# Transformation System

Federico Stagni

# What's this?

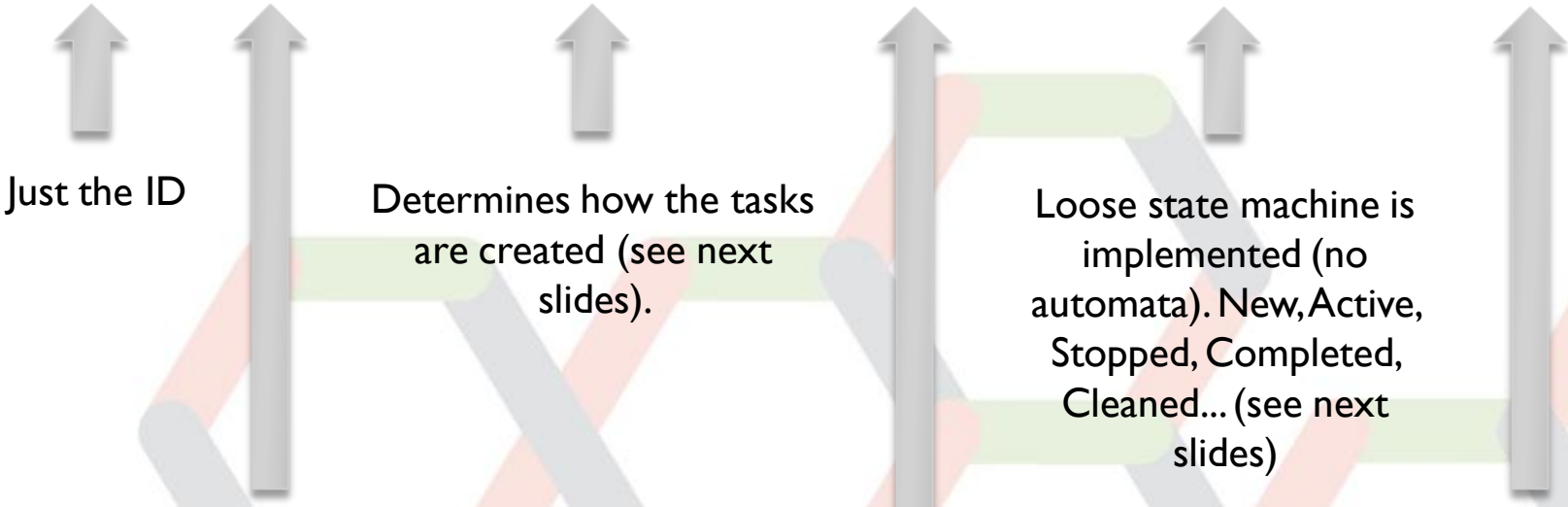
- A DIRAC system, as usually comprising:
  - MySQL tables
  - DIRAC Services
  - DIRAC Agents
  - Clients
  - Scripts
  - A set of **Plugins**
- Used for "repetitive" work
  - Can be used for different workflows:
    - Productions (MC, reconstructions, etc...)
    - Data management (Data replication, deletion, etc...)
- In LHCb we built our Production system on it (see presentation tomorrow morning)

# How it works

1. A new **transformation** is created (typed)
2. There are (or there aren't) **input** files (e.g: MC, reconstruction, replication)
3. Each transformation has a **plugin** (e.g.: ByReplica, BySize, Broadcast)
4. **Tasks** are created based on the specified plugin (e.g.: task including a number of inputs, or )
5. Tasks are treated by an **external system** (e.g: WMS, DMS)

# 1: Transformation

| ID | Type | Plugin | AgentType | Status | GroupSize |
|---|---|---|---|---|---|



Just the ID

Whatever transformation type you want. E.g. :MonteCarlo, DataReconstruction, replication

Determines how the tasks are created (see next slides).

