LHCb Production System

Federico Stagni, on behalf of the LHCbDirac team

Overview

- What's this
- DIRAC workflows
- LHCbDirac Bookkeeping
- Application Steps and Production Requests
- Transformation System
  - LHCb extensions
- Production system in action
- Issues, Limitations, ToDo

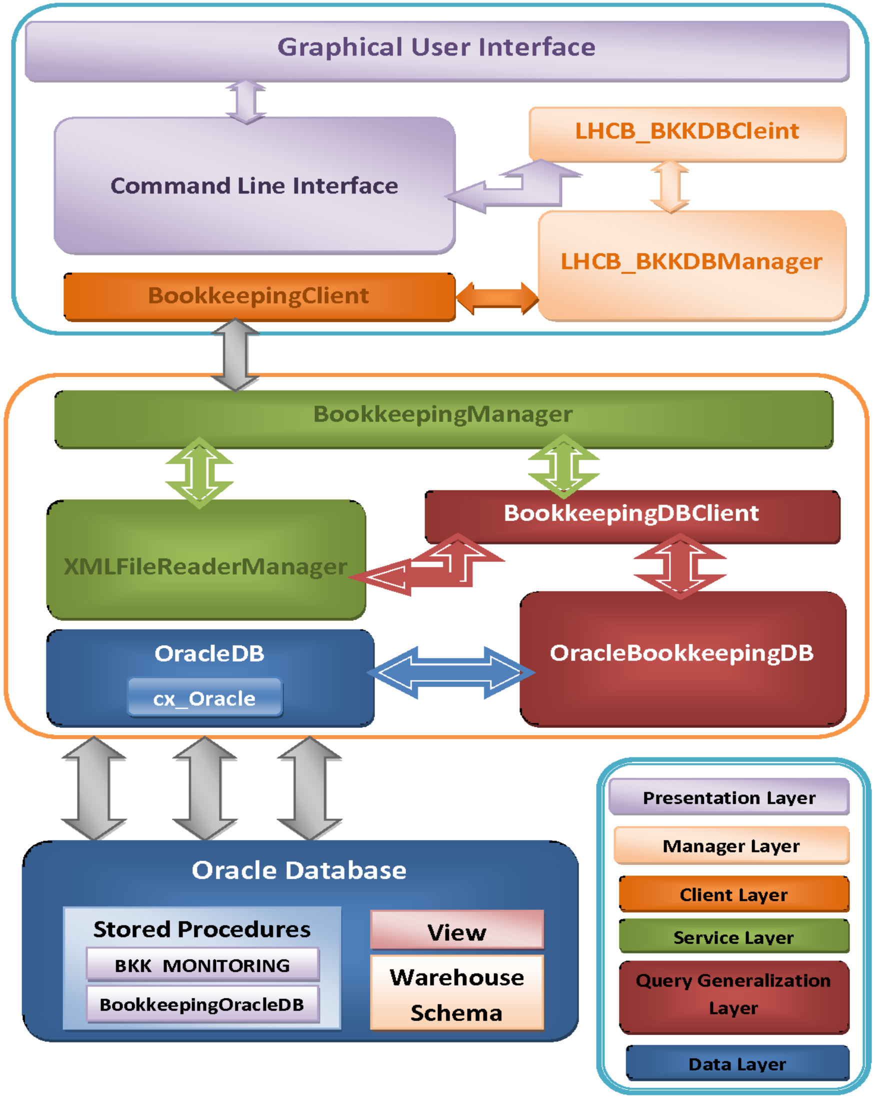
What's this

- The LHCb “Production system” is a large part of LHCbDirac
- Handles different types of productions:
  - MonteCarlo Simulations (MC)
  - Reconstruction (and Reprocessing)
  - Stripping (Selection of Physics events)
  - Merging
  - Working Groups analysis
- Not only an extension of the Transformation System

Bookkeeping

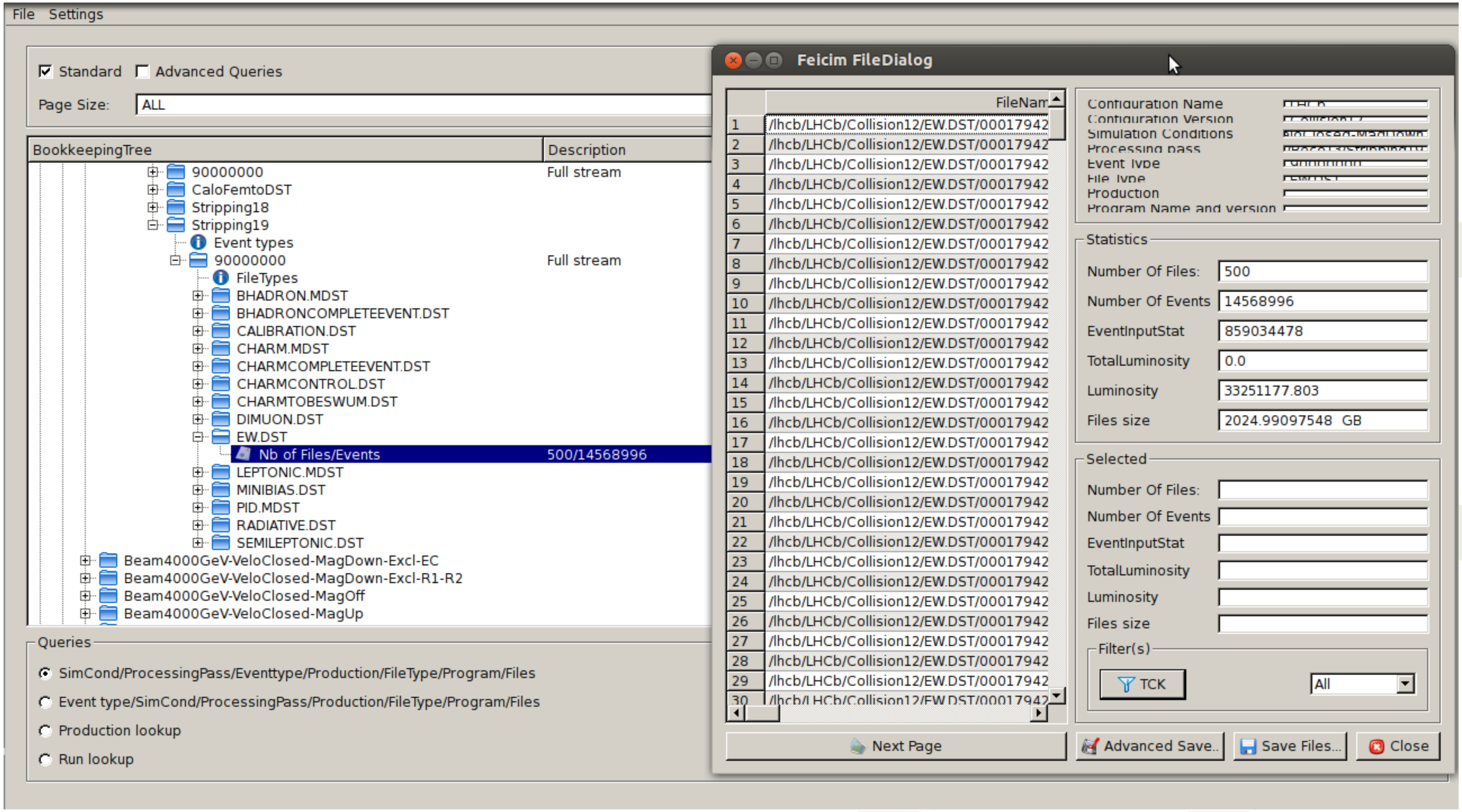
- Data provenance and dataset retrieval
- Not necessarily a tool for distributed computing
- Retrieving datasets for:
  - users (for analysis) and production system
    - Conditions (data taking, simulation)
    - Processing (applications, detector condition parameters)
    - Event type
    - File type
- Integrated in LHCbDIRAC as a “Catalog” (LHCbDIRAC.Resources.Catalog)
- Oracle backend (DIRAC.Core.Utilities.OracleDB)

