myOptions.py
New options file myPFNOptions.py successfully created
lxplus2891 ~ > more myPFNOptions.py
#-- GAUDI jobOptions generated on Wed Dec 10 15:28:19 2008
⊭_− Contains event types :
from Gaudi.Configuration import *
EventSelector().Input = [
  DATAFILE='castor://castorlhcb.cern.ch:9002/?svcClass=lhcbdata&castorVersion=2&path=/castor/cern.ch/qrid/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001592
2.dst' TYP='POOL ROOTTREE' OPT='READ'".
  DATAFILE='castor://castorlhcb.cern.ch:9002/?svcClass=lhcbdata&castorVersion=2&path=/castor/cern.ch/grid/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001601
_2.dst' TYP='POOL_ROOTTREE' OPT='READ'".
  DATAFILE='castor://castorlhcb.cern.ch:9002/?svcClass=lhcbdata&castorVersion=2&path=/castor/cern.ch/grid/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001608
_2.dst' TYP='POOL_ROOTTREE' OPT='READ'".
  DATAFILE='castor://castor!hcb.cern.ch:9002/?svcClass=lhcbdata&castorVersion=2&path=/castor/cern.ch/arid/lhcb/production/DC06/v1r0/00002069/DST/0000/00002069_00001615
_2.dst' TYP='POOL_ROOTTREE' OPT='READ'"1
```





DIRAC Monitoring web portal





General information

- Entry point to the DIRAC web portal
 - http://dirac.cern.ch
- Web implementation of (almost) a full desktop application
 - Monitoring of productions / jobs
 - Accounting (jobs, data management)
 - Allows to take actions on jobs
- Authentication / authorisation is mandatory
 - Anonymous access gives minimal access
 - Get a certificate and load it in our in your browser https://twiki.cern.ch/twiki/bin/view/LHCb/FAO/Certificate
