CNAF status and production plan

Armando Fella on behalf of Italian distributed computing group

Production tools

Two Production suites under development

1) site LCG/EGEE enabled: VO enabled, Gridftp, LCG-Utils

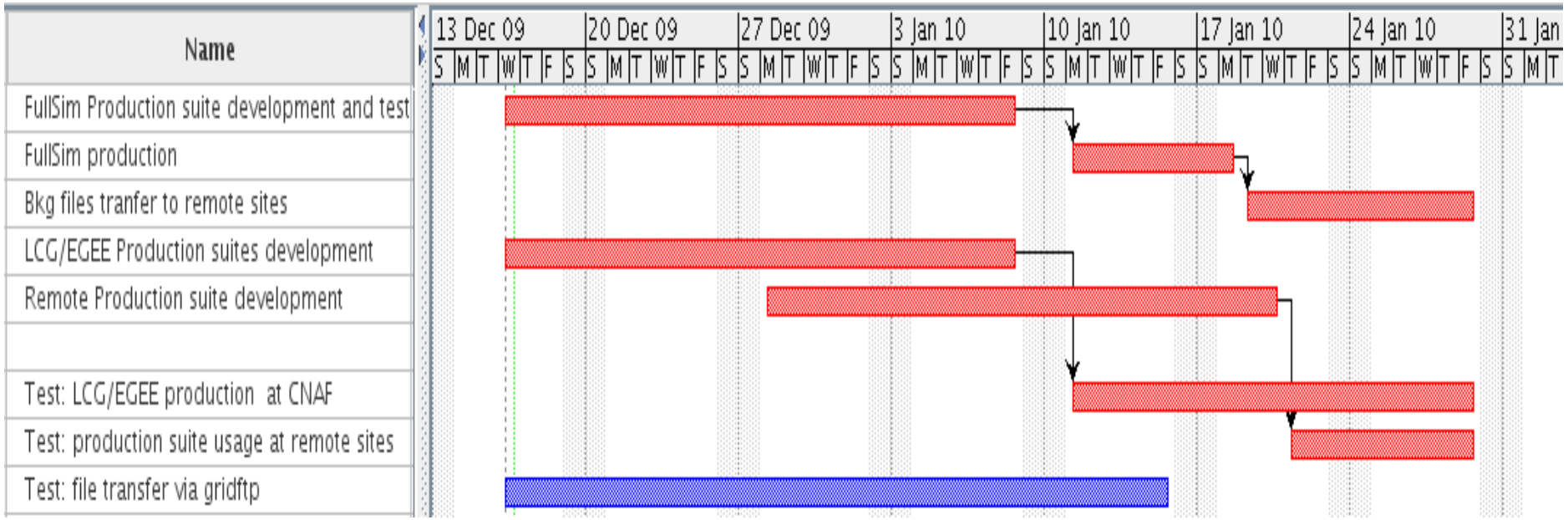
- input bkg files will be transferred to the site
- the production will be lauched from CNAF
- the job will transfer the output directly to CNAF via LCG-Utils
- the DB update will be performed offline or online (under investigation)
- no need of Production suite installation on the site

2) site OSG or not LCG/EGEE enabled:

- production suite installation on the site, submission via batch system
- the transfer of input and output should be performed via gridftp
- the DB update will be performed offline

The production suite flavors are under development, need site info

Production time line



- Start FullSim production within 10 Jan
- Start of bkg files transfer to remote sites: 16 Jan

Site setup procedure notes

- Transfer tool Gridftp is the only one supported by CNAF center,
  to use the tool you need to be a member of superbvo.org.
  A new wiki requirement was added: “Transfer test page”

- A production service test procedure is under development,
  please report on “Site status” wiki page the requested info, in particular
  the Storage Element hostname for LCG-Utils/Gridftp test

- The Grid way (OSG/LCG) is the preferred, please try to solve the
  VO site enabling “bureaucratic” steps. To achieve the goal is necessary
  to have a profitable communication with Grid people in each site.

