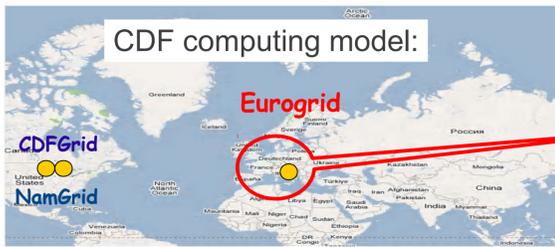


EUROGRID: a new glideinWMS based portal for CDF data analysis

S.Amerio^{1,6}, D.Benjamin², G.Compostella^{1,3,4}, J.Dost⁵, D.Lucchesi^{1,3}, I.Sfiligoi⁵

¹ INFN Padova (amerio@pd.infn.it), ² Duke University, ³ University of Padova, ⁴ Now with Max-Planck-Institut fuer Physik (MPG), ⁵ University of California San Diego
⁶This research was supported by a Marie Curie International Outgoing fellowship within the 7th European Community Framework Program

Introduction



Resources:

- dedicated farm at CNAF Tier1
- LCG resources in Italy, Spain, France and Germany

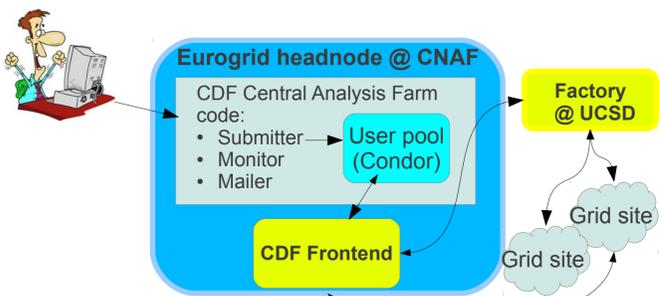
- Central farm @ FNAL (CDFGrid)
- Two access points to grid resources (NamGrid, Eurogrid)



Eurogrid is based on glideinWMS, the glidein-based workload management system.

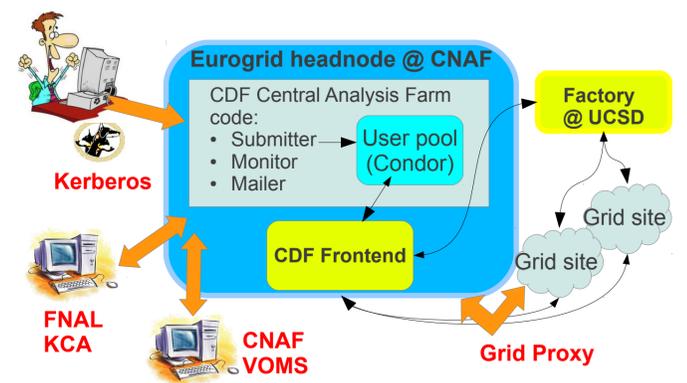
- GlideinWMS underlying batch system is based on Condor as the CDF Central Analysis Farm code (CAF) → easy integration and more efficient code management for the future.
- The use of pilot jobs masks site related errors or malfunctions from the job
- Minimum manpower is required for maintenance on the experiment side: CDF is responsible for the FrontEnd (FE) only, submission to the grid is taken care of by the Factory @ UCSD.

Job submission and execution



- CDF user sends job submission request through CAF UI
- The request is sent to Eurogrid headnode, prepared for Condor and submitted to the local Condor pool
- The FrontEnd polls the local Condor pool and makes requests for glideins to the Factory
- The Factory receives requests from the FE and submits glideins to grid sites
- Once a grid site matching job requirements is found, the CDF job wrapper on the site:
 - it fetches the job tarball;
 - it runs the job;
 - it regularly renews authentication
 - upon job completion, it sends the output to the user's area.

Authentication



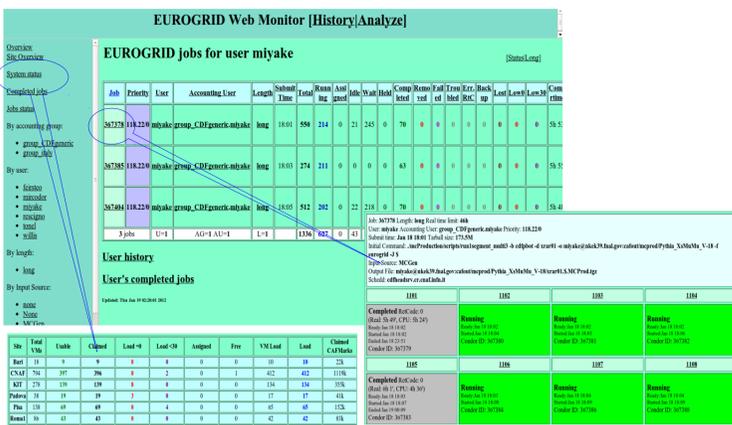
The headnode contacts:

- the Kerberized authentication authority (KCA) : from Kerberos ticket to X.509 Grid-like certificate;
- the CNAF Virtual Organization Membership Service (VOMS) to setup a valid proxy that can be used to submit to the grid.

