



# The Grid distributed data analysis at CMS



F.Fanzago – CERN/CNAF; F.Farina – INFN Milano Bicocca/CERN  
 M.Cinquilli – INFN Perugia; G.Codispoti – INFN Bologna; F.Fanfani – INFN Bologna  
 S.Lacaprarra – INFN LNL; E.Miccio – CERN/CNAF;  
 D.Spiga – Università Perugia/CERN; E.Vaandering - FNAL

## CMS computing model and grid infrastructure

### The CMS experiment

The Compact Muon Solenoid (CMS) is one of the four particle physics experiment at LHC.

It will collect data aiming to discover the Higgs boson.

A large amount of data will be produced (~2PB/year).

Data will be stored in many computing centres belonging to the CMS collaboration.

Data will be available for analysis to world-wide distributed physicists.

Analysis results shall be easily shared in the physics community

### Why the Grid

Huge amount of data to analyze

Large community of physicists which wants to access data

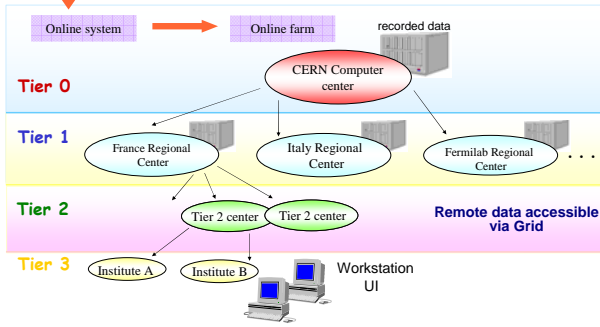
Many distributed sites where data will be stored and analyzed

CMS adopts a distributed architecture based on Grid infrastructure

- to ensure remote resources availability
- to assure remote data access to authorized user (belonging to CMS Virtual Organization)
- to distribute evenly the workload among computing sites



The CMS offline computing system is arranged in four Tiers which are geographically distributed



### Data distribution

During data acquisition data from detector that overhead different trigger level will be sent, stored and first step reconstructed at Tier-0.

Then they will be spread over some Tiers depending on the kind of physics data

The Grid infrastructure guarantees also enough computing power for simulation, processing and analysis data.

Until real data are not available, simulated data are used to study the detector response, and to get experience with management and analysis

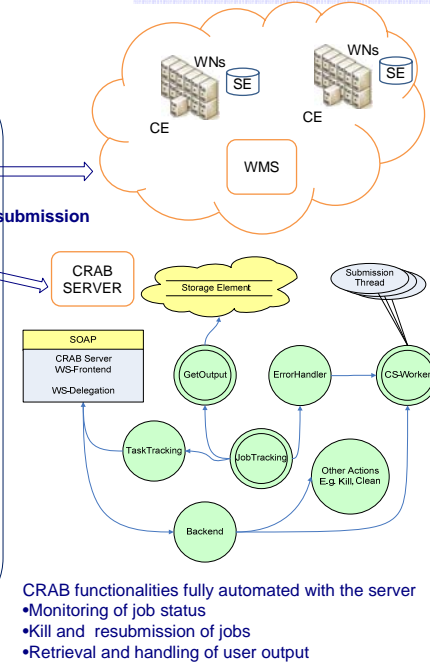
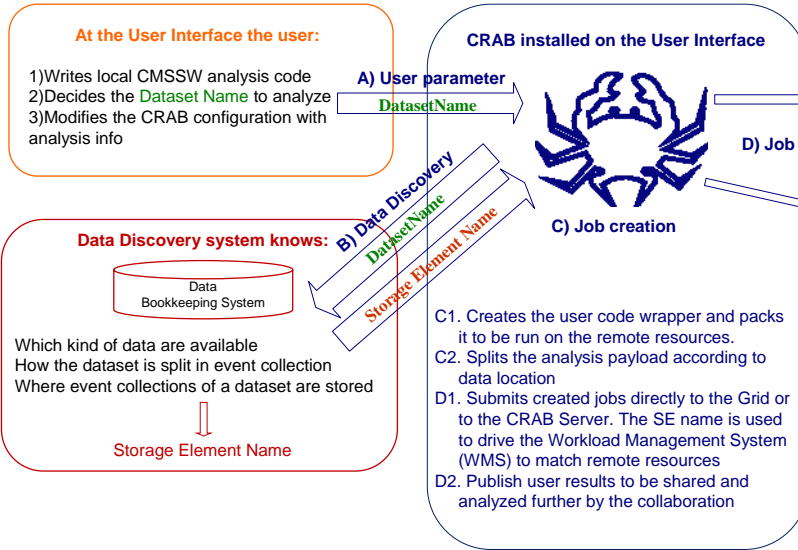
## CRAB and the CMS distributed analysis

