Integration of the ATLAS Production System with the Distributed Data Management on the EGEE Grid Infrastructure

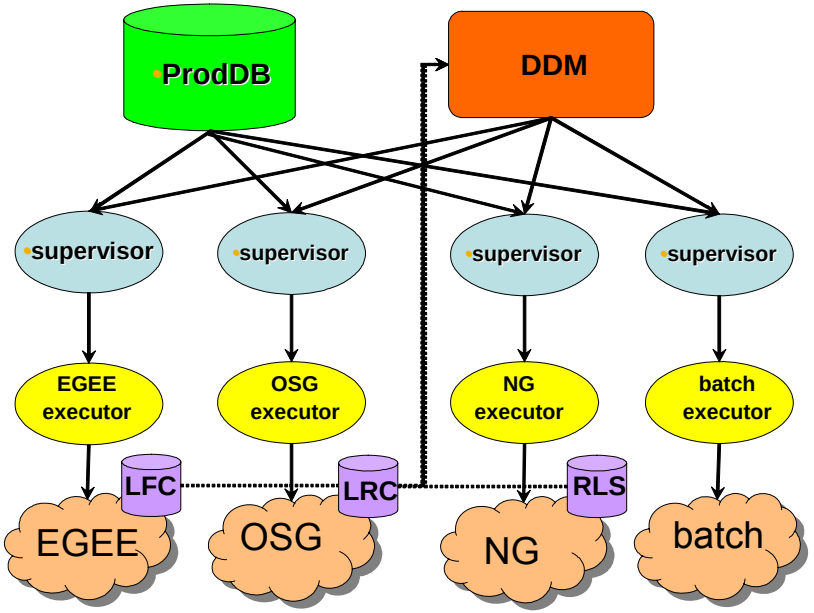
Simone Campana, CERN/INFN
David Rebatto, Rodney Walker, Miguel Branco

EGEE Conference, Geneva 2006

Information Society

The ATLAS production system