Monitoring

Web monitor

It periodically requests info from monitoring daemons running on the headnode.



Command line monitor from user's desktop

```
lamer@pd01 ~$ cdfmon --farm eurogrid jobs
Analysis Farm: eurogrid Host: cdfheadsrv.cr.cnaf.infn.it
-----
Job Group From To Status
343684 Long 280 317 Completed
343684 Long 318 318 Running
343684 Long 319 399 Completed
-----
343684 Long Completed: 199 Running: 1
343684 Long Completed: 99% Running: 0%
-----
Completed: 199 Running: 1 Total: 200
Completed: 99% Running: 0%
```

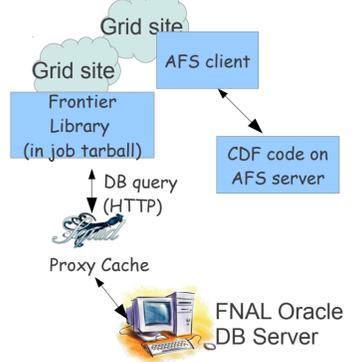
It exploits glideinWMS virtual monitoring machines on the nodes.

Users can have direct access on files on the worker node, with usual unix commands (ls, ps, cat,...)

Database access and code distribution

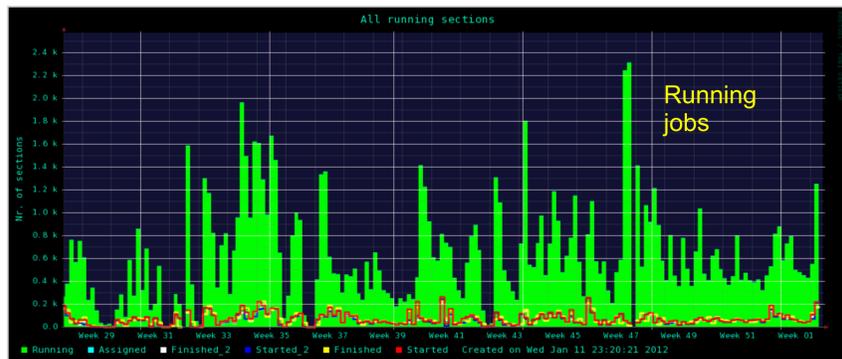
Database access at FNAL to retrieve run conditions

- DB queries are translated into HTTP requests using Frontier
- Squid proxies are used as caches to improve scalability and performance



CDF software available in most of Eurogrid sites through AFS

Performance



Very good response from users:

- On average 400 jobs running/day
- Used regularly by 40 CDF users
- On average 5 users/day
- The most used offsite CDF computing resource

Reliable:

- Negligible failure rate due to site related errors and malfunctioning.
- Three major downtimes (1 day) during the first year due to:
 - Maintenance work at CNAF;
 - Hardware failure;
 - Authentication problems;
- All jobs were safely resumed after the downtime.

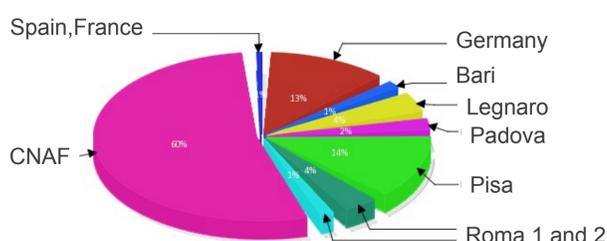
Matching resources and user needs



- GlideinWMS *match_expressions* are used to match jobs and sites
- glideinWMS custom scripts set the correct environment, compatibility libraries and kerberos libraries before running CDF job wrapper on the node.

Completely transparent to users

Used resources per site



Conclusions

Eurogrid is a new portal to access and analyze CDF data using European grid resources, official since June 2011.

It is based on glideinWMS, easy to install and maintain even for non experts → best solution in terms of required manpower for CDF future.

It got really good feedback from users: it is easy to use (nothing new with respect to other CDF job submission portals) and very reliable.

It will be maintained for long term CDF data preservation project.