

A GRID COMPUTING MODEL FOR AGATA

DATA MANAGEMENT AND DATA PROCESSING

Based on the document
“The AGATA Grid Computing Model for Data Management and Data Processing”

M. KACI
IFIC / CSIC-UV – Valencia, SPAIN
Grid & e-Science Group
11 / 24 / 2010

kaci@ific.uv.es

WHY A GRID COMPUTING MODEL FOR AGATA ?



GRID ?

Coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations... (I.Foster et al. *)

AGATA ?

...problem solving ?

**The AGATA project is very challenging for the collaboration
Data Management and Data Processing are among the key problems**

multi-institutional virtual organizations ?

**Large collaboration of people,
Multi-institutional, geographically distributed over the world,
VO AGATA created yet since 2007,**

Coordinated resource sharing ?

MISSING...

PROPOSAL :

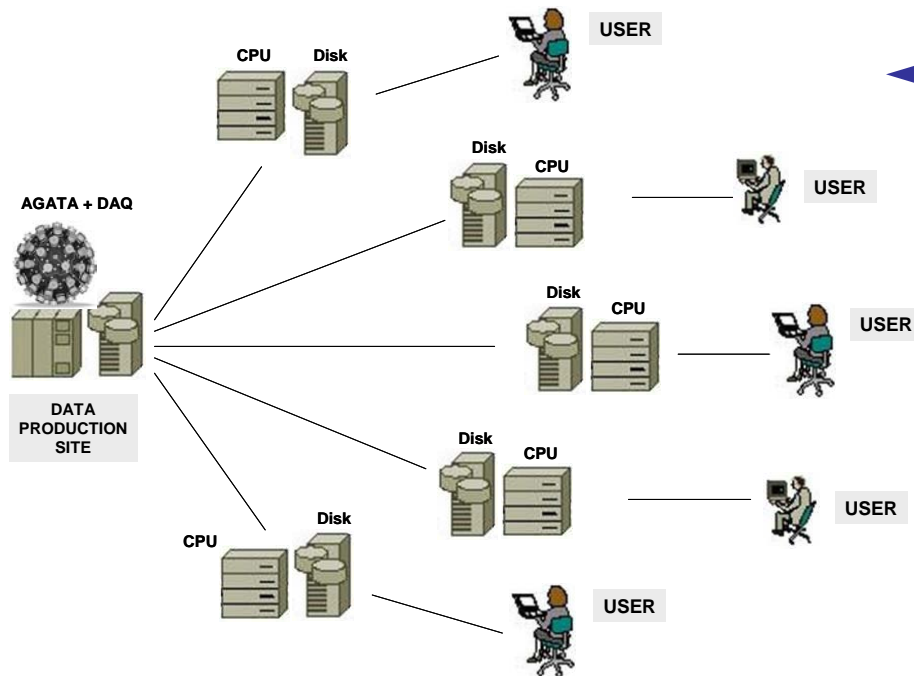
A GRID COMPUTING MODEL FOR AGATA DATA MANAGEMENT AND PROCESSING

(*): The GRID Blueprint for a New Computing Infrastructure, ISBN 1-55860-475-8

COORDINATED RESOURCE SHARING...