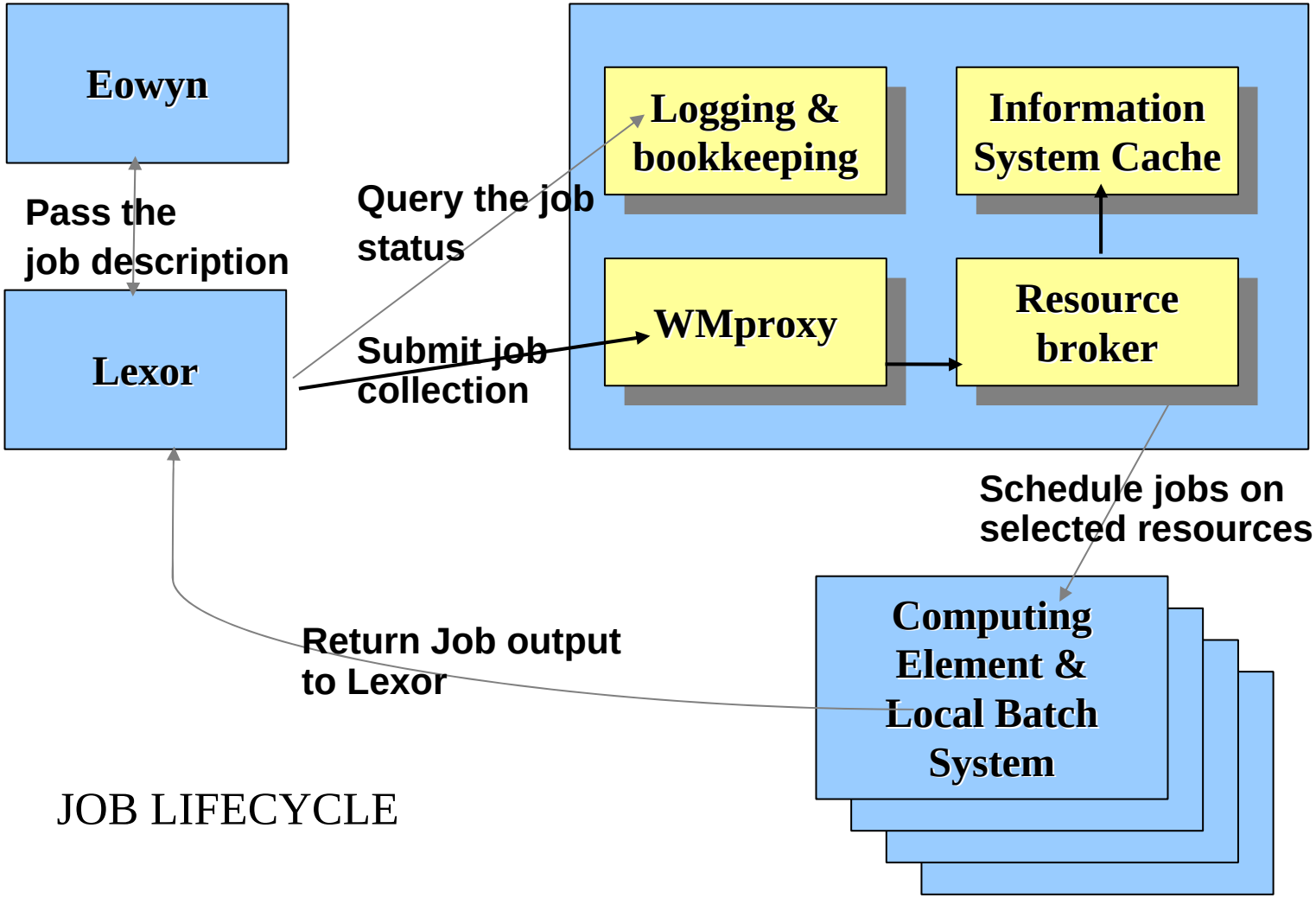
- **An ATLAS central database holds Grid-neutral information about jobs.**
  - ✓ Jobs can be of different type
    - Event generation
    - Simulatio+Digitization
    - Reconstruction
  - ✓ Each type of the chain produces output for the subsequent one



