Optimization of Italian CMS Computing Centers via MIUR funded Research Projects

T.Boccali[e], G.Donvito[h,e], A.Pompili[h,e], G.Della Ricca[b,e], E.Mazzoni[e], S.Argiro' [a,e], C.Grandi[e], D.Bonacorsi[c,e], L.Lista[e], F.Fabozzi[g,e]L.M.Barone[f,e], A.Santocchia[d,e], H. Riahi[d,e], A.Tricomi[i,e], M.Sgaravatto[e], G.Maron[e]

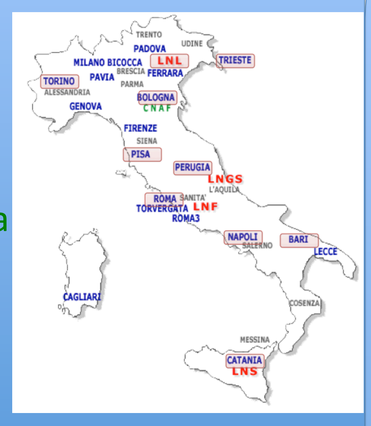
a b c d e f g h i

In 2012, 14 Italian Institutions participating LHC Experiments (10 in CMS) have won a grant from the ITALIAN MINISTRY OF RESEARCH (MIUR), to optimize Analysis activities and in general the Tier2/Tier3 infrastructure. We report on the activities being researched upon, classified under the broad-brush categories:

1. **National level Xrootd federation**
2. **Optimization of interactive systems**
3. **Distributed analysis tools**
4. **Cloud Systems**
5. **New technologies**

The participating sites for CMS are:

- INFN (Pisa, Laboratori di Legnaro)
- Univ. Trieste
- Univ. Torino
- Univ. Bologna
- Univ. Perugia
- Univ. Roma Sapienza
- Univ. Napoli
- Politecnico di Bari
- Univ. Catania



Optimization of computing centres for (interactive) analysis

While GRID enabled access to the resources is well established in our sites, the final step of physics analyses is less specified in the CMS computing Model. The activities which are under study are:

- **"User Interface on demand"** via LSF/PBS sharing with Worker Nodes, to allow for a variable number of interactive machines depending on the request. This increases resource usage, since we can avoid to reserve a large number of User Interfaces, to stay mostly idle, and can use them as Worker Nodes for most of the time.
- **Italy wide login on all User Interfaces**: this has been implemented via AAI (Authentication/ Authentication INFN system), and is currently tested on a few sites. Every Italian user, registered centrally (at the INFN Administration) as a CMS member, can login on a selected number of User Interfaces without any direct interaction with the local site.
- **PROOF deployment**: either on large (64 core) machines, or on the existing GRID clusters. Tests with Prood on Demand are being evaluated.
- **Xrootd caching servers** at the frontiers of small analysis centers: in centers with small storage systems, pre-allocating large data samples is unpractical, and Xrootd access is preferred. On the other hand, the final analysis step is often repeated many times, and a Geographical Xrootd access cannot be optimal. The solution we implemented is based on Xrootd caching servers: in these sites, the whole Xrootd Federation is faked as a "tape backend" to the local storage: if a file is not found locally, it is "staged in" via the Federation, and made to reside locally. Subsequent accesses will be local. Xrootd also takes care of purging the local storage when full, eliminating older files.