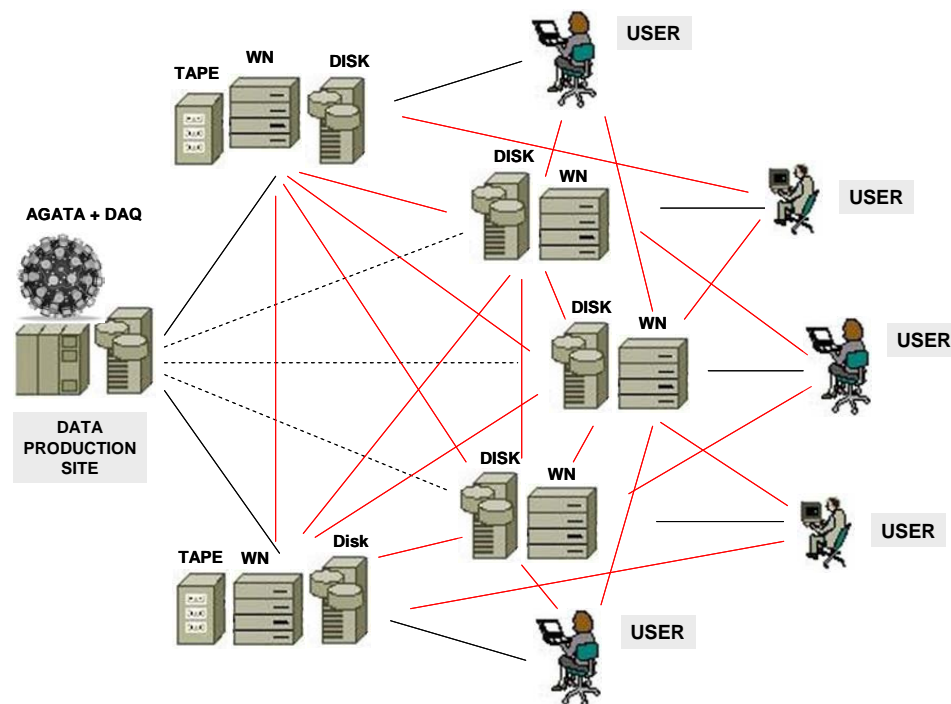
← **MIGRATE FROM...**

The classical situation where each user / institution uses his / its own computing resources and takes a copy of the experimental data at the local institution...



TOWARDS... →

A new situation where the users share computing resources and data in a coordinated way (policy) within a Virtual Organization...



OBJECTIVE :

Reduce the operational cost and the time of the Data processing for the AGATA collaboration by sharing the computing resources :

Sharing the Data :

**Reduces the storage space by having only few copies of the Data,
Secures the Data by having these multiple copies on the GRID,**

Sharing the Computing Power :

**Availability of large number of CPU for Data processing by the users,
Reduces the time of the Data processing (PSA, tracking, analysis)**

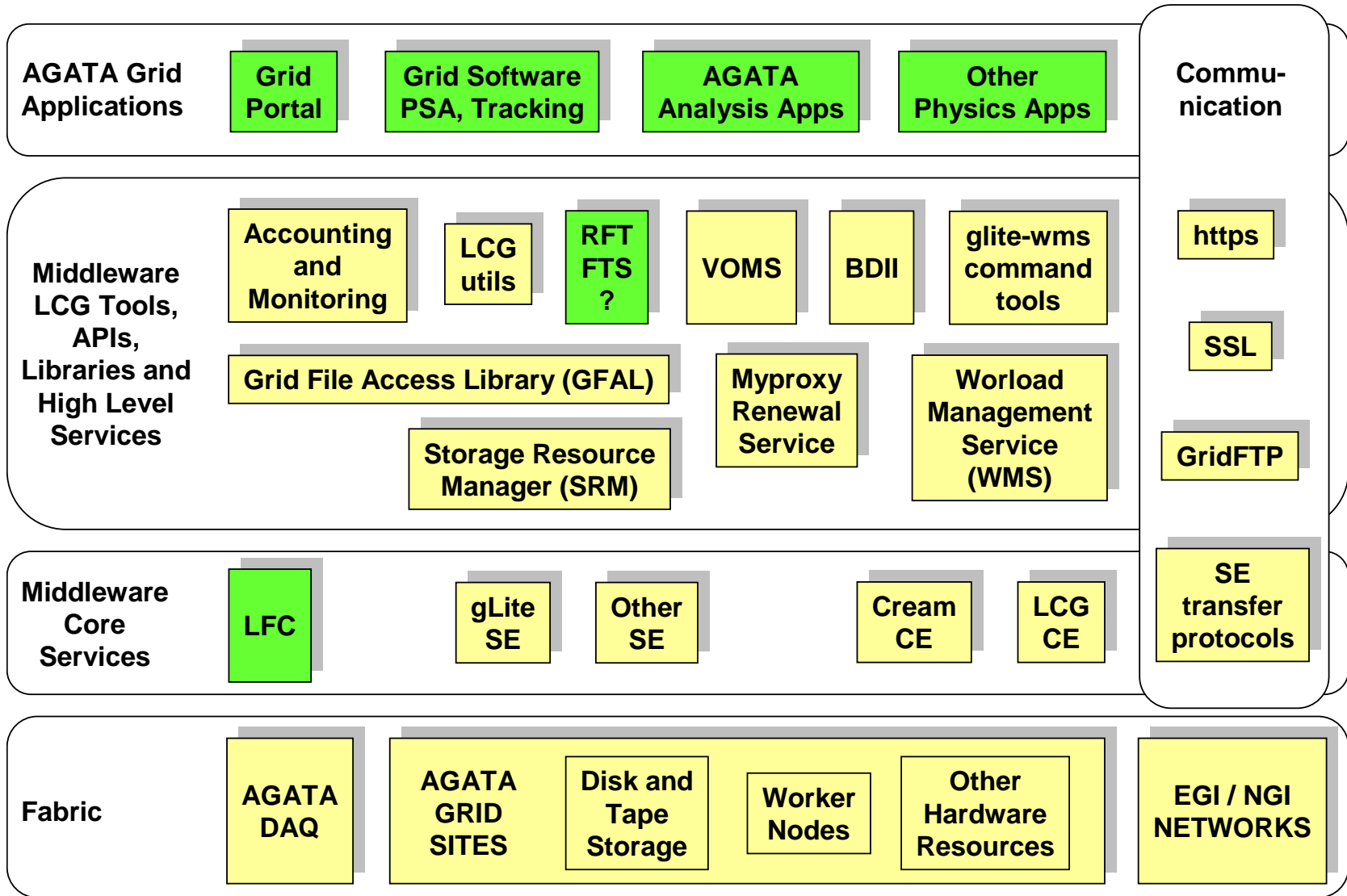
Provide the end user with an easy and light operational framework for Data Management and Data Processing on the Grid

