A multi-group and preemptive computing resource scheduling system based on HTCondor

Tuesday 11 October 2016 16:30 (15 minutes)

Virtual machines have many features —flexibility, easy controlling and customized system environments. More and more organizations and enterprises begin to deploy virtualization technology and cloud computing to construct their distributed system. Cloud computing is widely used in high energy physics field. In this presentation, we introduce an integration of virtual machines with HTCondor, which support resource management of multiple groups and preemptive scheduling policy. The system makes resources management more flexible and more efficient. Firstly, computing resources belong to different experiments, and each experiment has one or more user groups. All users of a same experiment have the access permission to all the resources owned by that experiment. Therefore, we have two types of groups, resource group and user group. In order to manage the mapping of user group and resource group, we design a permission controlling component to ensure jobs are delivered to suitable resource groups. Secondly, for elastically adjusting the resource scale of a resource group, it is necessary to schedule resources in the way of scheduling jobs. So we design a resource scheduler that focusing on virtual resources. The resource scheduler maintains a resource queue and matches an appropriate amount of virtual machines from the requested resource-group. Thirdly, in some conditions, one case that the resource may be occupied by a resource-group for a long time, it needs to be preempted. This presentation adds the preemptive feature to the resource scheduler based on the group priority. Higher priority leads to lower preemption probability, and lower priority leads to higher preemption probability. Virtual resources can be smoothly preempted, and running jobs are held and re-matched later. The feature is based on HTCondor, storing the held job, releasing the job to idle status and waiting for a secondary matching. We built a distributed virtual computing system based on HTCondor and Openstack. This presentation also shows some use cases of the JUNO and LHAASO experiments. The result shows that multi-group and preemptive resource scheduling perform well. Besides, the permission controlling component are not only used in virtual cluster but also in the local cluster, and the amount of experiments which it supports are expanding.

Tertiary Keyword (Optional)

High performance computing

Secondary Keyword (Optional)

Distributed workload management

Primary Keyword (Mandatory)

Cloud technologies

Primary author: JIANG, xiaowei

Co-authors: ZOU, Jiaheng (IHEP); SHI, Jingyan (IHEP); CHENG, Yaodong (IHEP)

Session Classification: Posters A / Break

Track Classification: Track 7: Middleware, Monitoring and Accounting