Tests of new computing technologies

We are investing manpower and resources in the test of technologies which may become relevant to CMS Computing in the longer run.
One strength of the CMSSW Software Stack is its complete independence from any proprietary code. A library / algorithm / tool, to enter the stack, must allow complete code distribution and patching.
In this way, we can recompile the full stack on virtually any POSIX platform with a c++ compiler, and even more easily on platforms which support g++.
We are currently performing benchmarking and porting activities on

- Xeon Phi
  - 3 machines available in Pisa, Bologna
- ARM architectures
  - Single "consumer level" boards like the HardKernel Odroid-U2 (a naked Samsung Galaxy S3)
  - Server-grade ARM cluster-in-a-box (Dell Copper)
  - Waiting to get hold on the first 64 bit ARMv8 chips

(Italian) CMS Xrootd Federation

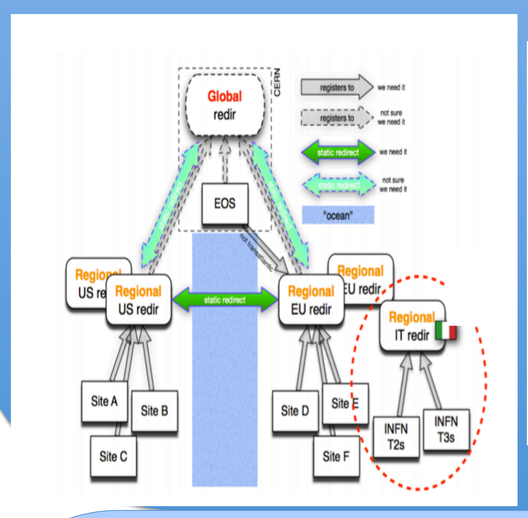
For the next LHC Run (2015-2017) the LHC Experiments, and CMS in particular, are planning to rely more and more on direct access to remote data. The **Italian federation of CMS Tier1/Tier2s** has been the second overall, and the first in Europe, to allow for remote Xrootd access.
Italy (Bari) hosts the European level redirector, which serves all European Tier2s and CERN. This is matched on the US side by the Nebraska redirector.

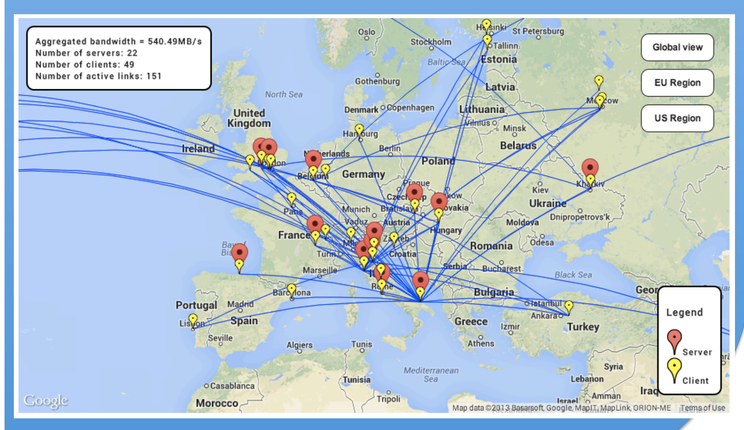
In order to allow easy exploitation of the resources in Italian Tier3 sites (Trieste, Napoli, Torino, Bologna, Perugia and Catania in this project), a specific redirector has been setup to include all Italian resources (Tier1, Tier2s, Tier3s…) as a single entity.

A specific development has been carried on to allow Xrootd access from the Tier1, at CNAF: the Tier1 uses GPFS +TSM as disk and tape systems, with automatic recalls from tape when a file is requested in GPFS. In conjunction with Xrootd developers, a specific plugin to "protect" the tape backend from chaotic recalls has been designed, implemented and tested, and is currently in production. In this way, only files already on the Disk Buffer are made available to the analysis tasks, with no possible stress on the tape drivers.

The Xrootd federation also provides a more robust computing infrastructure for CMS in Italy. A site's storage downtime does not automatically imply stop of the site activities: local CPUs can continue processing data received transparently from the other sites in the federation.

Majority of the European Xrootd, traffic originates from Italy (see Xrootd monitoring map below). In the current Computing Model, accesses via Xrootd are restricted to specific and "rare" cases, but this operation mode is going to become more and more used in the coming years.