WHAT TO DO FOR THAT ?

Little things, most of the GRID services needed by AGATA are operational through the EGEE projet (EGI presently) and using the LHC infrastructures,

Only few GRID components and applications have to be installed/developed for the optimal operation of the Grid-AGATA

THE AGATA GRID ARCHITECTURE



LEGEND:

General EGI / NGI components

Adapted components

THE EXISTING GRID RESOURCES FOR AGATA



STORAGE SYSTEM

Recording the line-shape
of the output signal...
TeraBytes of Data...
Hundreds of Data Files
per experiment...

```
-bash-3.2$ date  
Tue Nov 23 11:50:40 CET 2010
```

```
-bash-3.2$  
-bash-3.2$ lcg-infosites --vo vo.agata.org all
```

#CPU	Free	Tot	Jobs	Run	Wait	ComputingElement
656	55	0	0	0	0	lyogrid07.in2p3.fr:8443/cream-pbs-vo.agata.org
656	55	0	0	0	0	lyogrid02.in2p3.fr:2119/jobmanager-pbs-vo.agata.org
1216	18	0	0	0	0	sbgce2.in2p3.fr:8443/cream-pbs-vo.agata.org
536	134	0	0	0	0	ipngrid04.in2p3.fr:8443/cream-pbs-sdj
426	24	0	0	0	0	ipngrid04.in2p3.fr:8443/cream-pbs-agata
1216	18	0	0	0	0	sbgce1.in2p3.fr:2119/jobmanager-pbs-vo.agata.org
536	134	0	0	0	0	ipnls2001.in2p3.fr:2119/jobmanager-pbs-sdj
426	24	0	0	0	0	ipnls2001.in2p3.fr:2119/jobmanager-pbs-agata

CPU POWER

Off-line Pulse Shape Analysis
is time consuming...
Complex γ -ray tracking also...
Complex algorithms
(PSA, tracking)

Avail Space(Kb)	Used Space(Kb)	Type	SEs
4990008918	10796826	n.a	lyogrid06.in2p3.fr
10737418240	3262799712	n.a	ccsrm02.in2p3.fr
3740182493	2617242908	n.a	ipnsedpm.in2p3.fr
21445619256	72272689927	n.a	ipnsedpm.in2p3.fr
100000000000	500000000000	n.a	srn-v2.cr.cnaf.infn.it
36511000000	321996000000	n.a	srn-v2.cr.cnaf.infn.it
67730137056	380504621652	n.a	sbgse1.in2p3.fr

OTHER GRID SERVICES

UI, WMSPROXY, BDII,...

FILE CATALOGUE

Management of thousands of Data Files...

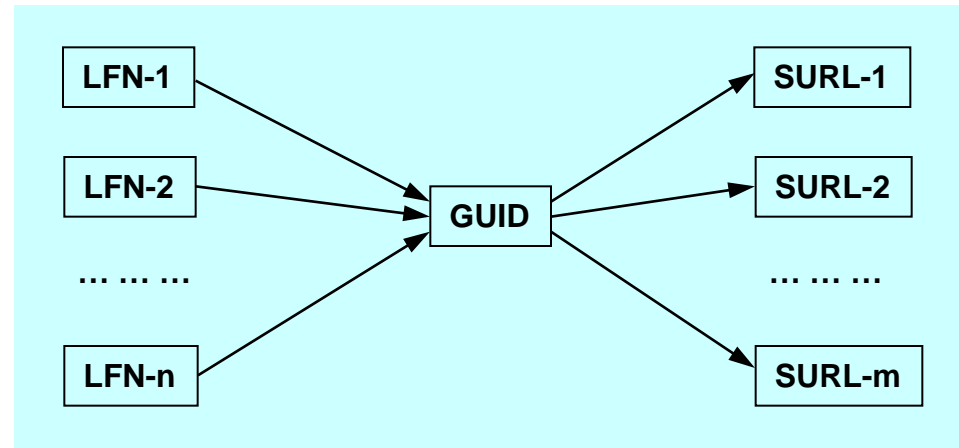
Multiple copies of the Data files (replicas) are well identified in the Grid

Easy location of Data Files by users, for their data processing...