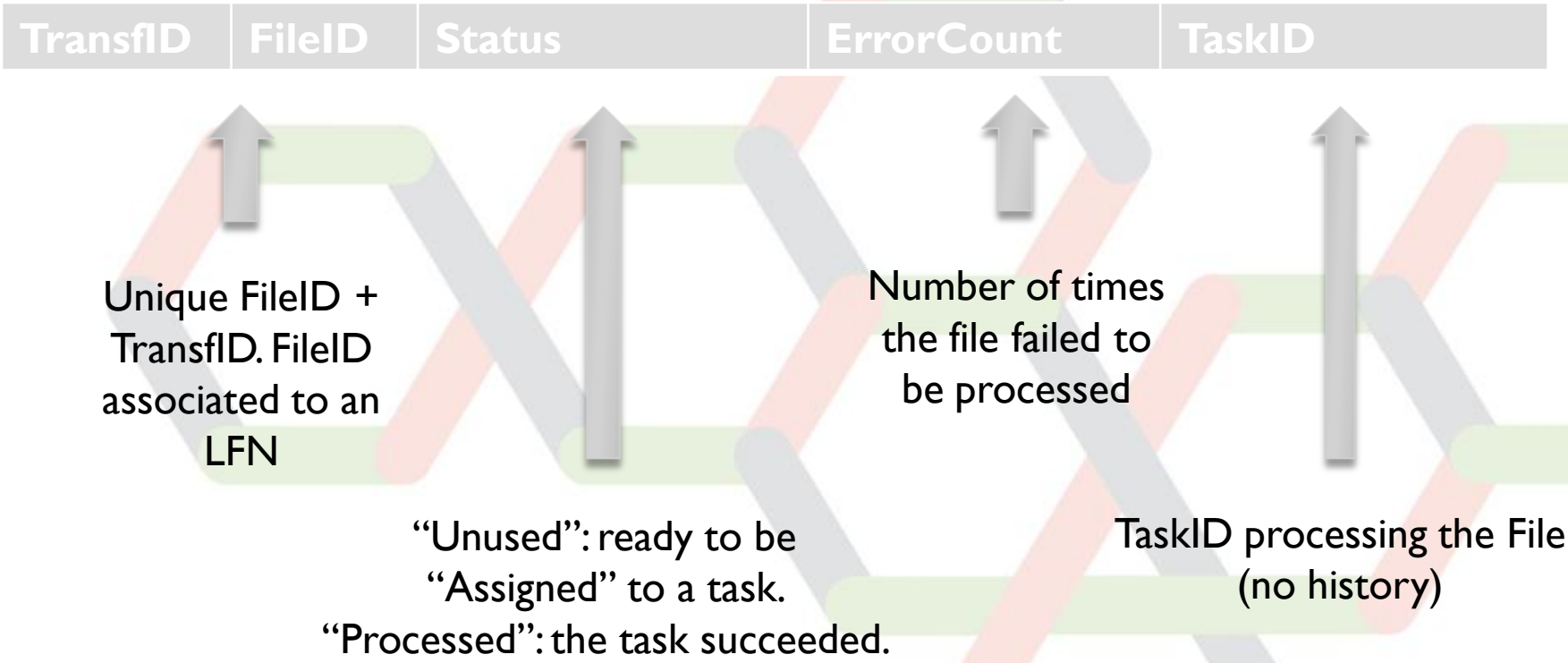
Choose between "manual" and "automatic".

Loose state machine is implemented (no automata). New, Active, Stopped, Completed, Cleaned... (see next slides)

Double meaning: size of the output (for merging transformations) or number of files in input per job

# 2: Input files

- Each Transformation can have a number of input files
  - E.g.: Data Reconstruction transformations



# 3: Plugins

- Standard():
  - Group files by replicas (tasks create based on the file location)
- BySize():
  - Group files until they reach a certain size (Input size in Gb)
- ByShare()
  - Group files given the share (specified in the CS) and location