Tools for Distributed Analysis

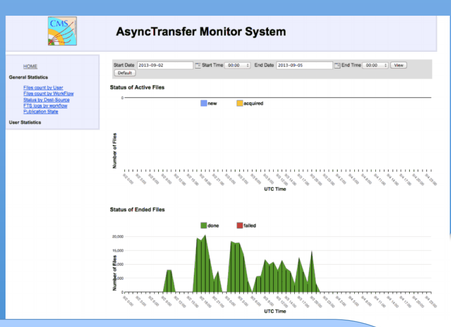
Italy is committed to the design, development and integration of the next generation tool for CMS Distributed Analysis tool, CRAB3. For the distributed analysis tool sustainability, CRAB3 is integrated with the distributed Analysis tool of ATLAS, PAnDA, into a Common Analysis Framework. The efforts are undertaken by CMS, ATLAS and the CERN/IT department. Italy is responsible for design and development of main components in the framework:

- **TaskManager backend**: needed to provide into PanDA the concept of Analysis Tasks.
- **Users Data Management system (AsyncStageOut)**: it manages and monitors the transfer and publication of CMS analysis jobs outputs.
- **CMS component in PanDA**: it handles the CMS jobs metadata to allow the later transfer and publication of the outputs, and also to create analysis reports to end-users.

Italy has ensured also crucial activities for the integration of the framework:

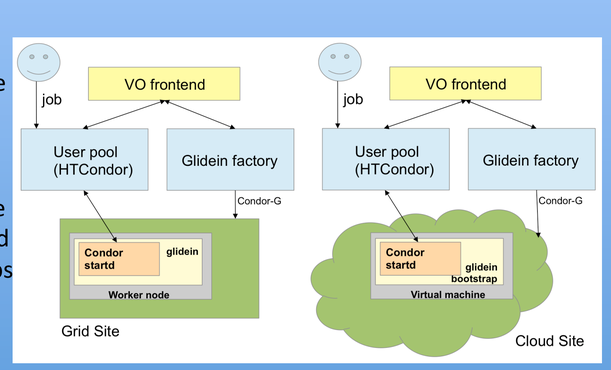
- **AsyncStageOut scale tests**: during the commissioning of the framework, scale test of the AsyncStageOut independently of underlying services has been performed. Figure shows the satisfactory performances demonstrated by the tool.
- **Alpha and Beta testing**.

In the future, we plan to add more commonalities to the framwork for CMS and ATLAS, such as the user data management, the analysis job splitting or the framework deployment.
Actually, the AsyncStageOut can interact only with the Workload management tools such as PanDA. The AsyncStageOut will be exposed also to users for the management of their files.