Organization of the Data per experiment...

Management of the Data Access by ACLs

Only Groups of people belonging to the same experiment share the same data



FILE TRANSFER

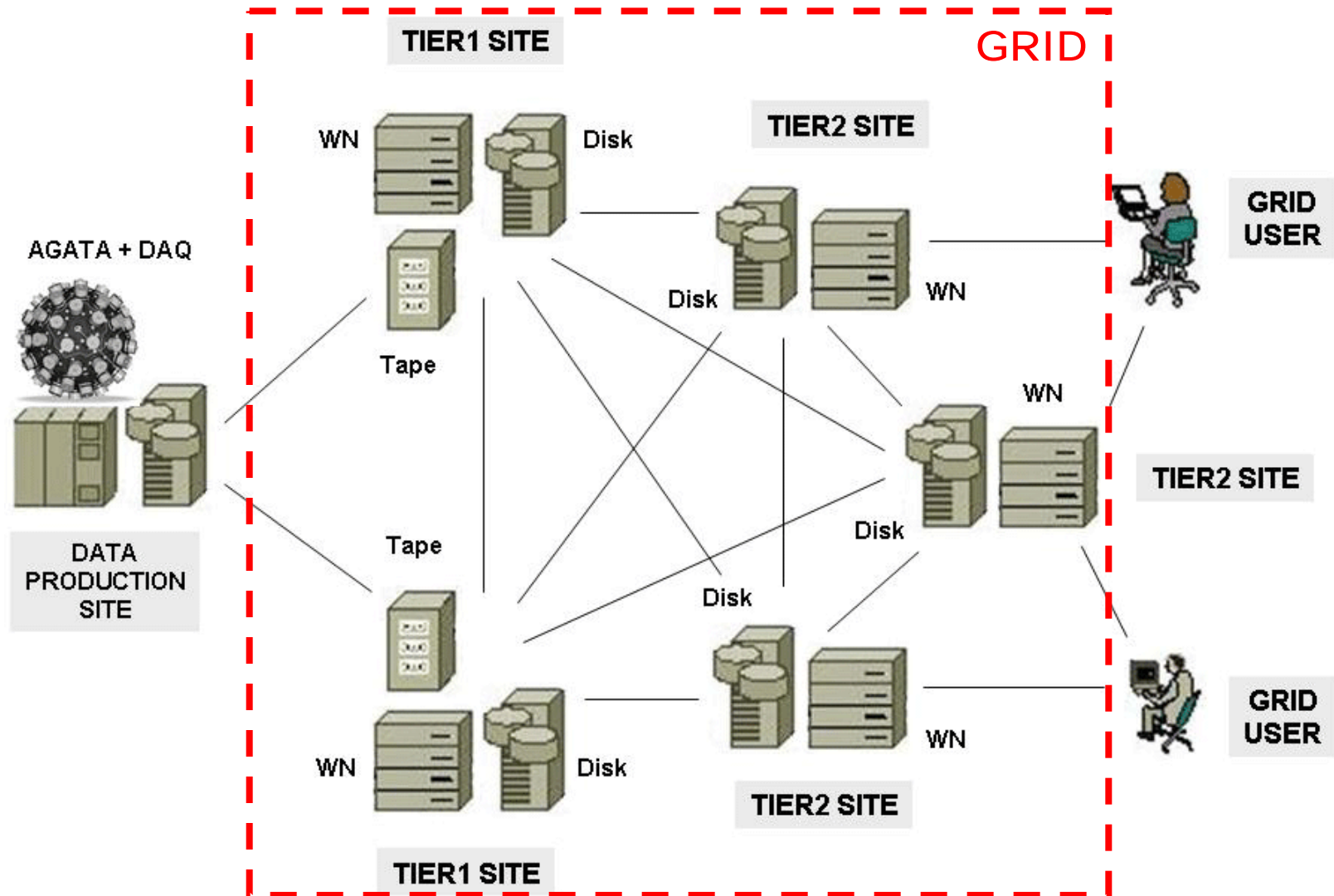
Move Data Files from DAQ to Grid...