BKK: design



- Layered design
- Focus on scalability
- One independent GUI, and a page integrated in the web portal

BKK: GUI



☑ Standard ☐ Advanced Queries

Page Size: ALL

BookkeepingTree	Description
90000000	Full stream
CaloFemtoDST	
Stripping18	
Stripping19	
Event types	
90000000	Full stream
FileTypes	
BHADRON.MDST	
BHADRONCOMPLETEEVENT.DST	
CALIBRATION.DST	
CHARM.MDST	
CHARMCOMPLETEEVENT.DST	
CHARMCONTROL.DST	
CHARMTOBESWUM.DST	
DIMUON.DST	
EW.DST	
Nb of Files/Events	500/14568996
LEPTONIC.MDST	
MINIBIAS.DST	
PID.MDST	
RADIATIVE.DST	
SEMILEPTONIC.DST	
Beam4000GeV-VeloClosed-MagDown-Excl-EC	
Beam4000GeV-VeloClosed-MagDown-Excl-R1-R2	
Beam4000GeV-VeloClosed-MagOff	
Beam4000GeV-VeloClosed-MagUp	

Queries

● SimCond/ProcessingPass/Eventtype/Production/FileType/Program/Files

○ Event type/SimCond/ProcessingPass/Production/FileType/Program/Files

○ Production lookup

○ Run lookup

Feicim FileDialog

	FileNam
1	/lhcb/LHCb/Collision12/EW.DST/00017942
2	/lhcb/LHCb/Collision12/EW.DST/00017942
3	/lhcb/LHCb/Collision12/EW.DST/00017942
4	/lhcb/LHCb/Collision12/EW.DST/00017942
5	/lhcb/LHCb/Collision12/EW.DST/00017942
6	/lhcb/LHCb/Collision12/EW.DST/00017942
7	/lhcb/LHCb/Collision12/EW.DST/00017942
8	/lhcb/LHCb/Collision12/EW.DST/00017942
9	/lhcb/LHCb/Collision12/EW.DST/00017942
10	/lhcb/LHCb/Collision12/EW.DST/00017942
11	/lhcb/LHCb/Collision12/EW.DST/00017942
12	/lhcb/LHCb/Collision12/EW.DST/00017942
13	/lhcb/LHCb/Collision12/EW.DST/00017942
14	/lhcb/LHCb/Collision12/EW.DST/00017942
15	/lhcb/LHCb/Collision12/EW.DST/00017942
16	/lhcb/LHCb/Collision12/EW.DST/00017942
17	/lhcb/LHCb/Collision12/EW.DST/00017942
18	/lhcb/LHCb/Collision12/EW.DST/00017942
19	/lhcb/LHCb/Collision12/EW.DST/00017942
20	/lhcb/LHCb/Collision12/EW.DST/00017942
21	/lhcb/LHCb/Collision12/EW.DST/00017942
22	/lhcb/LHCb/Collision12/EW.DST/00017942
23	/lhcb/LHCb/Collision12/EW.DST/00017942
24	/lhcb/LHCb/Collision12/EW.DST/00017942
25	/lhcb/LHCb/Collision12/EW.DST/00017942
26	/lhcb/LHCb/Collision12/EW.DST/00017942
27	/lhcb/LHCb/Collision12/EW.DST/00017942
28	/lhcb/LHCb/Collision12/EW.DST/00017942
29	/lhcb/LHCb/Collision12/EW.DST/00017942
30	/lhcb/LHCb/Collision12/EW.DST/00017942

Configuration Name	LHCb
Configuration Version	Collision12
Simulation Conditions	...loClosed-MagDown
Processing pass	...Reco13/Stripping19...
Event Type	90000000
File Type	EW.DST
Production	
Program Name and version	

Statistics

Number Of Files:	500
Number Of Events	14568996
EventInputStat	859034478
TotalLuminosity	0.0
Luminosity	33251177.803
Files size	2024.99097548 GB

Selected

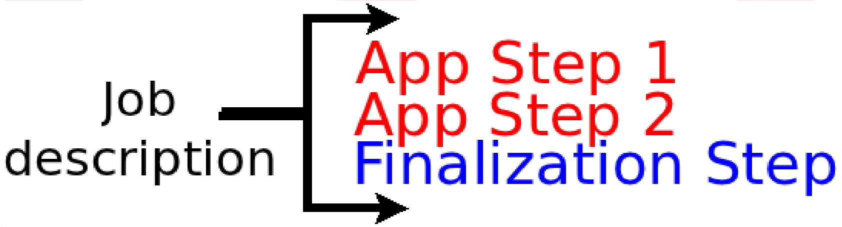
Number Of Files:	
Number Of Events	
EventInputStat	
TotalLuminosity	
Luminosity	
Files size	

Filter(s)

DIRAC Workflows

DIRAC user meeting 2011 (click!)

