21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015

Contribution ID: **386**   Type: **poster presentation**

Clusteralive

"CLUSTERALIVE"

"Clusteralive" is an integrated system developed in order to monitor and manage few important tasks in our HPC environment.

We have also other management systems, but now, with "Clusteralive" we can know immediately, just seeing our screen, if Clusters are up and running and we are sure that the most important functionality are well instanced.

"Clusteralive" is a php scripts suite able to monitor and perform tasks of automatic management about services, parameters and basic activities related in particular to computing nodes of the HPC clusters but also of the entire computing infrastructure.

At this link: http://farmsmon.pi.infn.it/clusteralive/monitoring.php you have a look at the specific application of "Clusteralive" dedicated to our principal projects.

In particular "Clusteralive" controls the cluster (dedicated to theoretical physics) called 'Zefiro' and funded by the SUMA project (SUper MAssive computing project, link: http://vh2.pi.infn.it/ a special project approved by Italian Research Ministery). Zefiro (2048 cores total: AMD Opteron 6380-2.5GHz) consists of 32 machines each one with 512 GB of RAM and 4 processors (16 cores for processors fo 64 cores total, grouped into 2 jobslot). Nodes are linked via Infiniband QDR connections operated with Mellanox IS5100 switch with 108 ports. The accesses are regulated by the IBM LSF (V.9) scheduler.

"Clusteralive" has also been recently extended to all the resources of the HPC in Pisa used for academic and industrial collaborations (more than 4000 computing cores total).

The monitoring system allows, via web browser, the view of essential informations about the status of each compute nodes and about a specific service and status of the entire HPC infrastructure.

For each computing node are displayed few informations like: the state (used / free / closed for maintenance), users that have assigned resources on it (used or reserved), information about the communication between the machines (ping) via Ethernet and via Infiniband, the percentage of used disk space, the used or reserved resources, the CPU load, the loked processes called zombie (due to a termination or to a bug in the application), daemons for user's autentication (nscd / nslcd), communication daemons (sshd) and node status daemons (gmond / hsflowd).

For the entire cluster the status of jobs on the specific queues are shown , specifically the running/pending and suspended jobs end in last two cases, it is possible to know reasons

for the specific status. The monitoring system also shows the physically turned off and turned on computing nodes and the nodes with the file system unmounted.

The automatic management system, has been developed for the recovery of some services in case of their malfunction. Specifically, an automatic restart of stopped and blocked services and an automatic closing system for all compute nodes in case of malfunction of the filesystem which supports the entire cluster HPC has been implemented togheter a system of automatic recovery of all the services necessary for communication between computing nodes and user identification.

An additional feature is going to be implemented and will permit the automatic cleaning of processes (zombie) after an unexpected termination or at the end of specific job.

Primary author: Mr CARUSO, Giuseppe (INFN - Pisa)

Co-authors: Dr CIAMPA, Alberto (INFN - Pisa); Mr FORMUSO, Antonino (INFN - Pisa); Mrs AREZZINI, Silvia (INFN - Pisa)

Presenter: Mrs AREZZINI, Silvia (INFN - Pisa)

Track Classification: Track8: Performance increase and optimization exploiting hardware features