Replication of Data Files within the Grid...

Retrieve user Data from Grid...

THE AGATA GRID COMPUTING STRUCTURE

The AGATA Grid Computing Model uses the LHC Tiers structure...
But, the Tier1s and Tier2s have similar roles in AGATA...
Tier1s provide Tape Storage,



THE AGATA DATA MANAGEMENT (I)



DATA MANAGEMENT POLICY

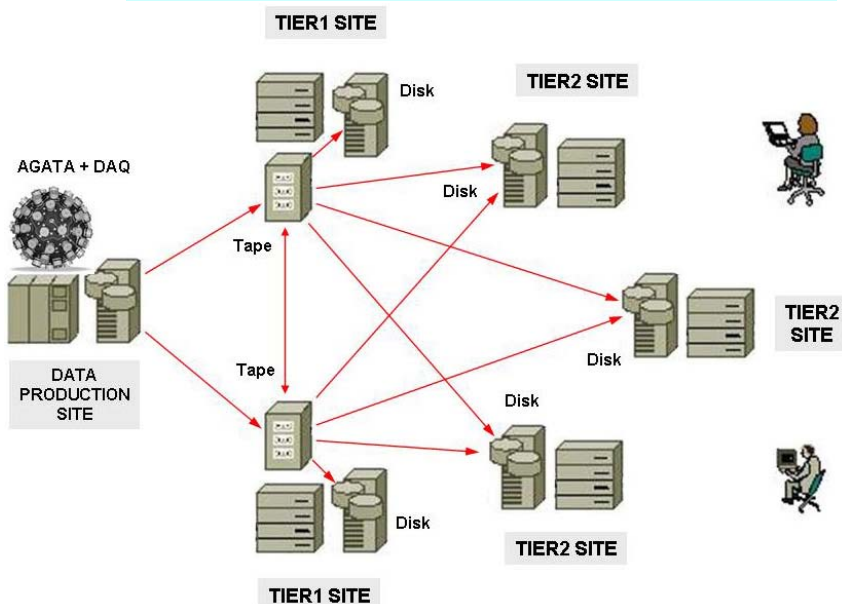
Policy for Data Access, Delete, Replicate, Use

DATA ORGANIZATION

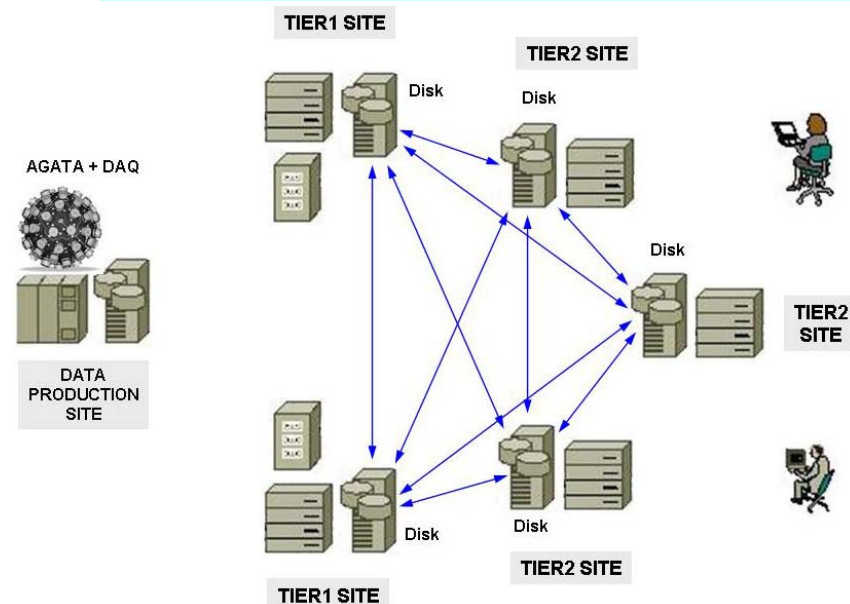
Size of the AGATA Data Files
Nomenclature for the Filenames of the AGATA Data
The Data Distribution and Placement
File Catalogue to organize the Data and the Access to them