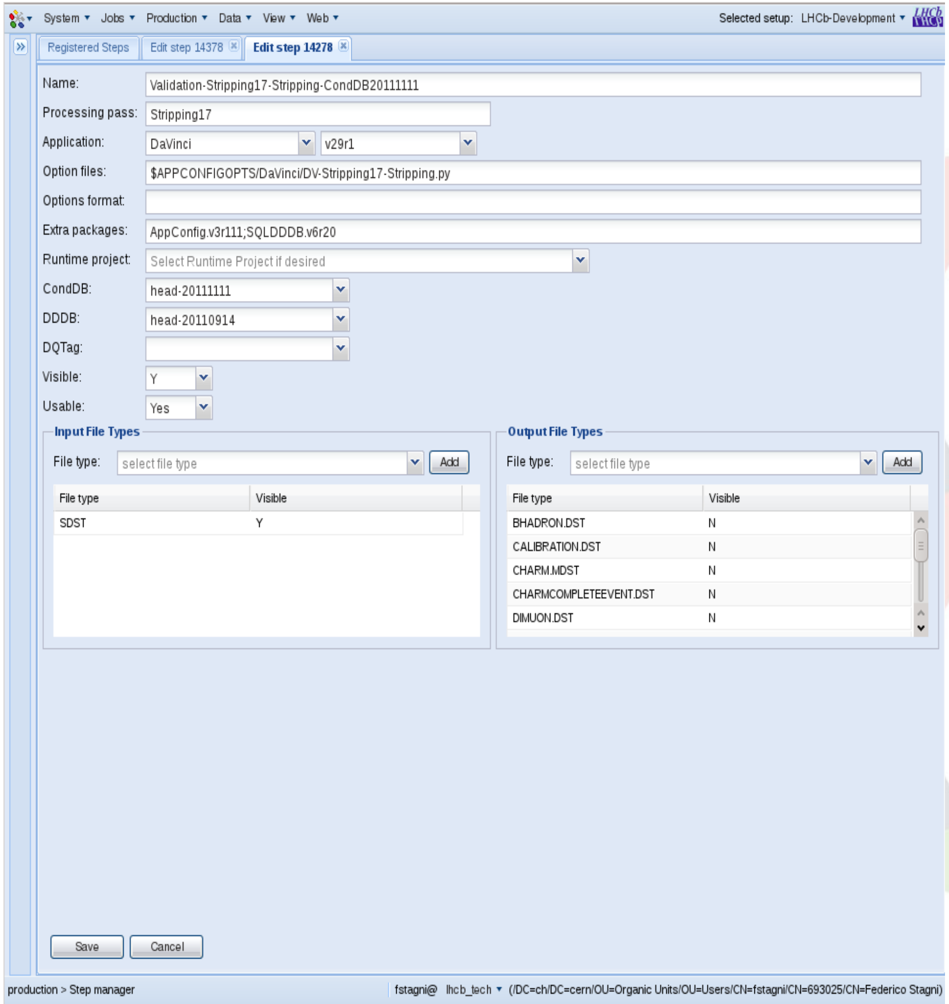
- Xml ↔python dict
- A workflow connects steps together
- dirac-jobexec aWorkflow.xml



Jdl:

- `Executable = "$DIRACROOT/scripts/dirac-jobexec";`
- `Arguments = "jobDescription.xml -o LogLevel=verbose";`

Production Request System



Name:	Validation-Stripping17-Stripping-CondDB20111111
Processing pass:	Stripping17
Application:	DaVinci v29r1
Option files:	$APPCONFIGOPTS/DaVinci/DV-Stripping17-Stripping.py
Options format:	
Extra packages:	AppConfig.v3r111;SQLDDDB.v6r20
Runtime project:	Select Runtime Project if desired
CondDB:	head-20111111
DDDB:	head-20110914
DQTag:	
Visible:	Y
Usable:	Yes

Input File Types

File type: select file type Add

File type	Visible
SDST	Y

Output File Types

File type: select file type Add

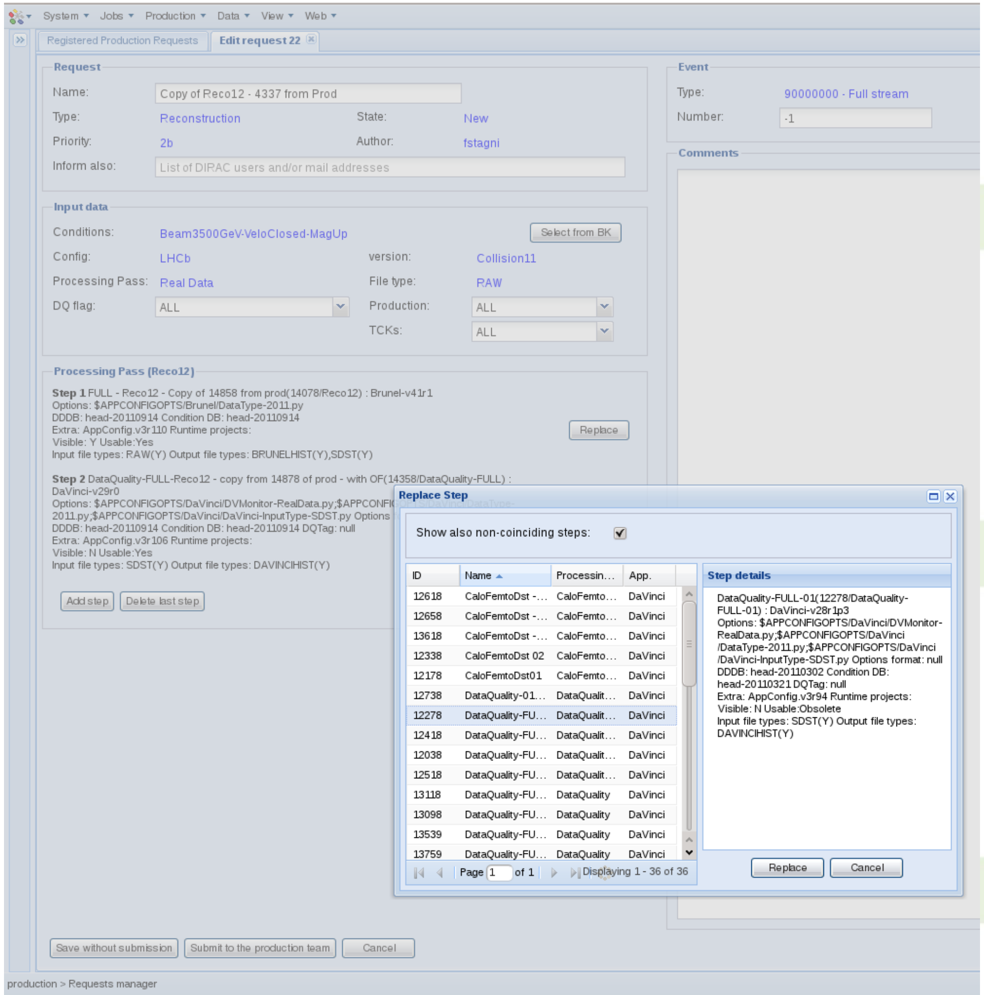
File type	Visible
BHADRON.DST	N
CALIBRATION.DST	N
CHARM.MDST	N
CHARMCOMPLETEEVENT.DST	N
DIMUON.DST	N

production > Step manager

fstagni@ lhcb_tech (/DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=fstagni/CN=693025/CN=Federico Stagni)

- Application Managers defines application steps
- “What to run” to go from X to Y
- LHCb application
- A step “translates” in a workflow application step

Production Request System /2



Request

Name: Copy of Reco12 - 4337 from Prod
Type: Reconstruction
State: New
Priority: 2b
Author: fstagni
Inform also: List of DIRAC users and/or mail addresses

Event

Type: 90000000 - Full stream
Number: -1

Comments

Input data

Conditions: Beam3500GeV-VeloClosed-MagUp
Config: LHCb
version: Collision11
Processing Pass: Real Data
File type: RAW
DQ flag: ALL
Production: ALL
TCKs: ALL