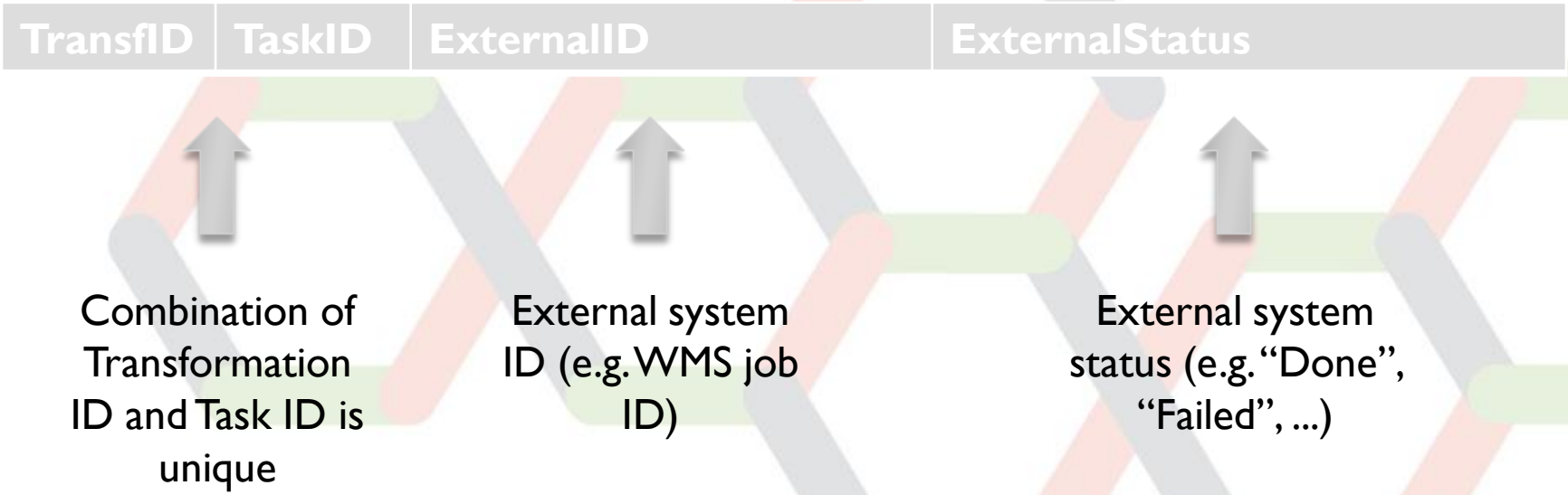
For replication:

- Broadcast()
  - Takes files at the source SE and brodcast to a given number of locations