- **A "supervisor" agent**
  - ✓ distributes jobs to Grid-specific agents called "executors"
  - ✓ follows up their status, validates them in case of success or flags them for resubmission.

- **The executors offer an interface to the underlying Grid middleware.**
  - ✓ Two different Executors can submit jobs to the EGEE infrastructure
    - Lexor
    - CondorG

Submitting jobs via the gLite WMS

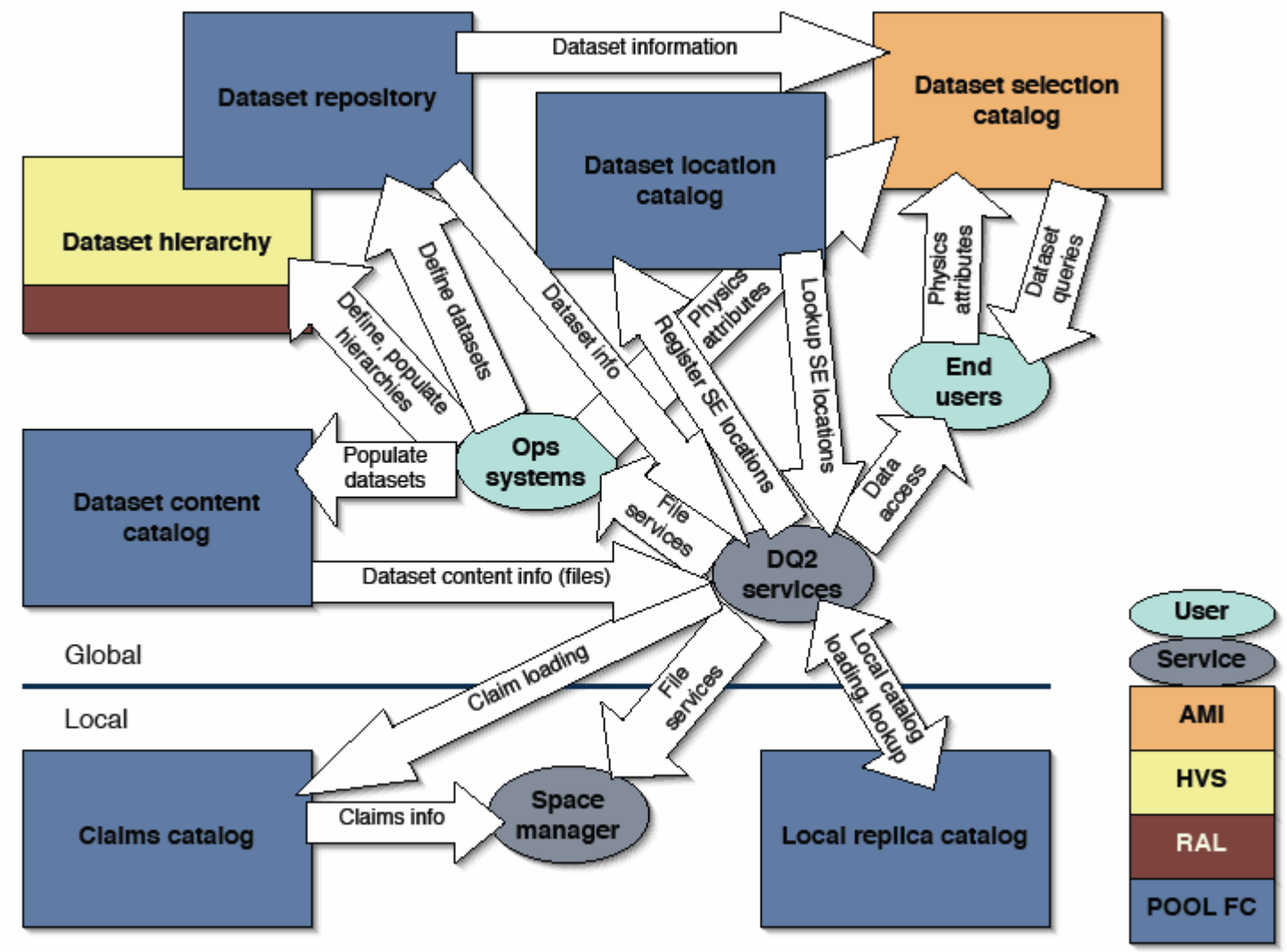


JOB LIFECYCLE

The Common Executor

- **The two executors differ only in Job submission and handling**
  - ✓ Every other aspect is common
  - ✓ Lexor and CondorG are in fact classes of a Common Executor
- **The Common Executor includes**
  - ✓ The wrapper around the ATLAS application
  - ✓ Staging in of input data
  - ✓ Staging out of output data
- **StageIn/Out of data from/to the WN to/from a SE is the only Data Management operation at runtime**
  - ✓ Every other operation is asynchronous and is performed by the Distributed Data Management system

