



Contribution ID: 387

Type: poster presentation

## Managing competing elastic Grid and Cloud scientific computing applications using OpenNebula

Elastic cloud computing applications, i.e. applications that automatically scale according to computing needs, work on the ideal assumption of infinite resources. While large public cloud infrastructures may be a reasonable approximation of this condition, scientific computing centres like WLCG Grid sites usually work in a saturated regime, in which applications compete for scarce resources through queues, priorities and scheduling policies, and keeping a fraction of the computing cores idle to allow for headroom is usually not an option. In our particular environment one of the applications (a WLCG Tier-2 Grid site) is much larger than all the others, so a possible strategy is to make it shrink and release resources when smaller, higher priority ones require them.

The implementation of said model in our infrastructure, based on the OpenNebula cloud stack, will be described and the very first operational experiences with a small number of strategies for timely allocation and release of resources will be discussed.

Such strategies, which aim to be non-invasive with respect to our most common class of applications, i.e. Grid jobs, include for example the tuning of virtual Worker Node parameters (number of cores and lifetime) to match the statistical distribution of job durations, or the balance between the amount of resources statically pinned to an application and the amount left available for competitive seizing by elastic applications.

**Primary authors:** BERZANO, Dario (CERN); MASERA, Massimo (Universita e INFN (IT)); VALLERO, Sara (Universita e INFN (IT)); Dr BAGNASCO, Stefano (I.N.F.N. TORINO); LUSSO, Stefano (INFN-TO)

**Presenter:** Dr BAGNASCO, Stefano (I.N.F.N. TORINO)

**Track Classification:** Track7: Clouds and virtualization