DATA FLOW

Data Model: flow of the Raw Data



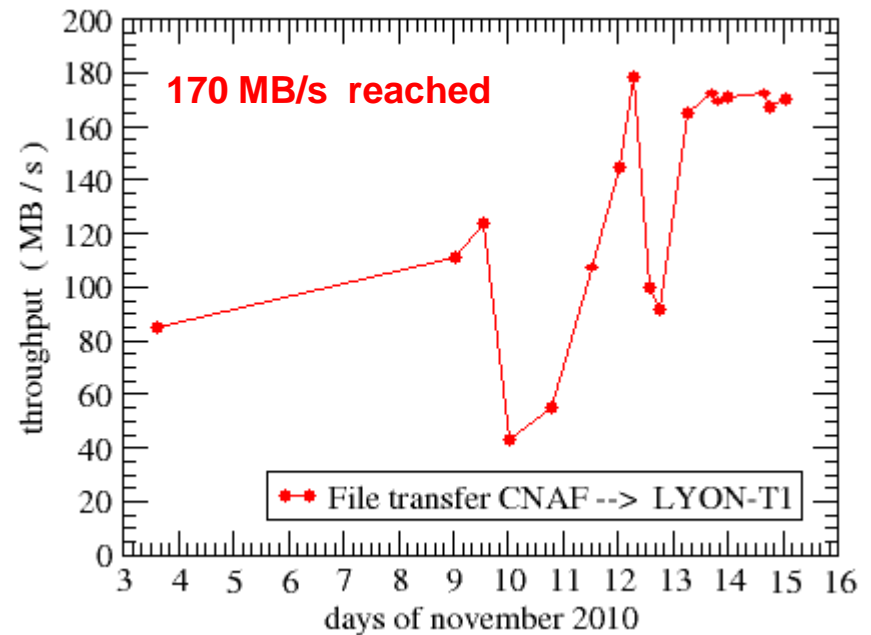
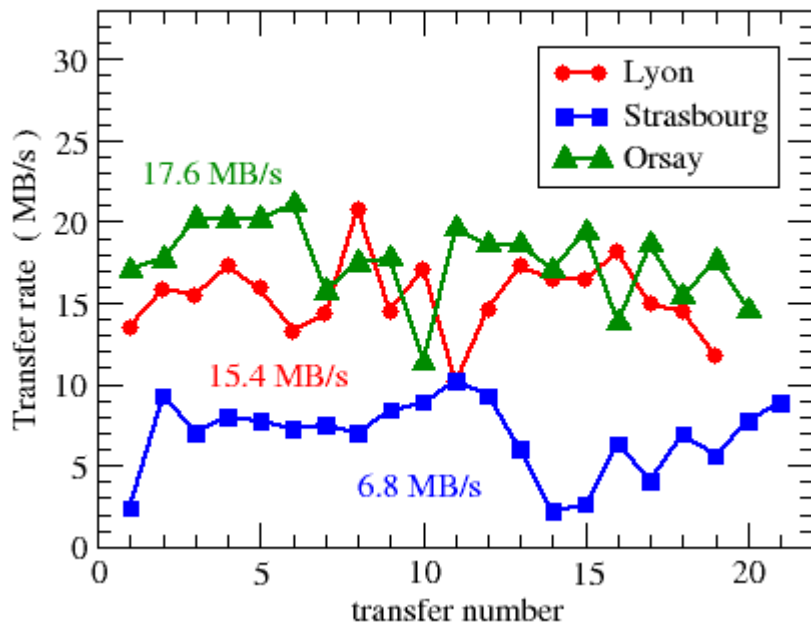
Data Model: flow of the Physics Data



DATA TRANSFER

- Moving Data from the DAQ Farm to the Grid
- Replicating Data between the AGATA Grid Sites
- Downloading User's Data from the Grid to the local Storage

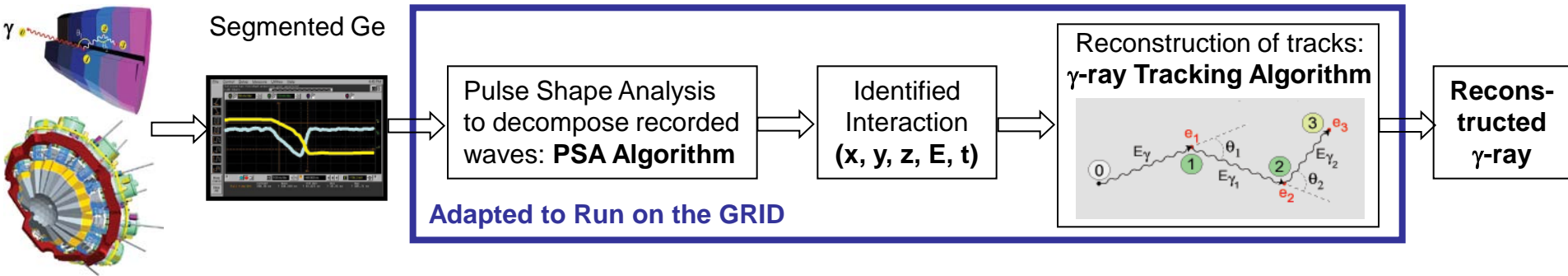
Transfer 1.4 GB file from CNAF-TI to WNs of France