The Distributed Data Management



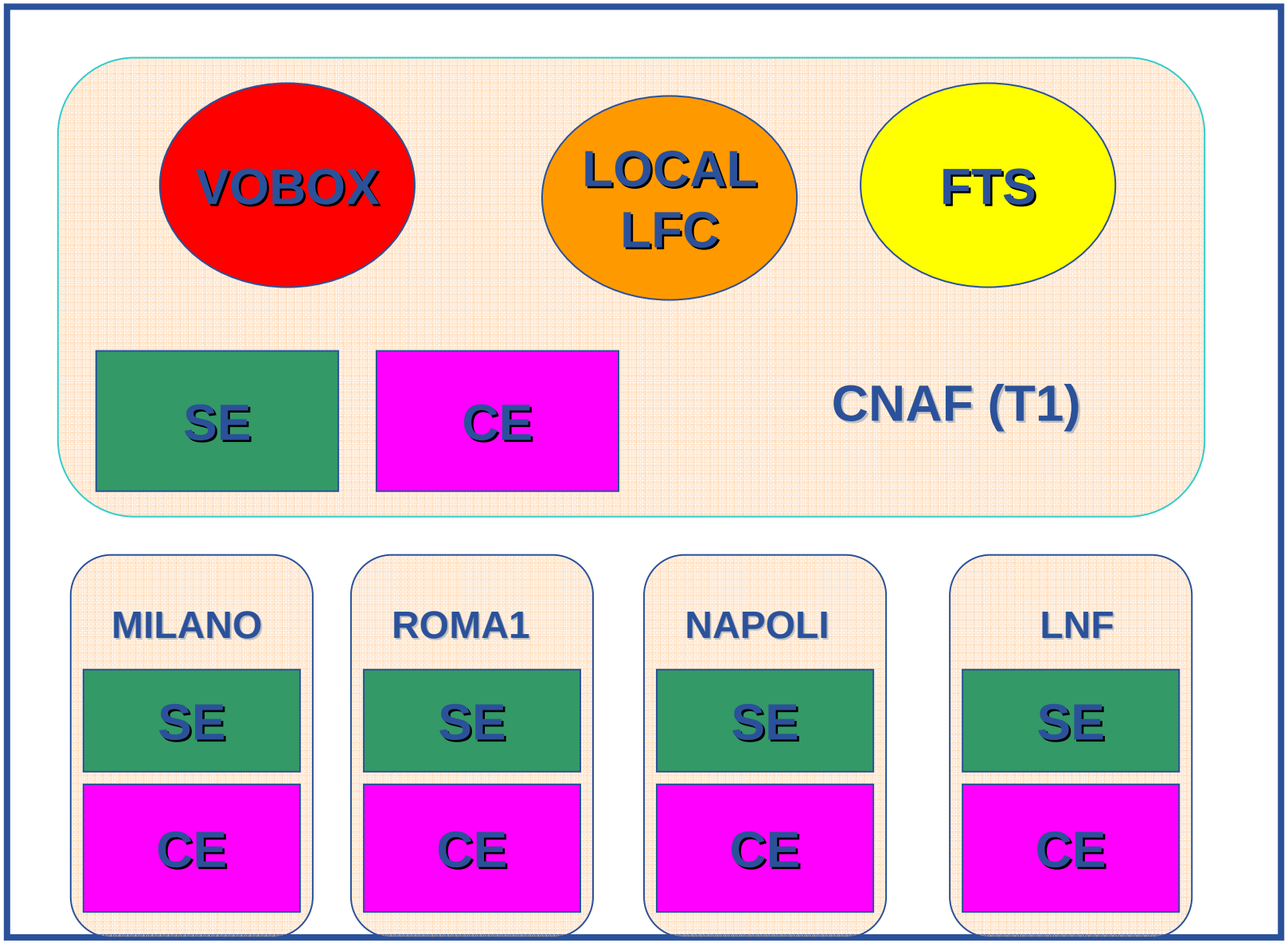
In simpler words: The DDM ...

- … enforces the concept of dataset
  - ✓ Logical collection of files
  - ✓ Datasets are defined in central catalogs
    - Dataset Repository: hold dataset name and unique ID
    - Dataset Hierarchy: holds dataset namespace information
    - Dataset Content: maps dataset with its logical content files
    - Dataset Location: store dataset locations
  - ✓ Local catalogs at the sites
    - Provide logical to physical file name mapping
    - LCG File Catalog in EGEE
- … based on a subscription model
  - ✓ Datasets are subscribed to sited
  - ✓ A series of services enforce the subscription
    - Lookup data location
    - Trigger data movement
    - Validate file transfer

TiersOFAtlas

- **Atlas T1s and T2s are logically organized in CLOUDS**
  - ✓ Mostly, this reflects geographical closeness and EGEE ROC organization
    - ❖ Not necessarily this is always the case
  - ✓ Motstly driven by the network topology
- **A cloud includes one T1 and several T2s**
  - ✓ Every T1 and T2 provides both CPU and Storage capacity for ATLAS
- **The T1 runs central services for the Cloud, including the DDM Site Services for dataset subscription handling**

The Italian Cloud (IT)