Processing Pass (Reco12)

Step 1 FULL - Reco12 - Copy of 14858 from prod(14078/Reco12) : Brunel-v41r1
Options: $APPCONFIGOPTS/Brunel/DataType-2011.py
DDDB: head-20110914 Condition DB: head-20110914
Extra: AppConfig.v3r110 Runtime projects:
Visible: Y Usable:Yes
Input file types: RAW(Y) Output file types: BRUNELHIST(Y),SDST(Y)

Step 2 DataQuality-FULL-Reco12 - copy from 14878 of prod - with OF(14358/DataQuality-FULL) : DaVinci-v29r0
Options: $APPCONFIGOPTS/DaVinci/DVMonitor-RealData.py;$APPCONFIGOPTS/DaVinci/DataType-2011.py;$APPCONFIGOPTS/DaVinci/DaVinci-InputType-SDST.py Options
DDDB: head-20110914 Condition DB: head-20110914 DQTag: null
Extra: AppConfig.v3r106 Runtime projects:
Visible: N Usable:Yes
Input file types: SDST(Y) Output file types: DAVINCIHIST(Y)

Replace Step

Show also non-coinciding steps: ☑

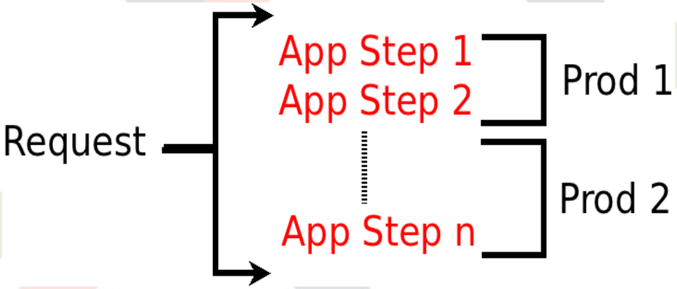
ID	Name ▲	Processin...	App.
12618	CaloFemtoDst -...	CaloFemto...	DaVinci
12658	CaloFemtoDst -...	CaloFemto...	DaVinci
13618	CaloFemtoDst -...	CaloFemto...	DaVinci
12338	CaloFemtoDst 02	CaloFemto...	DaVinci
12178	CaloFemtoDst01	CaloFemto...	DaVinci
12738	DataQuality-01...	DataQualit...	DaVinci
12278	DataQuality-FU...	DataQualit...	DaVinci
12418	DataQuality-FU...	DataQualit...	DaVinci
12038	DataQuality-FU...	DataQualit...	DaVinci
12518	DataQuality-FU...	DataQualit...	DaVinci
13118	DataQuality-FU...	DataQuality	DaVinci
13098	DataQuality-FU...	DataQuality	DaVinci
13539	DataQuality-FU...	DataQuality	DaVinci
13759	DataQuality-FU...	DataQuality	DaVinci

Page 1 of 1 Displaying 1 - 36 of 36

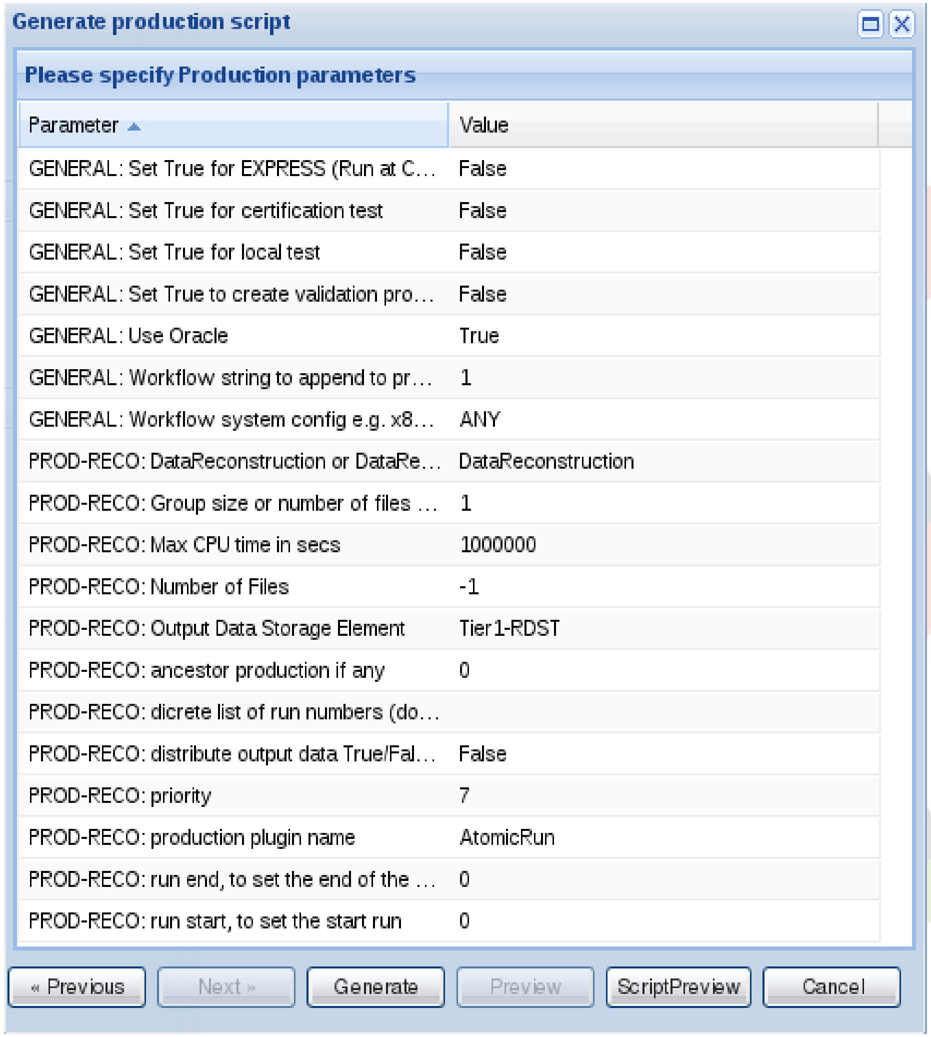
Step details

DataQuality-FULL-01(12278/DataQuality-FULL-01) : DaVinci-v28r1p3
Options: $APPCONFIGOPTS/DaVinci/DVMonitor-RealData.py;$APPCONFIGOPTS/DaVinci/DataType-2011.py;$APPCONFIGOPTS/DaVinci/DaVinci-InputType-SDST.py Options format: null
DDDB: head-20110302 Condition DB: head-20110321 DQTag: null
Extra: AppConfig.v3r94 Runtime projects:
Visible: N Usable:Obsolete
Input file types: SDST(Y) Output file types: DAVINCIHIST(Y)

- Steps are combined in production requests (e.g. MC, or Reconstruction)