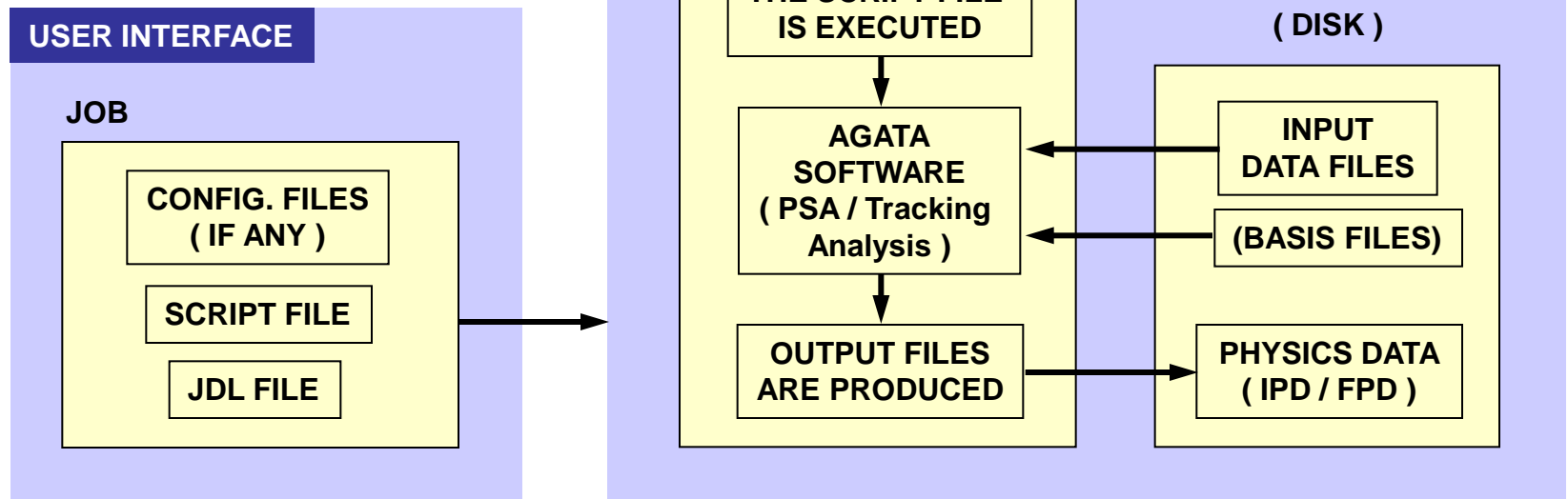
Improvement of the throughput using a distributed data transfer technique
More details in the Grid session on Thursday...

THE AGATA DATA PROCESSING (I)

ADAPTING THE AGATA SOFTWARE TO THE GRID :



RUNNING THE AGATA SOFTWARE ON THE GRID



THE AGATA DATA PROCESSING (II)



APPLICATION : RUNNING THE AGATA PULSE SHAPE ANALYSIS AND γ -RAY TRACKING ON THE GRID (**)

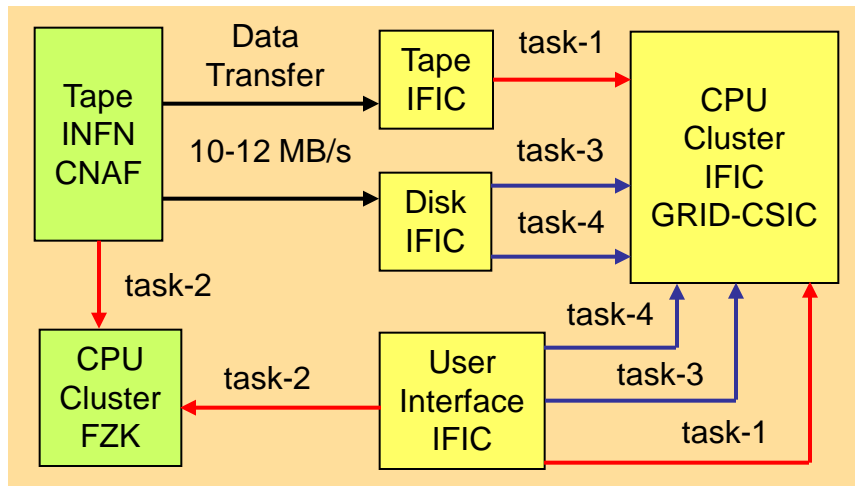
GRID RESOURCES USED :

GRID-CSIC : 50 cores (2GB per core, SLC5)
Sufficient Disk Space for storage (Lustre)

OTHERS : Additional storage 0.6 TB Tape (Castor)
Other EGEE clusters (CANF, FZK, MANCH.)