DDM: underlying services in EGEE

- **LCG File Catalog**
  - ✓ At every T1
  - ✓ Hold replicas information for files at T1 and connected T2s
    - ❖ See Cluod model next slide
- **gLite File Transfer Server**
  - ✓ Baseline service for file replication
  - ✓ At the T0
    - ❖ Ensures file transfers T0↔T1
  - ✓ At the T1s
    - ❖ Ensures file trasnfers T1↔T1 and T1↔connectedT2
- **SRM Storage Elements**
  - ✓ Not so much direct interaction at the moment
  - ✓ Could change in the future
    - ❖ Cache turnover, Claim Service
  - ✓ Need SRM2.2 for some of this

Prodsys-DDM integration

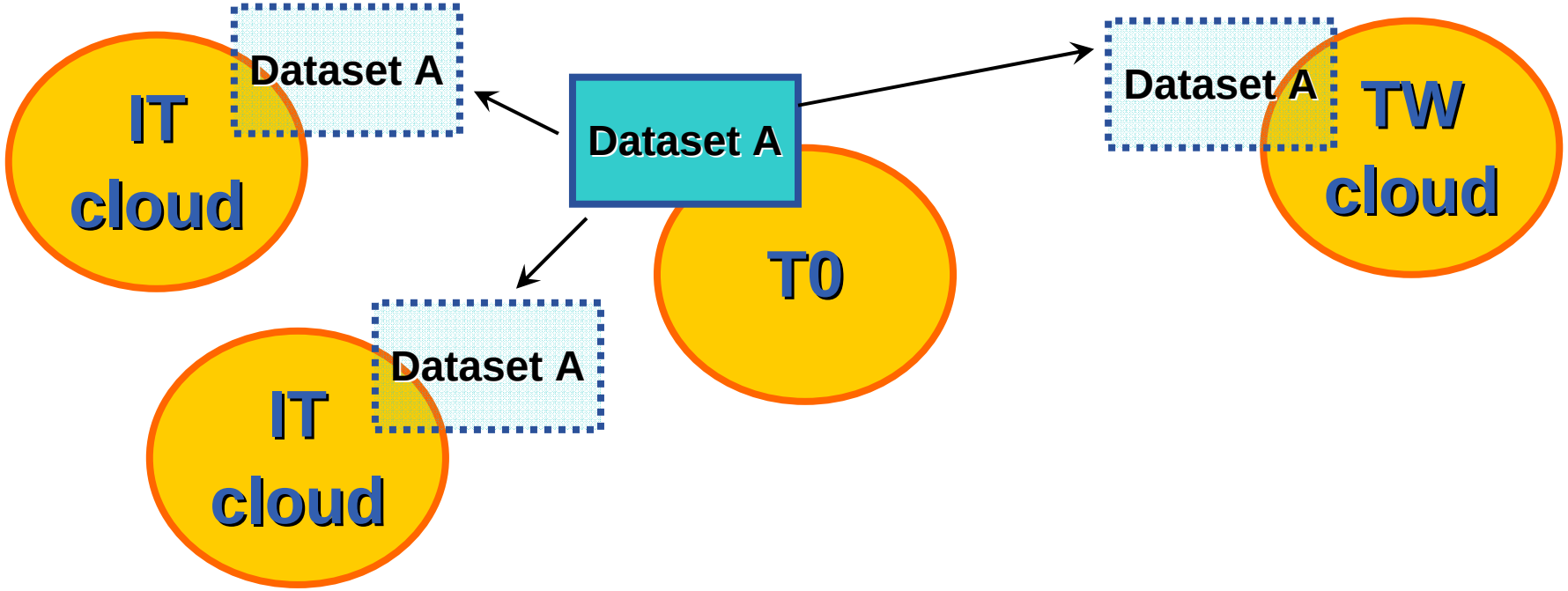
- **A Production Task is assigned to a CLOUD**
  - ✓ A task is a collection of jobs
  - ✓ Jobs of the same task process inputs from the SAME dataset
  - ✓ Jobs of the same task produce outputs in the SAME dataset
- **The gLite WMS or the CondorG submit the job to the "best" site of the cloud**
  - ✓ "Best" according to a ranking expression
- **This cloud must host the input dataset**
  - ✓ Files physically present in SRMs of the cloud and registered in T1 LFC
- **The jobs of a task store the output in a SRM of the cloud**
  - ✓ And registered in LFC at T1