Production Request System /3



Generate production script

Please specify Production parameters

Parameter ▲	Value
GENERAL: Set True for EXPRESS (Run at C...	False
GENERAL: Set True for certification test	False
GENERAL: Set True for local test	False
GENERAL: Set True to create validation pro...	False
GENERAL: Use Oracle	True
GENERAL: Workflow string to append to pr...	1
GENERAL: Workflow system config e.g. x8...	ANY
PROD-RECO: DataReconstruction or DataRe...	DataReconstruction
PROD-RECO: Group size or number of files ...	1
PROD-RECO: Max CPU time in secs	1000000
PROD-RECO: Number of Files	-1
PROD-RECO: Output Data Storage Element	Tier1-RDST
PROD-RECO: ancestor production if any	0
PROD-RECO: dicrete list of run numbers (do...	
PROD-RECO: distribute output data True/Fal...	False
PROD-RECO: priority	7
PROD-RECO: production plugin name	AtomicRun
PROD-RECO: run end, to set the end of the ...	0
PROD-RECO: run start, to set the start run	0

- Production requests are submitted using production templates
- e.g.: priority, which plugin, where the outputs are stored, DIRAC CPU, etc.
- Each production is created using the Production API

LHCbDirac TS

- Extension of the DIRAC TS, mostly for interacting with the BKK
  - DB:
    - Physics RUNs information
    - BKK queries (supersedes TransformationInputDataQuery)
  - Service and clients extended for the DB extension

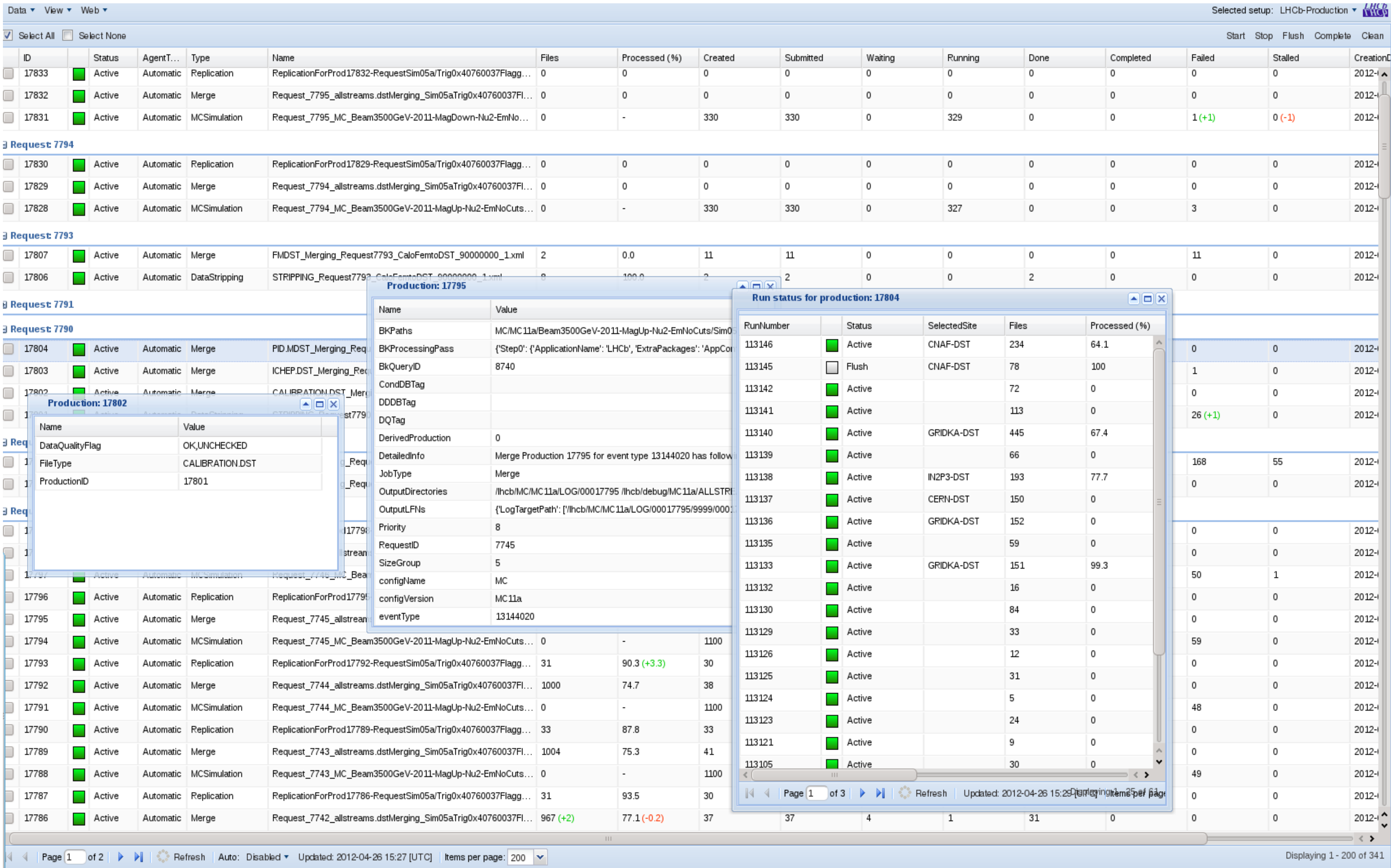
LHCbDirac TS /2

- Agents
  - BookkeepingWatchAgent
    - Looks for BKK queries, and fills the TransformationFiles table
    - Threaded, uses pickle file for caching
  - DataRecoveryAgent
    - Resets input files in “Unused” status, in case the jobs failed
    - A counter is kept, with a maximum of re-trials
  - Extensions for cleaning, and closing productions

LHCbDirac TS /3

- Plugins (LHCbDIRAC.TransformationSystem.Agent.TransformationPlugins)
  - Many LHCb plugins coded
    - e.g. ByRun, AtomicRun, with flushing...
    - This is where you want to extend
- TaskManager
  - Extended to handle the inputs the LHCb way