BACKUP

Distributed Computing human network

CNAF	Italian group
Caltech	Frank Porter, Piti Ongmongkolkul
SLAC	Steffen Luiz, Wei Yang
McGill	Steven Robertson
Queen Mary	Adrian Bevan
RAL	Fergus Wilson
LAL and Lyon	Nicolas Arnaud
Bari	Giacinto Donvito, Vincenzo Spinoso
Legnaro - Padova	Gaetano Maron, Alberto Crescente
Napoli	Silvio Pardi
Ferrara	Giovanni Fontana, Marco Ronzano
Pisa	Alberto Ciampa, Enrico Mazzoni, Dario Fabiani

Email list: **superb-grid-mng@lists.infn.it**

Production design proposals

Next MC productions involve the exploitation of distributed computing resources, two different distributed design systems are proposed:

the February production distributed system proposal include the installation at remote sites of tools used at CNAF during the Nov '09 test production.

the web production interface, local Bookkeeping DB communication, Data handling via direct access to file system

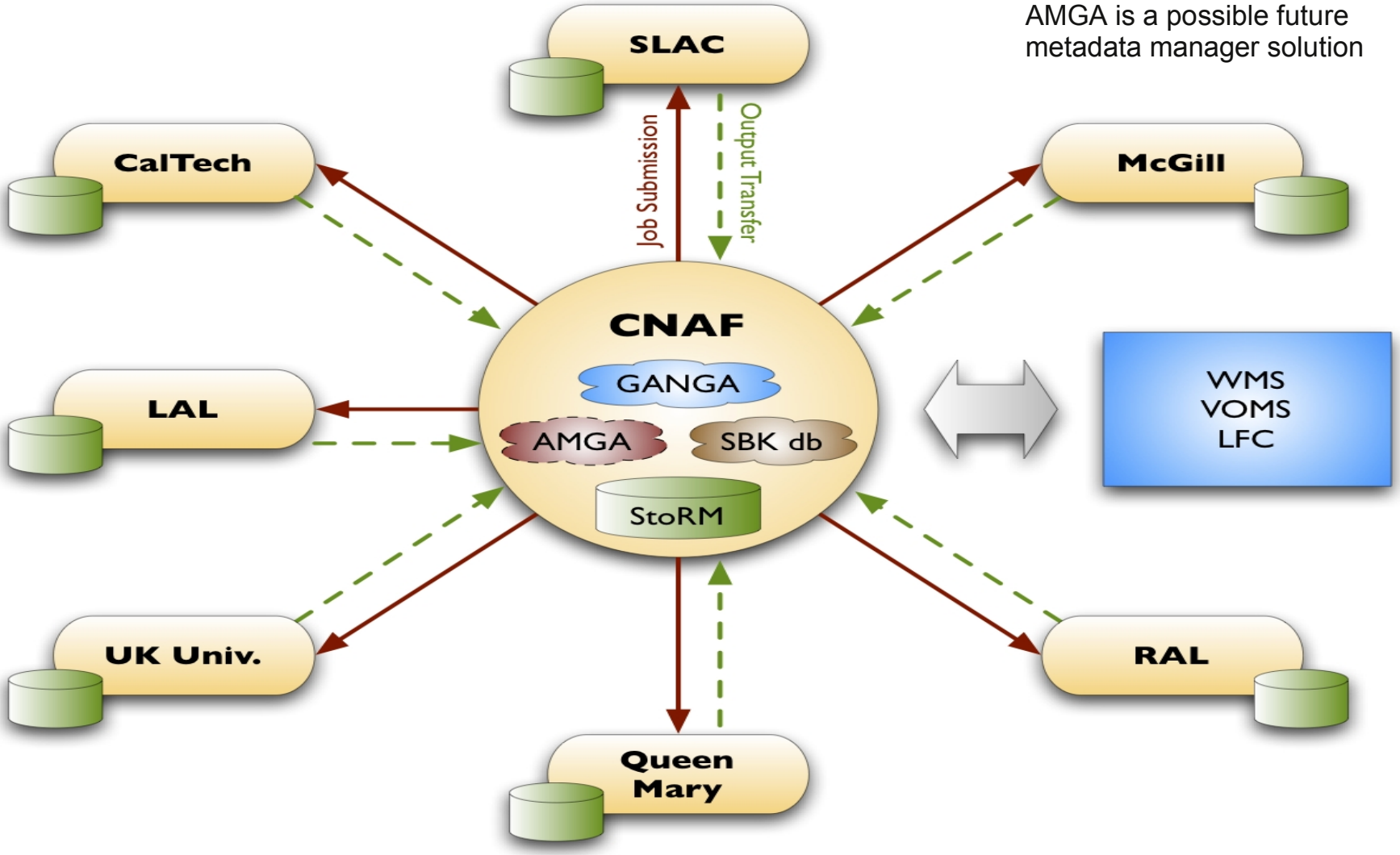
the setup of a full Grid compliant system will be proposed as solution for April '10 production