Asynchronous File Movement (via DDM)

- Task X needs Dataset A
- Dataset A is in UK, but UK cloud is very busy with other tasks
- Replicate Dataset A to FR cloud (via DDM subscription)
- Assign task X to FR cloud

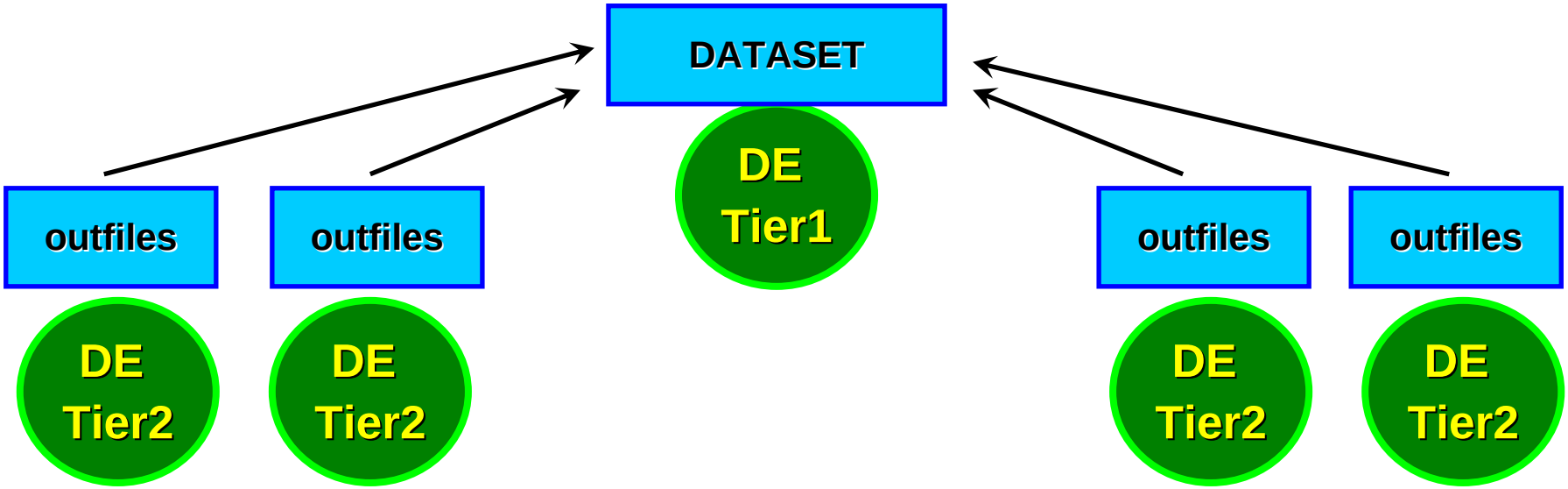
Asynchronous File Movement (via DDM)

- **Task X,Y,Z need dataset A**
- **Dataset A has been produced by physicists**
  - ✓ … and stored in CASTOR at CERN
  - ✓ … and registered correctly in DDM
- **Replicate Dataset A to IT,CA,TW cloud (via DDM subscription)**
  - ✓ And assign i.e. Task X to IT, Task Y to CA etc …