LHCb Transformation Monitor Web



Selected setup: LHCb-Production ▾

☑ Select All ☐ Select None

	ID		Status	AgentT…	Type	Name	Files	Processed (%)	Created	Submitted	Waiting	Running	Done	Completed	Failed	Stalled	CreationD
☐	17833	🟩	Active	Automatic	Replication	ReplicationForProd17832-RequestSim05a/Trig0x40760037Flagg…	0	0	0	0	0	0	0	0	0	0	2012-
☐	17832	🟩	Active	Automatic	Merge	Request_7795_allstreams.dstMerging_Sim05aTrig0x40760037Fl…	0	0	0	0	0	0	0	0	0	0	2012-
☐	17831	🟩	Active	Automatic	MCSimulation	Request_7795_MC_Beam3500GeV-2011-MagDown-Nu2-EmNo…	0	-	330	330	0	329	0	0	1 (+1)	0 (-1)	2012-
⊟ Request: 7794																	
☐	17830	🟩	Active	Automatic	Replication	ReplicationForProd17829-RequestSim05a/Trig0x40760037Flagg…	0	0	0	0	0	0	0	0	0	0	2012-
☐	17829	🟩	Active	Automatic	Merge	Request_7794_allstreams.dstMerging_Sim05aTrig0x40760037Fl…	0	0	0	0	0	0	0	0	0	0	2012-
☐	17828	🟩	Active	Automatic	MCSimulation	Request_7794_MC_Beam3500GeV-2011-MagUp-Nu2-EmNoCuts…	0	-	330	330	0	327	0	0	3	0	2012-
⊟ Request: 7793																	
☐	17807	🟩	Active	Automatic	Merge	FMDST_Merging_Request7793_CaloFemtoDST_90000000_1.xml	2	0.0	11	11	0	0	0	0	11	0	2012-
☐	17806	🟩	Active	Automatic	DataStripping	STRIPPING_Request7793_CaloFemtoDST_90000000_1.xml	8	100.0	2	2	0	0	2	0	0	0	2012-
⊟ Request: 7791																	
⊟ Request: 7790																	
☐	17804	🟩	Active	Automatic	Merge	PID.MDST_Merging_Requ									0	0	2012-
☐	17803	🟩	Active	Automatic	Merge	ICHEP.DST_Merging_Req									1	0	2012-
☐	17802	🟩	Active	Automatic	Merge	CALIBRATION.DST_Merg									0	0	2012-
☐	17801	🟩	Active	Automatic	DataStripping	STRIPPING_Request7790									26 (+1)	0	2012-
⊟ Req																	
☐	1					_Requ									168	55	2012-
☐	1					_Requ									0	0	2012-
⊟ Req																	
☐	1					d17798									0	0	2012-
☐	1					lstream									0	0	2012-
☐	17797	🟩	Active	Automatic	MCSimulation	Request_7746_MC_Bea									50	1	2012-
☐	17796	🟩	Active	Automatic	Replication	ReplicationForProd17795									0	0	2012-
☐	17795	🟩	Active	Automatic	Merge	Request_7745_allstream									0	0	2012-
☐	17794	🟩	Active	Automatic	MCSimulation	Request_7745_MC_Beam3500GeV-2011-MagUp-Nu2-EmNoCuts…	0	-	1100						59	0	2012-
☐	17793	🟩	Active	Automatic	Replication	ReplicationForProd17792-RequestSim05a/Trig0x40760037Flagg…	31	90.3 (+3.3)	30						0	0	2012-
☐	17792	🟩	Active	Automatic	Merge	Request_7744_allstreams.dstMerging_Sim05aTrig0x40760037Fl…	1000	74.7	38						0	0	2012-
☐	17791	🟩	Active	Automatic	MCSimulation	Request_7744_MC_Beam3500GeV-2011-MagUp-Nu2-EmNoCuts…	0	-	1100						48	0	2012-
☐	17790	🟩	Active	Automatic	Replication	ReplicationForProd17789-RequestSim05a/Trig0x40760037Flagg…	33	87.8	33						0	0	2012-
☐	17789	🟩	Active	Automatic	Merge	Request_7743_allstreams.dstMerging_Sim05aTrig0x40760037Fl…	1004	75.3	41						0	0	2012-
☐	17788	🟩	Active	Automatic	MCSimulation	Request_7743_MC_Beam3500GeV-2011-MagUp-Nu2-EmNoCuts…	0	-	1100						49	0	2012-
☐	17787	🟩	Active	Automatic	Replication	ReplicationForProd17786-RequestSim05a/Trig0x40760037Flagg…	31	93.5	30						0	0	2012-
☐	17786	🟩	Active	Automatic	Merge	Request_7742_allstreams.dstMerging_Sim05aTrig0x40760037Fl…	967 (+2)	77.1 (-0.2)	37	37	4	1	31	0	0	0	2012-

Page 1 of 2 | Refresh | Auto: Disabled ▾ | Updated: 2012-04-26 15:27 [UTC] | Items per page: 200 ▾ | Displaying 1 - 200 of 341

Production: 17802

Name	Value
DataQualityFlag	OK,UNCHECKED
FileType	CALIBRATION.DST
ProductionID	17801

Production: 17795

Name	Value
BKPaths	MC/MC11a/Beam3500GeV-2011-MagUp-Nu2-EmNoCuts/Sim0
BKProcessingPass	{'Step0': {'ApplicationName': 'LHCb', 'ExtraPackages': 'AppCo
BkQueryID	8740
CondDBTag	
DDDBTag	
DQTag	
DerivedProduction	0
DetailedInfo	Merge Production 17795 for event type 13144020 has follow
JobType	Merge
OutputDirectories	/lhcb/MC/MC11a/LOG/00017795 /lhcb/debug/MC11a/ALLSTRE
OutputLFNs	{'LogTargetPath': ['/lhcb/MC/MC11a/LOG/00017795/9999/0001
Priority	8
RequestID	7745
SizeGroup	5
configName	MC
configVersion	MC11a
eventType	13144020