Full GANGA based job management, Grid Security Interface authentication, central bookkeeping DB communication, data handling via lcg-utils.

the site resources exploitation during official production should be ruled:
- the production managers will have priority on systems utilization
- an agreement on submission policy with user community should be defined

Full Grid integrated production design

AMGA is a possible future metadata manager solution



Full Grid integrated production workflow

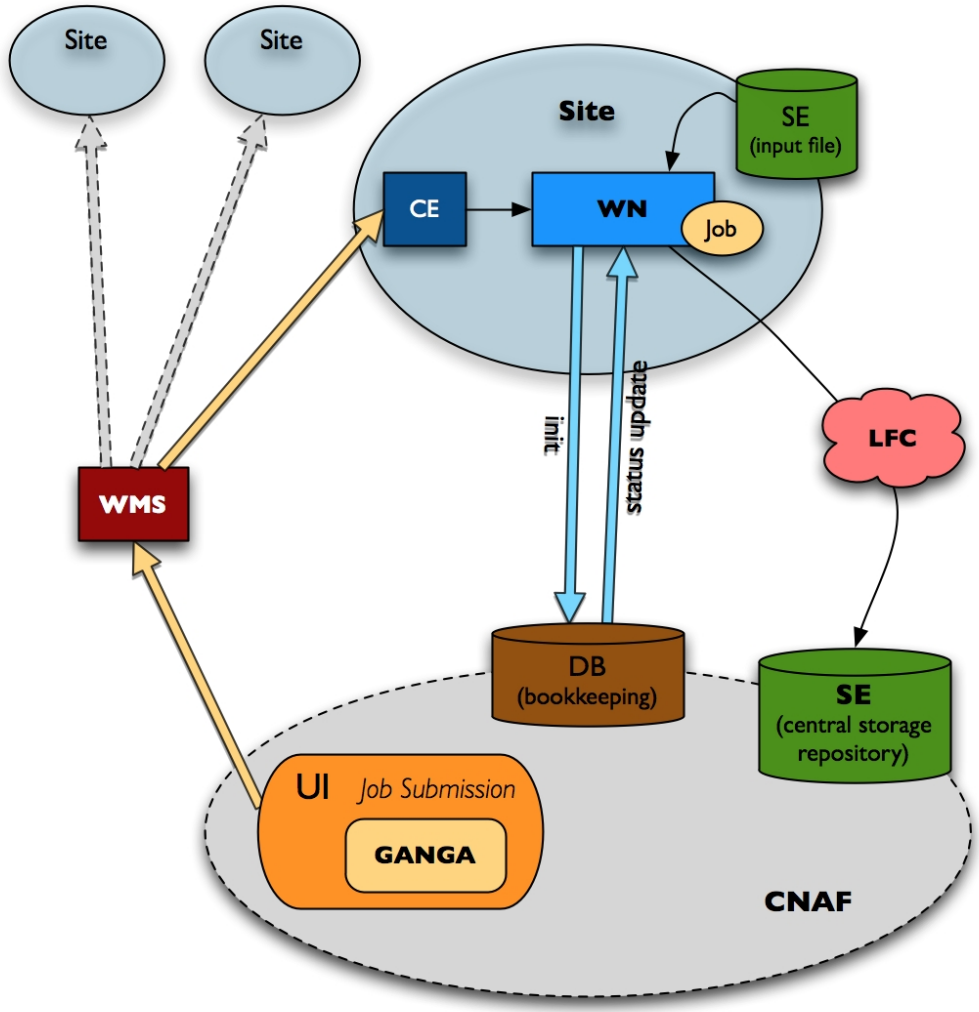
The job input files are transferred via
- LCG-Utils to the involved sites
- Storage Elements