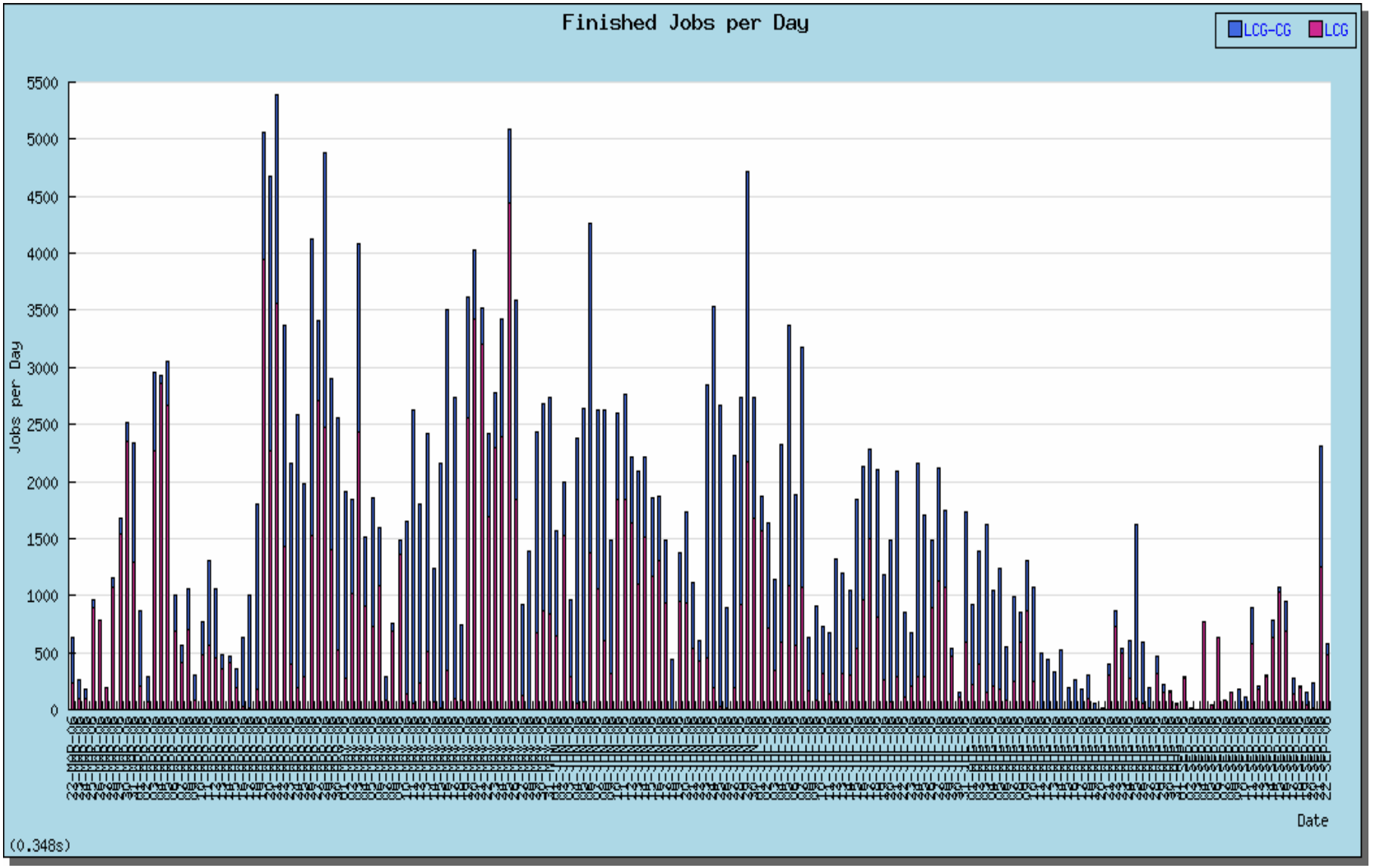


Asynchronous File Movement (via DDM)