 - DIRAC authorisation through "DIRAC groups"

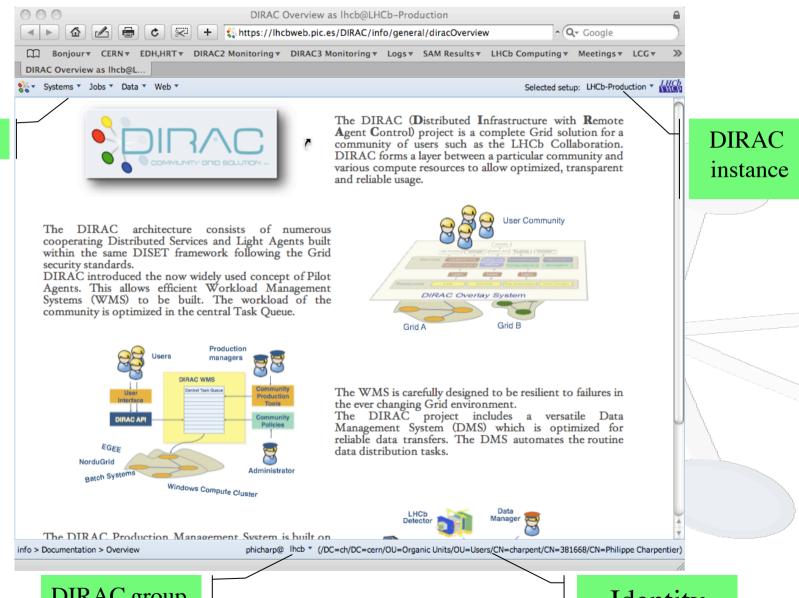
 - → Other groups: Ihcb_prod, dirac_admin...
 - ★ Future: specific groups per physics groups, PPG (for production authorisation)...
 - ★ Capabilities depends on the group





Menus

The DIRAC portal home page



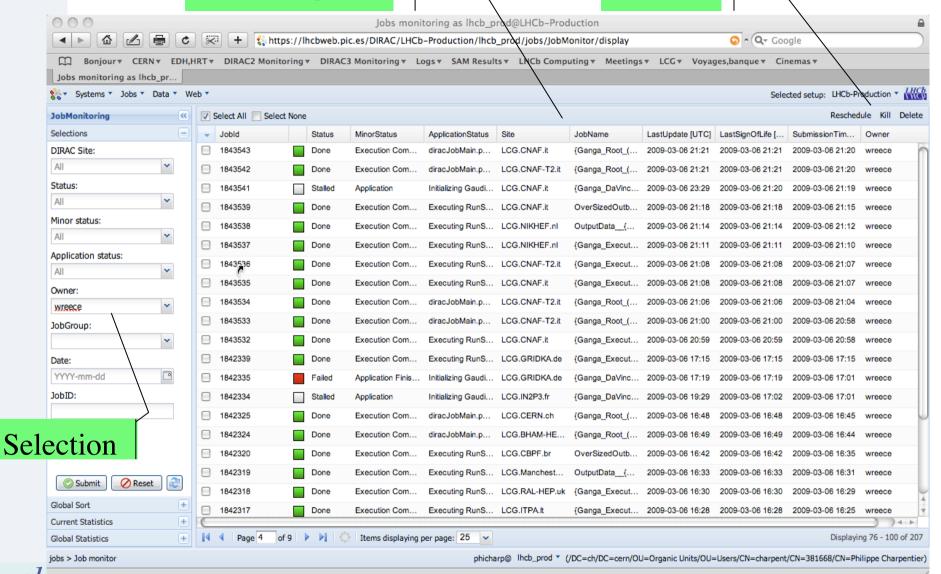




Job Monitoring

Monitoring info

Actions







Job Monitoring (cont'd)

- Selection
 - For group 1hcb_user, only see your own jobs
 - Can select with

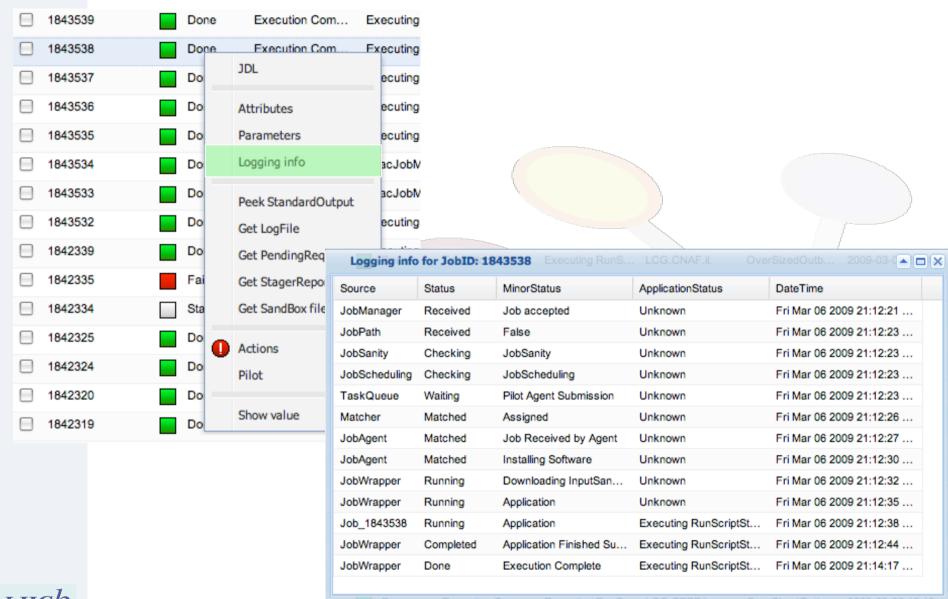
 - Site

 - ☆ ...
- o Columns
 - Can tailor the columns to be displayed
 - Clicking toggles the sorting in the column
- Rows
 - Jobs displayed in pages (default 25 rows, don't exceed 100)
 - Can scroll pages





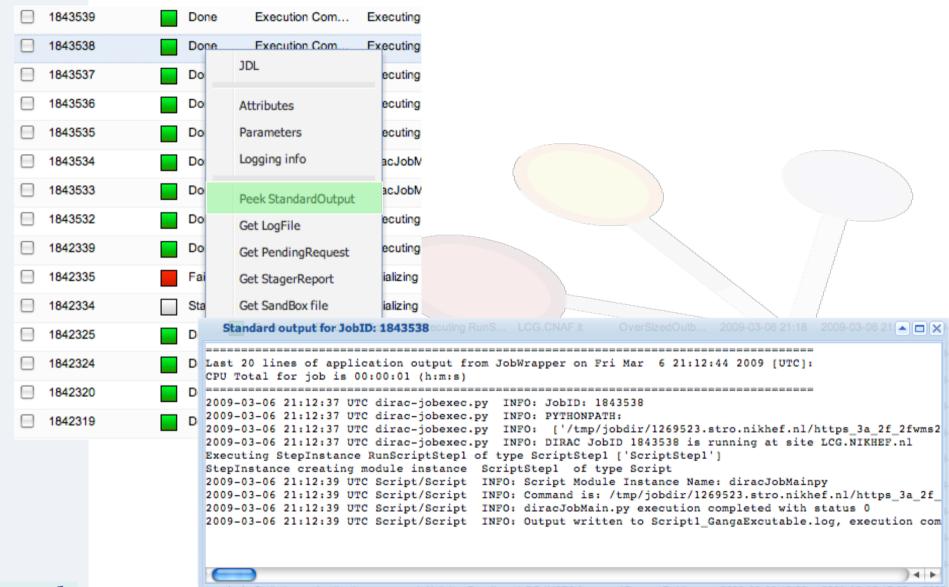
Logging info







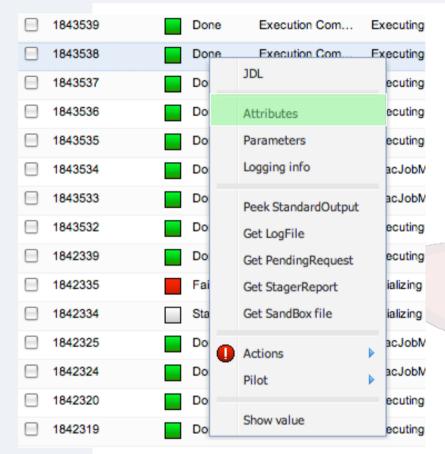
Output peeking







Attributes



Attributes for JobID: 1843	Value
SystemPriority	0
ApplicationNumStatus	0
EndExecTime	2009-03-06 21:12:44
JobID	1843538
VerifiedFlag	True
RetrievedFlag	True
Status	Done
StartExecTime	None
RescheduleCounter	0
JobSplitType	Single
MinorStatus	Execution Complete
ApplicationStatus	Executing RunScriptStep1
SubmissionTime	2009-03-06 21:12:21
JobType	user
MasterJobID	0
KilledFlag	False
RescheduleTime	None
DIRACSetup	LHCb-Production
FailedFlag	False
CPUTime	0.0
OwnerDN	/DC=ch/DC=cern/OU=Organic Unit
JobGroup	Ihob
JobName	OutputData{Ganga_Executable
AccountedFlag	False
LastUpdateTime	2009-03-06 21:14:18





Parameters