The job submission is performed by GANGA on User Interface at CNAF

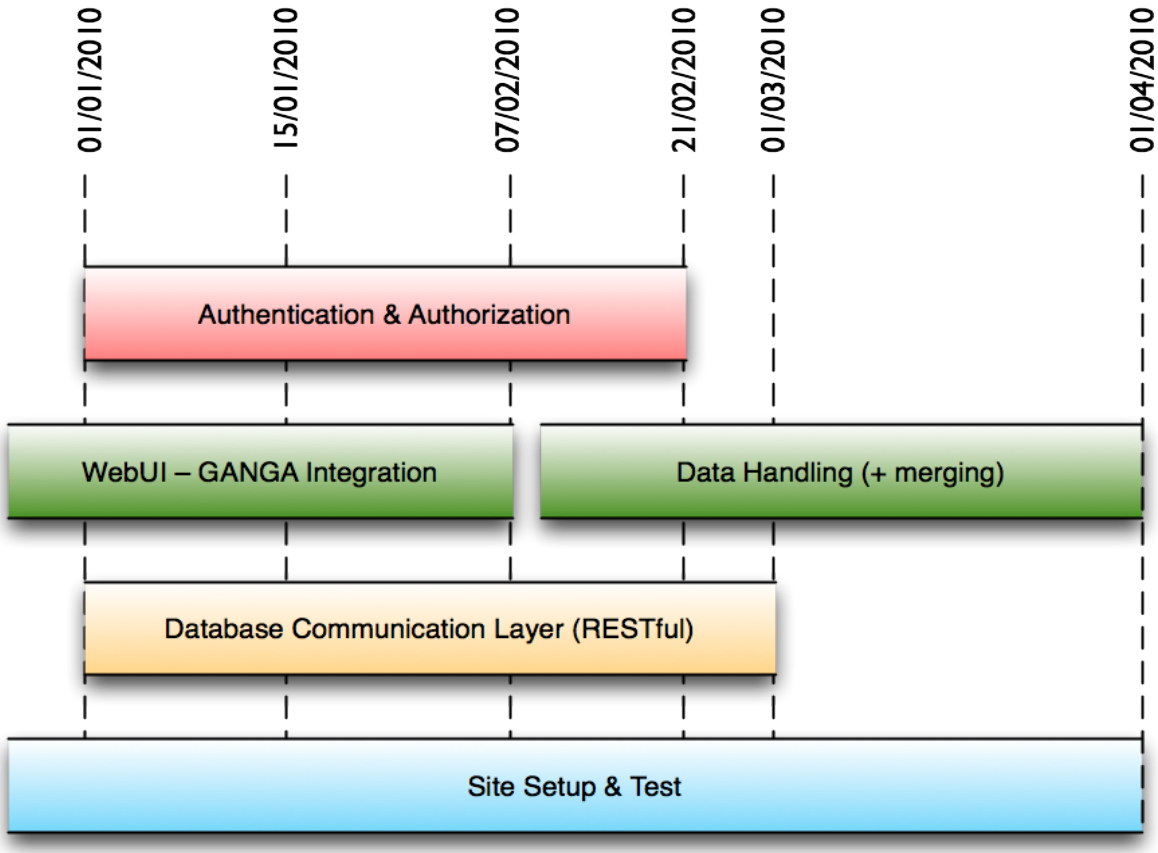
The WMS routes the jobs to the matched sites

The job is scheduled by the site Computing Element to a Worker Node

The job during running time
- accesses the DB for initialization and status update
- retrieves input files by local Storage Element
- transfers the output to the CNAF Storage Element



Full Grid integrated production timeline



February production design description

Production system components to be installed at sites:

Bookkeeping MySQL DB, phpmyadmin as management web interface (optional)
Apache web server
PHP 5 scripting language + apache php module
Web production and monitor interface software (submission to Grid and to batch)

Requirements:

One server host accessible from worker nodes where the above components should be installed (prod.server@your.site)
Disk space to contain input production files (local Storage Element)
Disk space accessible from WN, the output files should be stored in a system five like file system (Eg.: ext3, xfs, jfs, gpfs....). The output files should be copied back to CNAF central storage at end production time.

