The activities carried out by the institutions involved in the project PRIN2008MHENNA (by MIUR, Italy) have been designed having user analysis activities in mind. All the aspects important to increase user satisfaction level in analysis have been considering, studying and deploying solutions to improve the experience.

This has included:

- Networking optimizations
- Storage optimizations
- Study on diverse hardware components
- Optimization to the facilities’ access
- Improvement in user support

Results

The Bari farm is able to handle up to 40Gb/s of aggregate bandwidth among storage server and WN

The Pisa activity was focused on two distinct aspects:
- Client side studies on remote data access in HEP analysis
- Improvement of user support

Client side remote data access:
- Real analysis tests in HEP environment (CMS 2011 7 TeV data)
- Testing, using as targets:
  - Local disk (for reference)
  - Remote Dcache systems (like INFN-PISA)
  - Remote Storm systems (like INFN-BARI)
- In the case of interactive analysis, the performance drop due to the data non-locality is of the order of 10%; we evaluated that increase in ease of analysis largely outweighs the lost performance
- A modified CRAB client (the tool for grid submision of analyses in CMS) is going to be used to test also the non interactive analysis part. This is going to happen by not forcing the jobs to land where data is, but by allowing them to be run where CPUs are available, with data access remotely

User Support improvements:
- The activities were focused on preparing user support infrastructures, specifically for the use case of analysis in CMS
- User support has been followed via a shared Gmail account, via CMS Hypernews system, and via the Savannah system. Just to cite a number, the Hypernews forum has seen 6000 threads opened in the last two years, for a total number of messages at least three times bigger.

At INFN-Bari the activities were focused on:
- Testing several software solutions: dCache, xrootd, lustre, GPFS
- Lustre preferred for performance, TCO, and management
- Improving read-head and caching policy to improve CPU usage for data analysis application
- Configuring network topology for minimizing the bandwidth needed on the edge switch uplink
- By distributing storage systems and worker nodes on the same switch it is possible to exploit both uplink and downlink in a more efficient way
- Distribution of data via WAN protocols: the Bari Tier2 is hosting the Xrootd redirector that federated all the data source available
- Monitoring the access to the Lustre distributed file system: All the clients are accessing file using standard posix interface so it is important to have a monitoring that provide information on which file is accessed from which client
- We build a monitoring system that stores into a DB all the information needed, like: the client, the path of the file, the name and the id of the process that is using the file and the relative timestamps
- Optimization of scheduling algorithms in order to provide fast and interactive access to the batch resources to end users running the CMS analysis

INFN-Milano Bicocca activities were focused on the distributed analysis aspect in HEP analysis activities, for the CMS experiment in particular.

CMS uses since the start of the pp collisions in LHC (late 2009) CRAB as main tool for submitting analysis jobs to the GRID. The version currently operational is CRAB2, which consists in:

- A server which takes care of interaction with the GRID (OSG and AAI)
- A client to be installed on the user’s User Interface
- A server which takes care of interaction with the GRID (DSG and gLite flavors being supported)

The client supports also server-less operations, with direct interaction to the GRID; the benefits introduced by the server are:
- Speed: the user experiences much faster commands, since the jobs are sent to the GRID in asynchronous mode
- Automatic smart resubmission of jobs in case of failure (for example, excluding sites with excessive failure rate)

The group in Milano Bicocca focused on one side on the maintenance of CRAB2, introducing bug fixes and new features as requested by the CMS Experiment (like special job splitting features). On the other, Milano worked on the development of the successor CRAB3, completely rewritten to use a common interface with the other tool in CMS submitting jobs to the GRID (the one used for Monte Carlo processing and data reprocessing, WMAgent). CRAB3 is now a thin client, which does not store any information locally (stateless), but communicates with the WMAgent infrastructure.
CRAB3 is supposed to enter into production phase later this year, after some months of extensive testing from expert users.