Changing the batch system in a Tier 1 computing center: why and how

Andrea Chierici
INFN-CNAF
E-mail: andrea.chierici@cnaf.infn.it

Stefano Dal Pra
INFN-CNAF
E-mail: stefano.dalpra@cnaf.infn.it

Abstract. In this section we include some detailed informations and the most notable pictures that did not fit in the main paper due to space restrictions.

1. Grid Engine Setup
There are five different host types:

**Master Host** is the center of the cluster. This host runs the master daemon that stores all configuration data, runtime information and information about compute jobs. The scheduling component also resides on the master host and is responsible for all the planning tasks needed to distribute jobs into the cluster.

**Shadow Master host** Zero or more shadow master hosts can be setup in each cluster. This process provides backup functionality in case the master hosts fails.

**Submit Hosts** Submit hosts are used to submit jobs to GE and to control them. The master host is by default also a submit host.

**Admin Hosts** Operators and managers of GE can execute administrative commands on admin hosts. The master host is by default also an administrative host.

**Execution Hosts** Multiple execution hosts can exist in a cluster. Hosts running this process provide their compute resources to the corresponding cluster.

The daemons running on these hosts are (see figure 1a):

**sge_qmaster** Controls overall behavior of the cluster, holds most current state about jobs, queues and other GE objects, interoperates with sge_schedd for scheduling activities and delivers dispatched jobs to proper sge_execd daemons

**sge_schedd** Is the dedicated scheduling daemon, matching pending jobs to available resources and queues

**sge_execd** runs on execution host (worker node) and controls all running jobs on it, collecting job information and reporting it to sge_qmaster daemon
In our installation we share a network disk among the master servers and all the worker nodes (see figure 1b).

1.1. Classic vs BDB Spooling
Grid Engine supports two different spooling methods on the master host: classic spooling and BDB spooling (see figure 2).

2. How to make the switch
At the grid level (see figure 3a), we install Cream Computing Elements with GE extensions and point them to our Master Node, that shares the same WNs set. When all the WNs have turned to GE only job execution, its safe to remove the LSF client and turn the LSF CEs off. At worker node level (see figure 3b), both LSF and GE clients run at the same time on the node and the number of slots is modified in order for GE to take over the LSF slots. An alternative approach is to allocate a subset of WNs on the farm to GE, once all the LSF jobs on them are terminated.
3. **Slurm**

SLURM implements a fairly traditional cluster management architecture, rather similar to the one we found in LSF and GE (see figure 4).