February production procedure proposal

1) Background files production at CNAF :

The background files to be used as input for Fast Simulation are produced at CNAF

2) Background files transfer to sites :

The background files should be transferred to file system accessible from WN

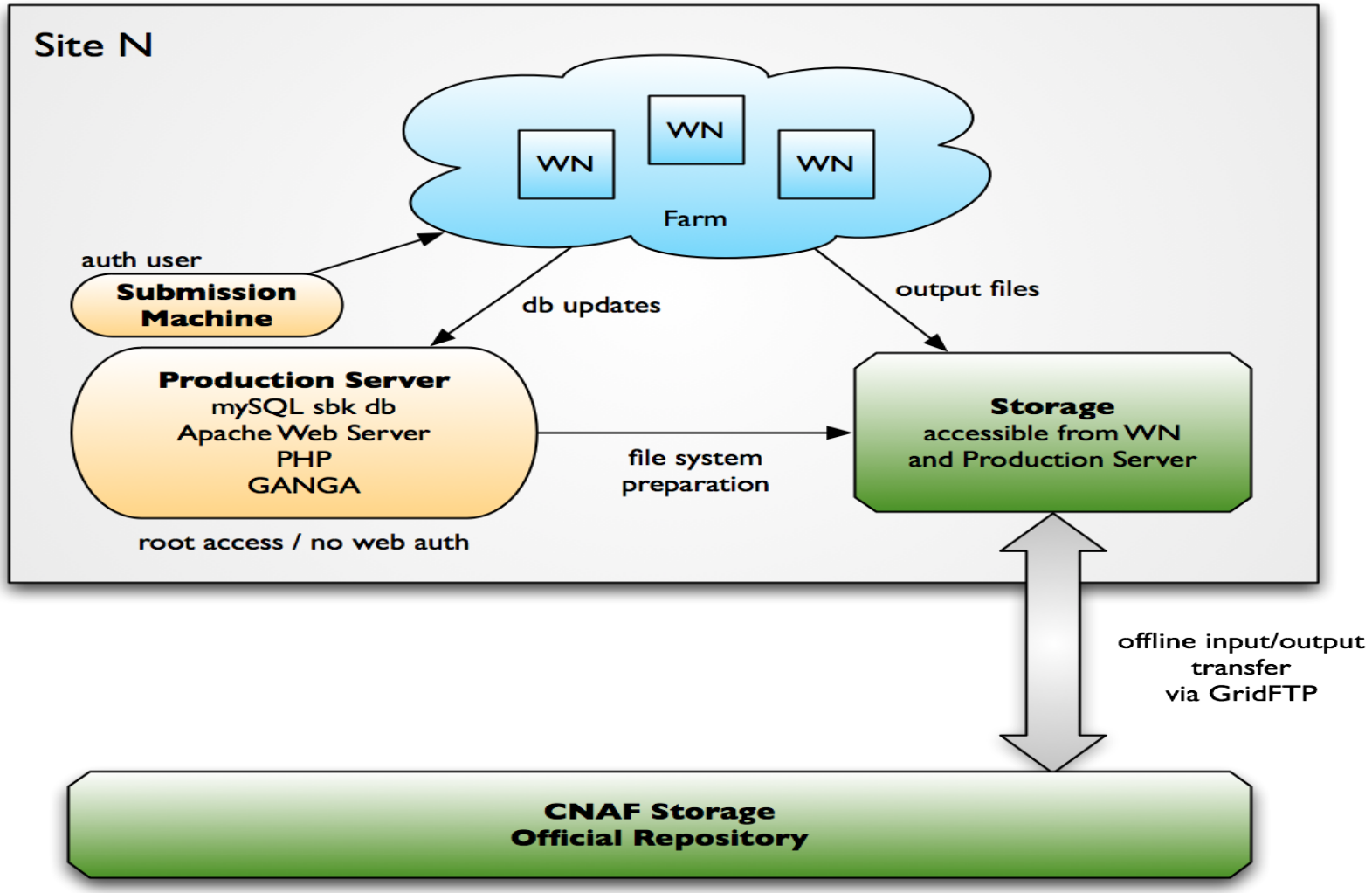
3) Perform the production at sites :

Use the web User Interface to produce the simulated data
Store the data into an accessible disk space

4) Output files transfer to CNAF central storage :

The job output files should be transferred to CNAF

February production procedure proposal



Conclusion

The February distributed solution is an hybrid solution looking toward the full Distributed (Grid compliant) solution.

The coordination/cooperation with involved site contacts is a key element

The time needed for the February production completion depends strictly by the number of sites ready at the production start time