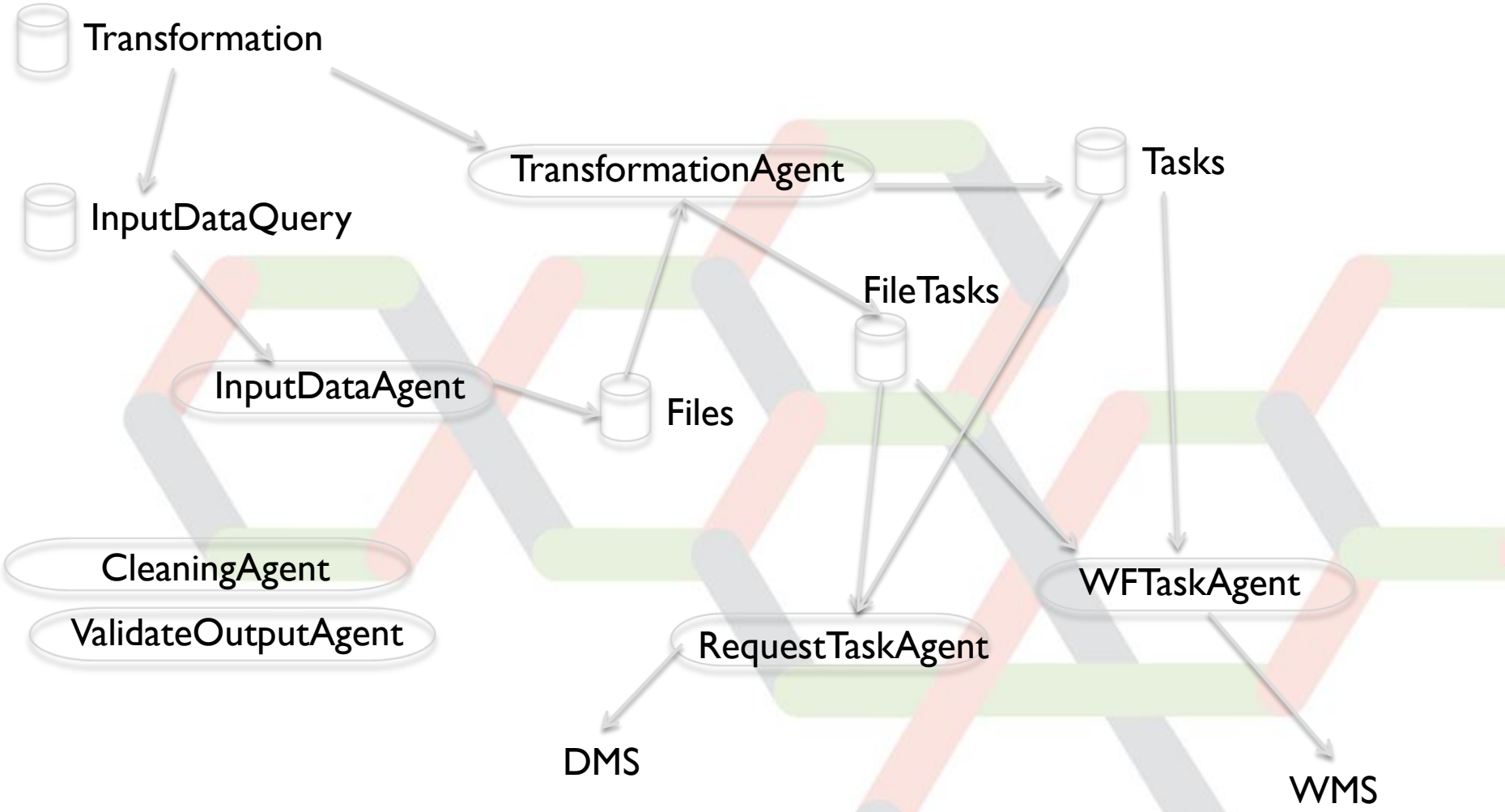
# 4: Tasks, 5: External system

Tasks are treated by an external system:

- WMS (all but replication and removal transformations)
- DMS (replication and removal transformations)



# Agents


Transformation
TransformationAgent
Tasks
InputDataQuery
FileTasks
InputDataAgent
Files
CleaningAgent
WFTaskAgent
ValidateOutputAgent
RequestTaskAgent
DMS
WMS