Run status for production: 17804

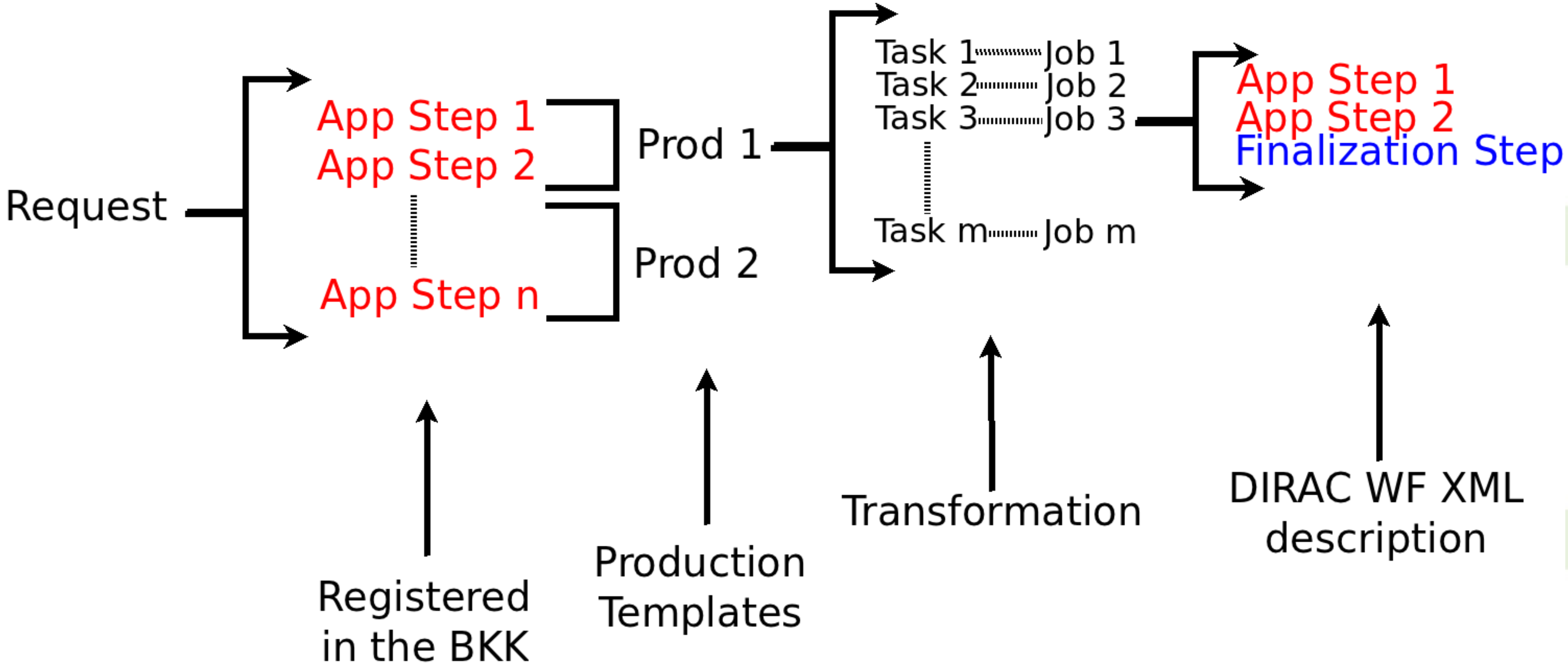
RunNumber		Status	SelectedSite	Files	Processed (%)
113146	🟩	Active	CNAF-DST	234	64.1
113145	☐	Flush	CNAF-DST	78	100
113142	🟩	Active		72	0
113141	🟩	Active		113	0
113140	🟩	Active	GRIDKA-DST	445	67.4
113139	🟩	Active		66	0
113138	🟩	Active	IN2P3-DST	193	77.7
113137	🟩	Active	CERN-DST	150	0
113136	🟩	Active	GRIDKA-DST	152	0
113135	🟩	Active		59	0
113133	🟩	Active	GRIDKA-DST	151	99.3
113132	🟩	Active		16	0
113130	🟩	Active		84	0
113129	🟩	Active		33	0
113126	🟩	Active		12	0
113125	🟩	Active		31	0
113124	🟩	Active		5	0
113123	🟩	Active		24	0
113121	🟩	Active		9	0
113105	🟩	Active		30	0

Page 1 of 3 | Refresh | Updated: 2012-04-26 15:29 [UTC] Displaying 1 - 25 of 61 Items per page

Production API

- Expose functionalities to connect together TS, BKK and Production Request System
- Use LHCbJob.py (extension of DIRAC.Interfaces.API.Job.py) to create a DIRAC workflow, whose xml is uploaded to the Transformation DB
- python modules are run within the workflow, grouped within steps. Application steps and Finalization steps are present

Putting concepts together



Request

App Step 1
App Step 2
App Step n

Prod 1
Prod 2

Task 1 Job 1
Task 2 Job 2
Task 3 Job 3
Task m Job m

App Step 1
App Step 2
Finalization Step

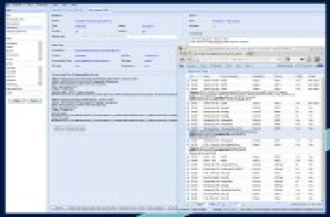
Registered in the BKK

Production Templates

Transformation

DIRAC WF XML description

Distributed Computing Coordinator



PPG/OPG

Physics and Operations Planning Group, follow activities from high level

Application Managers

Core software, application releases

Production Managers

Run, control and coordinate most of the production activities

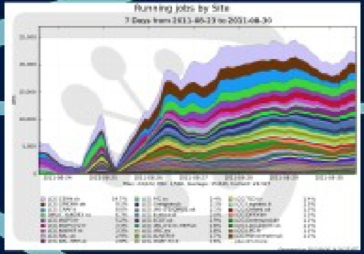
Productions Monitor

DQ

Data Quality shifter

Computing shifter

Mainly look after productions activities



Running jobs by Site

Data Manager

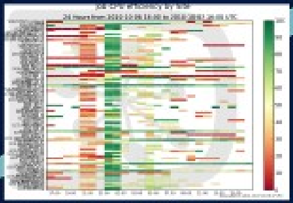
Data issues, distribution, removal

GEOC

Grid Expert On Call: referes to WLCG, run daily operations meeting, 1st contact for shifters

Resources Manager

Plans for resources



Job CPU efficiency by Site

Sites testing and monitoring

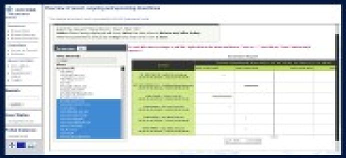
DAST

Distributed Analysis Support shifter

Dirac Experts

Contact for all DIRAC-related issues

DIRAC Config



Tier1 contacts

A contact person for each Tier 1

Sites Manager

Mostly handles T2s

The system in Action

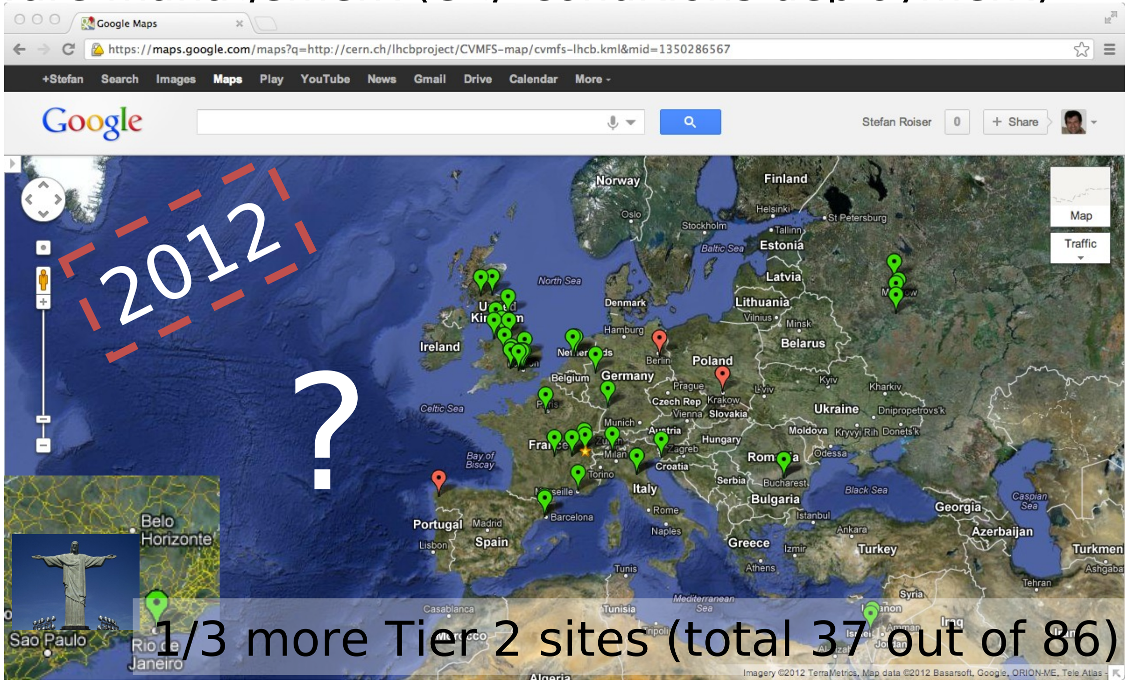
Taking the LHCb Reprocessing 2012 as example