CMS collaboration develops some tools, interfaced with Grid services provided by WLCG and OSG, to allow data analysis in a distributed environment

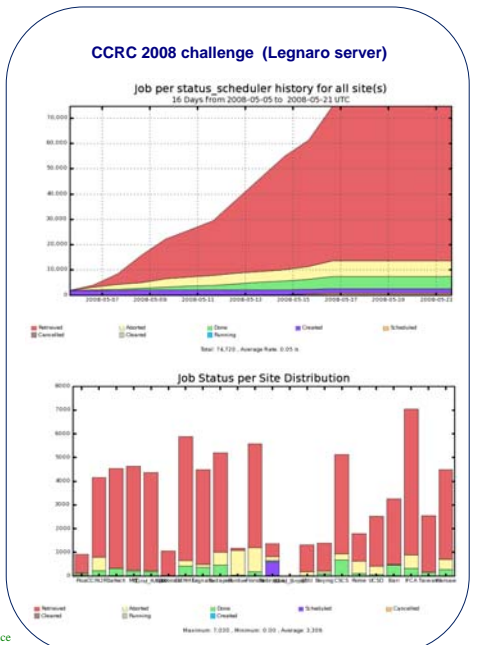
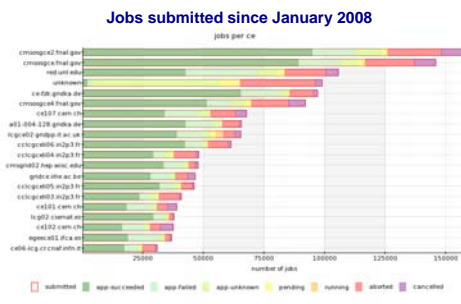
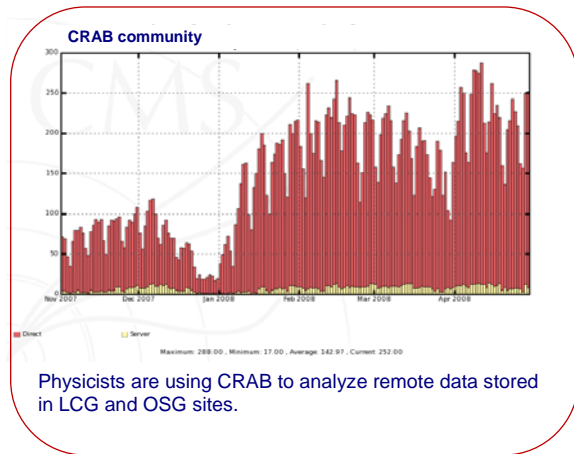
- Analysis software CMSSW
- Data distribution PhEDEx
- Data discovery DBS

CRAB (CMS Remote Analysis Builder) is a friendly interface to simplify the creation and the submission of analysis jobs to the Grid.

Its purpose is to allow users to orchestrate the Grid infrastructure to run the analysis code on remote data as easily as in a local environment and publish result.



## CRAB usage



>10<sup>7</sup> jobs were submitted to the Grid using the CRAB tool since its first release (April 2005)  
 When real data will be available the expected daily rate of submitted jobs will be ~100000.

CRAB tool is used to analyze remote data and to test distributed analysis chain also during CMS challenge

- Magnet test cosmic challenge
- CSA06
- CCRC 2008

CRAB proves that CMS users are able to use available Grid services and that the full analysis chain works in a distributed environment.

Data sources: CMS dashboard (<http://arda-dashboard.cern.ch>) and CRAB Server Monitoring Service