TESTS AND RESULTS :

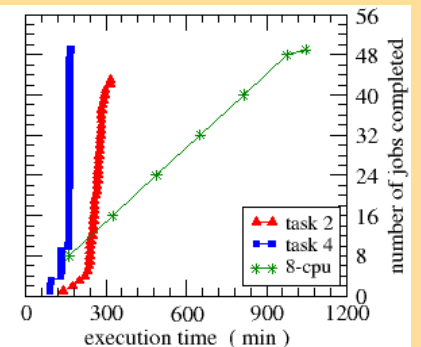
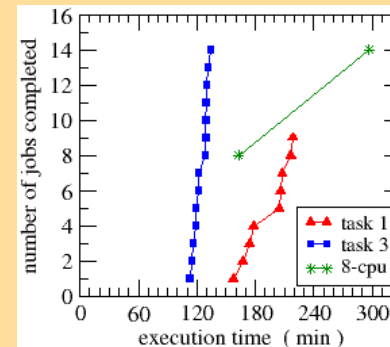


Task-1: 14 jobs, 0.6 TB data

Task-2: 40 jobs, 2.0 TB data

Task-3: 14 jobs, 0.6 TB data

Task-4: 49 jobs, 2.1 TB data



More results on using Grid remote access to Data on Thursday...

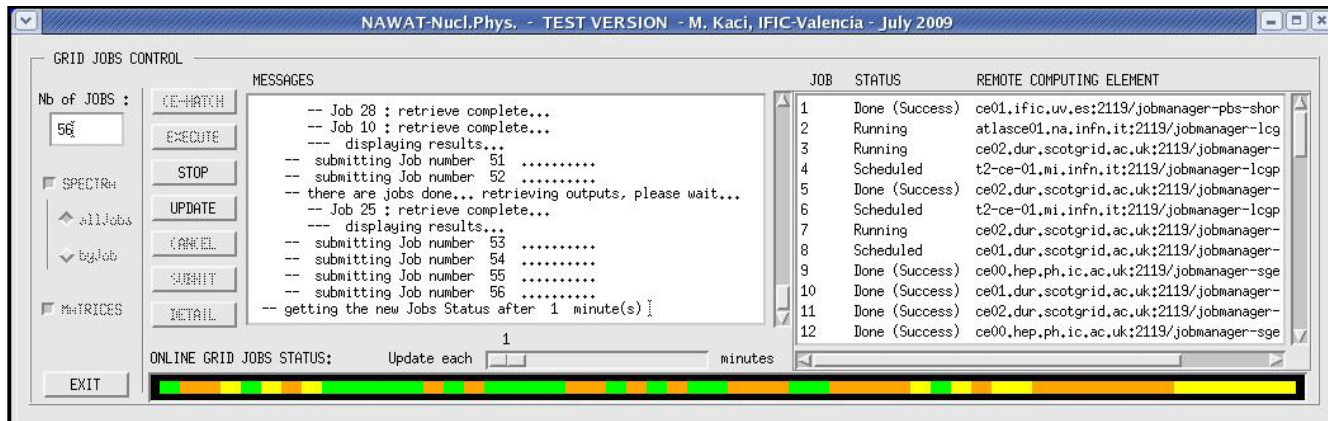
(**): M. KACI et al., 4th Iberian Grid Infrastructures Conference, Braga – Portugal, 22-26 May, 2010

THE END-USER IN THE AGATA GRID COMPUTING MODEL



The Computing Model would not be complete without a front end interface allowing the end users running transparently their data processing on the Grid...

A light Grid application with user friendly GUI is to be adapted for AGATA Data processing...



A support to the users for running their Data processing on the Grid is also contemplated in the proposed AGATA Grid Computing Model

OVERVIEW : PRESENT SITUATION IN AGATA GRID



The proposed AGATA Grid Computing Model is designed for the AGATA project and is oriented to the end users.

Last version of Narval Emulator is yet running on the Grid