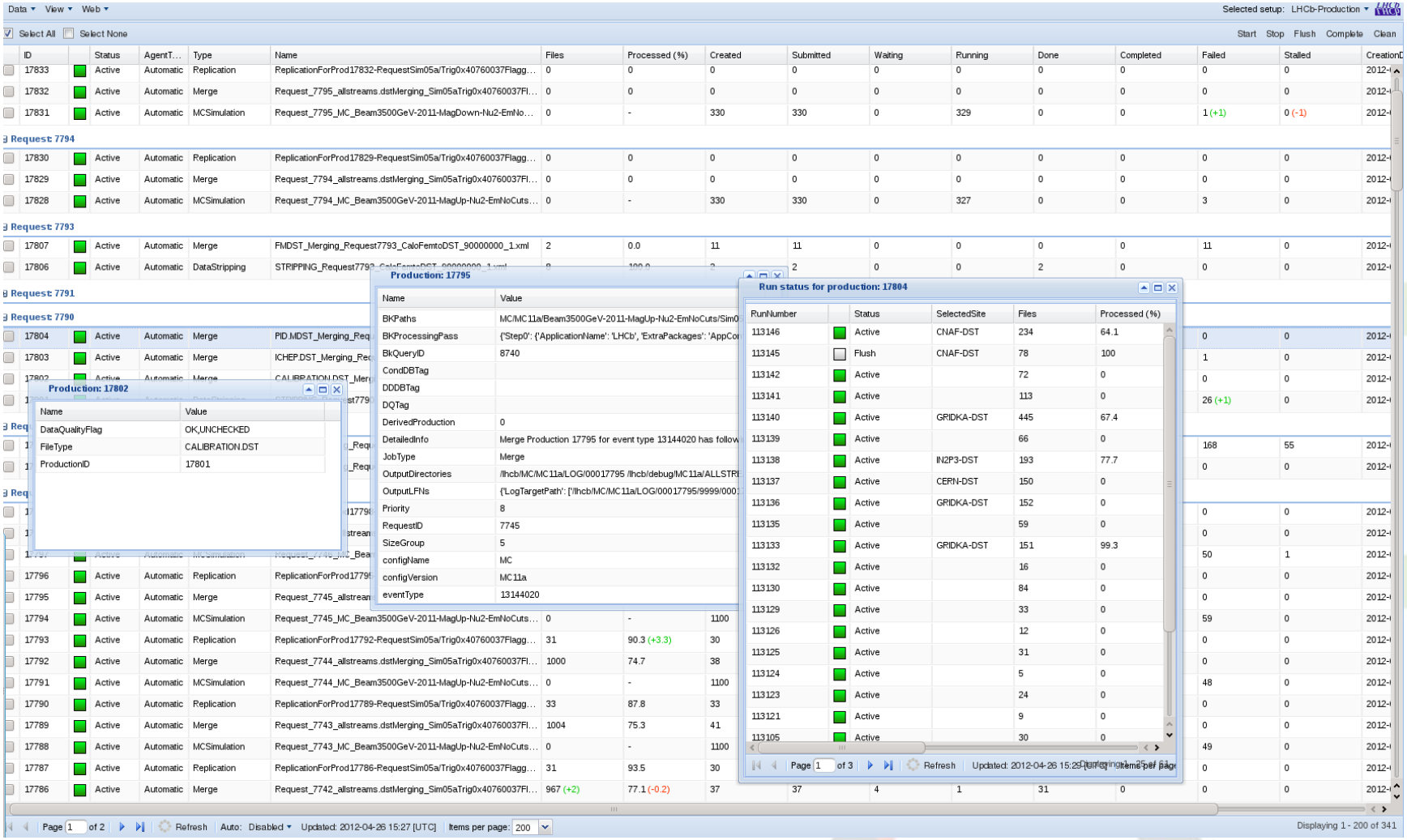
# Agents /2

- The TransformationAgent creates the tasks
  - For all the transformation types, or the ones configured
  - Version in v6r5 not really scalable (good for tests)
    - Version in v6r6 is a porting from LHCbDirac one
- The WorkflowTaskAgent submits production tasks to the WMS
- The RequestTaskAgent submits data management tasks to the DMS (using the RMS)

# Clients

- TaskManager.py:
  - Contains classes wrapping the logic of how to “transform” a task in a job/request
- TransformationClient.py:
  - Main client to the service/DB
    - Recently adapted for “big” productions
- Transformation.py
  - For Data Management transformations
- TransformationCLI.py:
  - A CLI for manipulations of the DB

# Web


Data View Web
Selected setup: LHCb-Production
Select All Select None
Start Stop Flush Complete Clean
Request 7794
Request 7793
Request 7791
Request 7790
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/ALLSTR
OutputLFNs {'LogTargetPath': ['/lhcb/MC/MC11a/LOG/00017795/9999/000
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
Page 1 of 2 Refresh Auto: Disabled Updated: 2012-04-26 15:27 [UTC] Items per page: 200
Displaying 1 - 200 of 341

# What you might want to extend

- For Data Management
  - Nowhere
- For pure MC:
  - Nowhere
- For DataManipulation productions
  - Most probably some parts
    - Especially if you have an external Catalog (e.g. Bookkeeping)
      - E.g.: where to get the list of input files?
      - See LHCb presentation tomorrow

# Questions