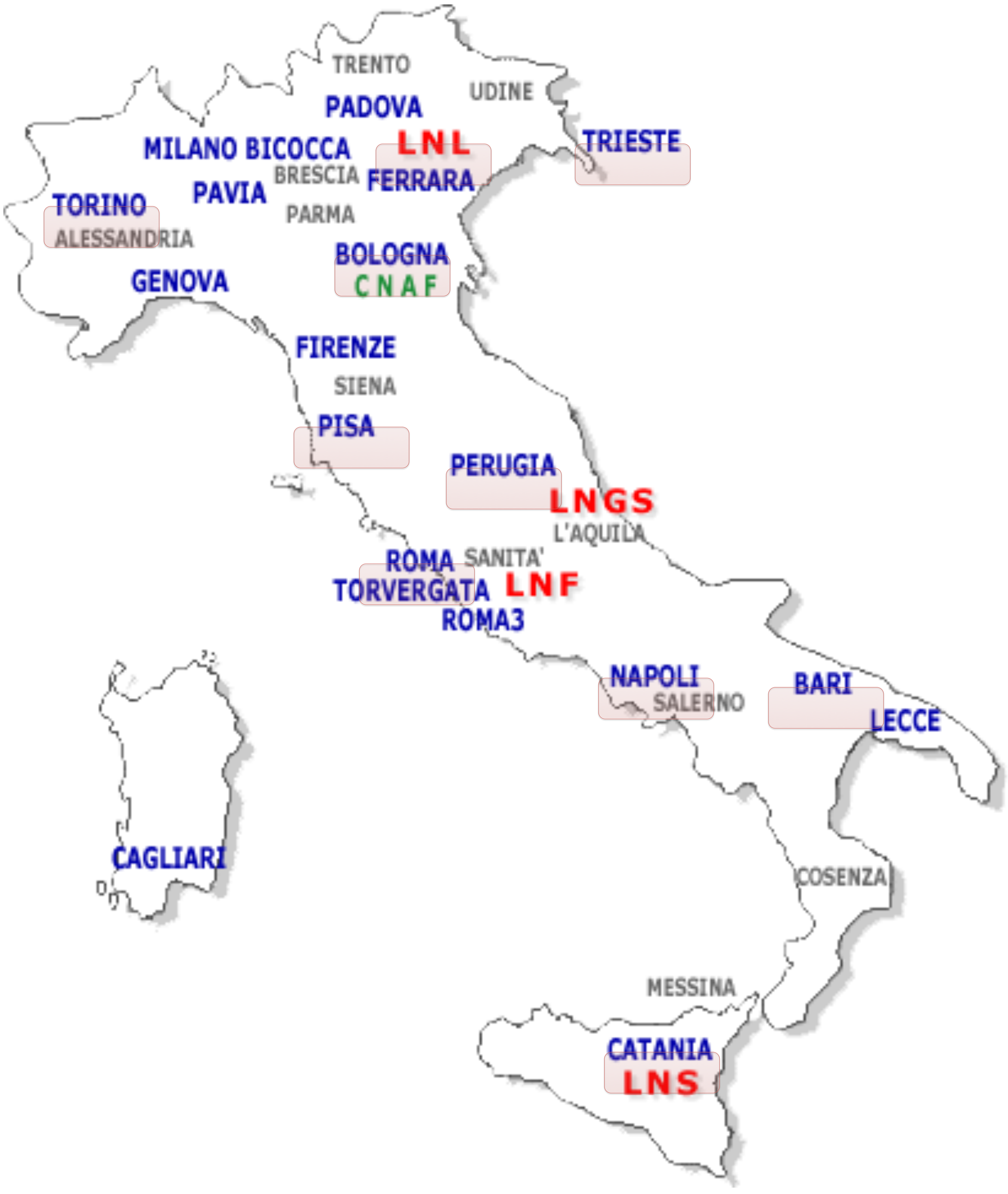


Dynamic provisioning on GRID and Cloud

In the job submission framework of the CMS experiment, resource provisioning is separate from resource scheduling. This is implemented by **pilot jobs**, which are submitted to the available Grid sites to create an overlay batch system where user jobs are eventually executed.
CMS is now exploring the possibility to use Cloud resources besides the GRID, basically considering the same architecture for what concerns the dynamic resource provisioning: instead of submitting pilot jobs, virtual machines (where the pilot jobs run) are created on demand.
At the Padova-Legnaro Tier2 a OpenStack Cloud based testbed has been set up, and here the model has been successfully demonstrated executing CMS CRAB analysis jobs.



The work has been partially funded under contract 20108T4XTM of 'Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale (Italy)'


TRENTO
UDINE
PADOVA
MILANO BICOCCA
LNL
TRIESTE
BRESCIA
FERRARA
PAVIA
TORINO
PARMA
ALESSANDRIA
BOLOGNA
CNAF
GENOVA
FIRENZE
SIENA
PISA
PERUGIA
LNGS
L'AQUILA
ROMA SANITA'
ROMA TORVERGATA
LNF
ROMA3
NAPOLI
SALERNO
BARI
LECCE
CAGLIARI
COSENZA
MESSINA
CATANIA
LNS