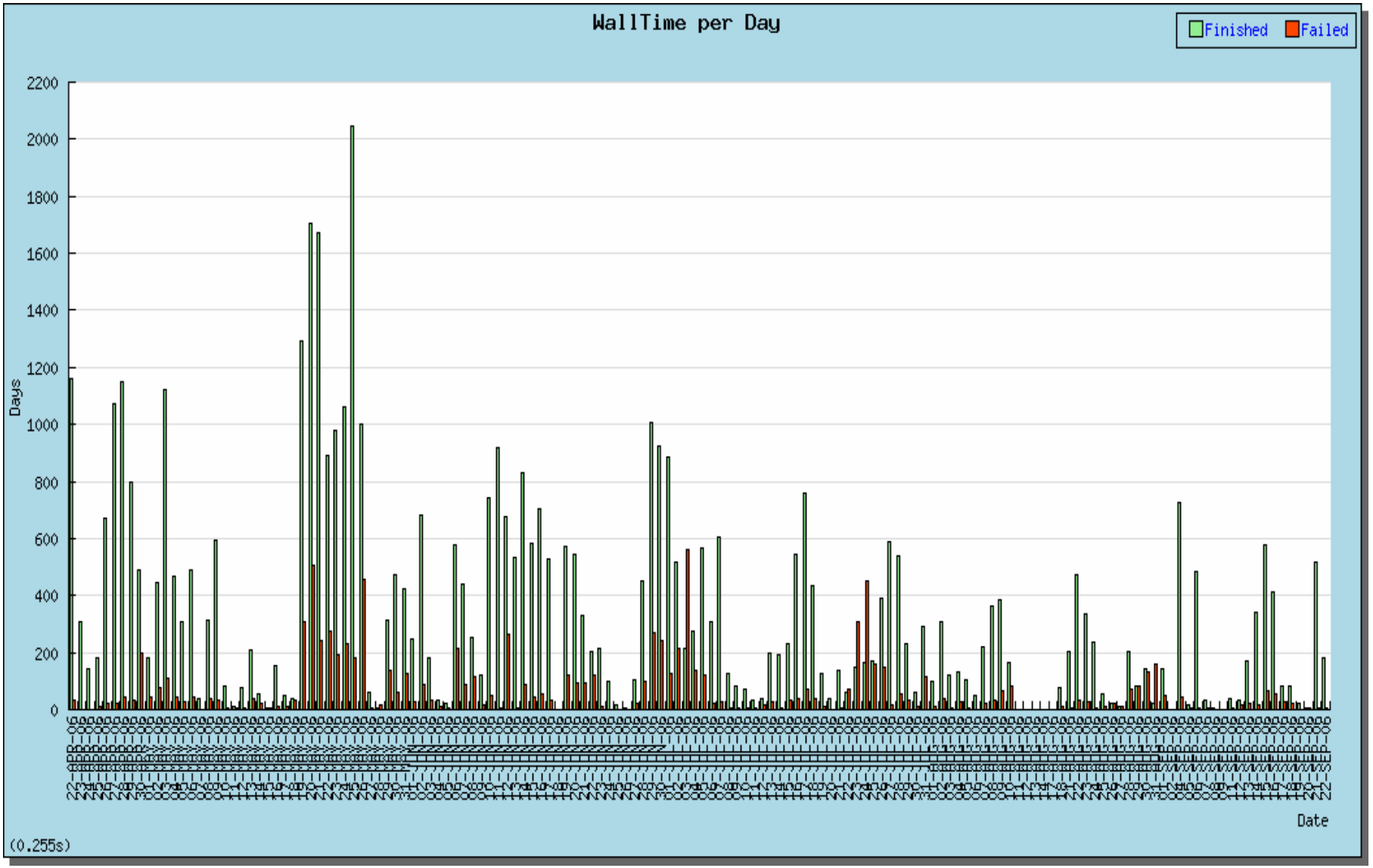
- **Task X is running in cloud DE**
  - ✓ producing output files
  - ✓ Files are stored in different DE Storage Elements
- **Subscribe output dataset to DE Tier1 Storage and aggregate output**
  - ✓ Streaming, as soon as files are produced
  - ✓ Particularly important for Analysys Data Files



From the ATLAS monitoring…



From the monitoring ...



Conclusion

- **ATLAS production system and Distributed Data Management have been fully integrated**
  - ✓ Need now to get experience operating the system
- **The integration will bring immediate benefits**
  - ✓ In the production activity
  - ✓ For future activities relying on produced simulated data i.e. analysis
- **Some improvements are foreseen**
  - ✓ Mostly to get more flexibility running jobs across different clouds
- **ATLAS prodsys need to scale up the throughput by a factor of 10 in the next year**
  - ✓ This will be the next challenge!
- **The EGEE infrastructure is CRUCIAL for ATLAS production**
  - ✓ Production with latest stable release run more than 60% of the jobs in EGEE