The Plan (source: Stefan Roiser)

- Reprocessing from mid Sep until Xmas break
  - In this time process all 2012 data up to Sep TS
  - Using a simplified workflow, only 1 output file of reco
- Operations on T1 sites + 20 % attached T2s (as in 2011)
- First pass processing at CERN + 2 T2s

Let’s think BIG (source: Stefan Roiser)

- Can we expand this model by using more Tier2s?
  - Preferably choose CVFMS sites
  - Because of easier software management (e.g. conditions deployment)
  - No preference on “site power”
  - Questions
  - Will the T1 storage sustain the load?
  - Network congestions?



2012

?

1/3 more Tier 2 sites (total 37 out of 86)

CVMFS site Non-CVMFS site

How?

- We “attached” T2s to T1s:
  - (output): /Resources/Sites/<yourSite>/AssociatedSEs
  - (input): /Operations/SiteLocalSEMapping

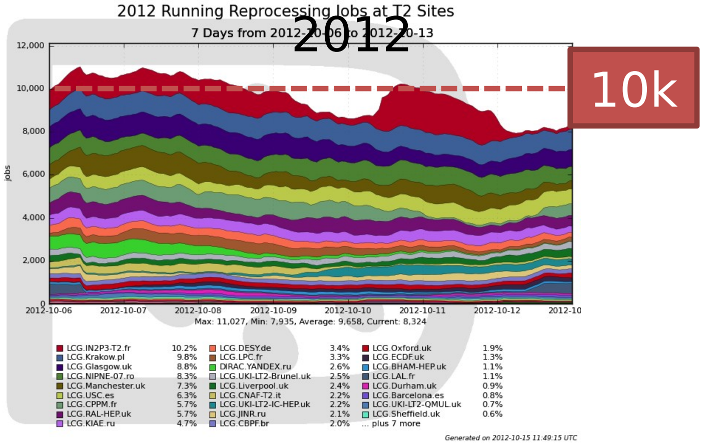
(some mix between Resources and Operations)

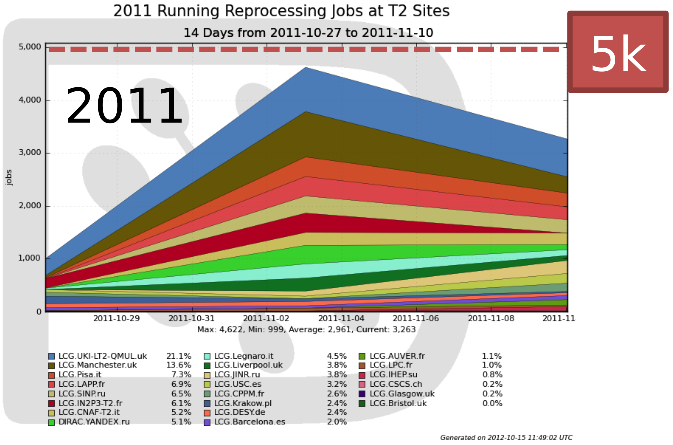
For scheduling we intervene on:

- /Operations/JobScheduling/MatchingDelay
- /Operations/JobScheduling/RunningDelay

Current Status (source: Stefan Roiser)

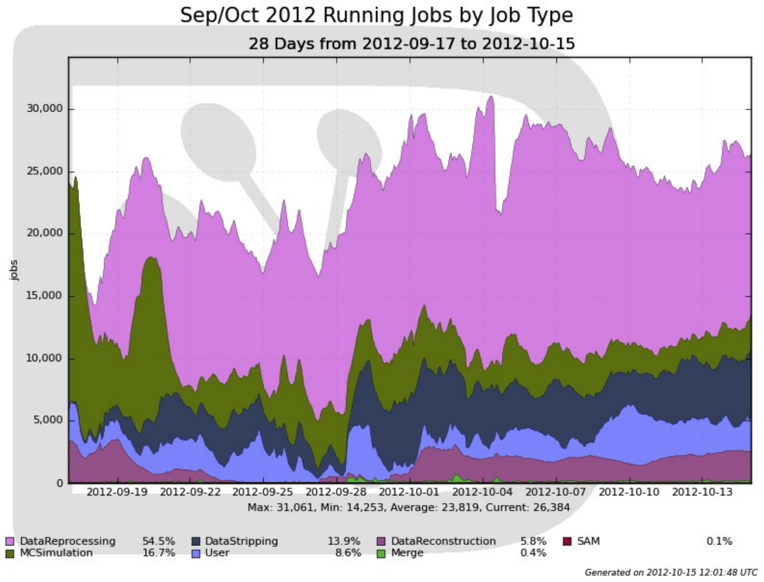
Running reconstruction jobs at T2s
during 2011 and 2012 reprocessing





All grid activities since start of the reprocessing campaign

- Running at “full steam”
- Enough room for user jobs
- Less Simulation jobs
- Peaks of 30k running jobs (usual 25k)



Problems (source: Stefan Roiser)

- It’s not all singing and dancing
  - Mostly scaling issues both at (LHCb)Dirac and sites
  - Because of high load, comparable to data post LS1
- Data moving is >just< following at some sites
  - Staging of tape data, moving between storages
- Several adjustments to Dirac to handle huge productions
  - Several module changes needed to speedup for handling large amount of jobs in a short period
  - Scaling also by multiplying data handling agents
- Production priorities didn't really seem to work!

Handling the scalability issues

- e.g. productions with ~400k input files
- 2 TransformationManager and 2 BookkeepingManager (services) running in parallel on different machines
- The TransformationClient splits large queries in peaces.
- Large use of caching in TransformationAgent and BookkeepingWatchAgent

From Barcelona 2011: shopping list! (what's remaining?)

- Core, Framework
  - Use MySQL Transactions? SQLAlchemy?
- Request System
  - Complete review started... maybe using the executors framework?
  - Few steps made for the monitoring
- Scripts
  - Still feeling like if there is a little mess... review?
- Testing and Certification
  - Made some steps in LHCb, adoption in DIRAC (jenkins)?
  - Documentation, documentation, documentation...

Questions

Setups

- DIRAC setups:
  - Development, Certification, Production
- LHCbDirac Production setup:
  - 1 machine for agents
  - 1 machine for security services
  - 2 machines for services
  - 1 for DIRAC SEs (logs, and Sandboxes)
  - 2 web servers (1 at CERN, 1 in Barcelona)
  - 6 MySQL DB machines (3 for accounting)
  - An Oracle DB maintained by CERN IT
  - 1 machine for special needs