

Opportunistic usage of the CMS online cluster using a cloud overlay

Friday, March 18, 2016 9:00 AM (30 minutes)

After two years of maintenance and upgrade, the LHC (Large Hadron Collider), the largest and most powerful particle accelerator in the world, has started its second three year run. Around 1500 computers make up the CMS (Compact Muon Solenoid) online cluster. This cluster is used for Data Acquisition of the CMS experiment at CERN, selecting and sending to storage around 20 TBytes of data per day that are then analysed by the WLCG (Worldwide LHC Computing Grid) infrastructure that links hundreds of data centres worldwide. 3000 CMS physicists can access and process data, and are always seeking more computing power and data. The backbone of the CMS online cluster is composed of 16000 cores which provide as much computing power as all CMS WLCG Tier 1 sites (352K HEP-SPECHS06 score in the CMS cluster versus 300K across Tier 1 sites). The computing power available in the CMS cluster can significantly speedup the processing of data, so an effort has been made to allocate the resources of the CMS online cluster to the GRID when it isn't used to its full capacity for data acquisition. This occurs during the maintenance periods when the LHC is non-operational, which happens one week out of every six. During 2016, the aim is to increase the availability of the CMS online cluster for data processing by making the cluster accessible during the time between two physics collision while the LHC and beams are being prepared. This is usually the case for a few hours every day, which would vastly increase the computing power available for data processing. Work has already been undertaken to provide this functionality, as an Openstack cloud layer has been deployed as a minimal overlay that leaves the primary role of the cluster untouched. This overlay also abstracts the different hardware and networks that the cluster is composed of. The operation of the cloud (starting and stopping the virtual machines) is another challenge that has been overcome as the cluster has only a few hours spare during the aforementioned beam preparation. By improving the virtual image deployment and integrating the Openstack services with the core services of the Data Acquisition on the CMS Online cluster it is now possible to start a thousand virtual machines within 10 minutes and to turn them off within seconds. This presentation will explain the architectural choices that were made to reach a fully redundant and scalable cloud, with a minimal impact on the running cluster configuration while giving a maximal segregation between the services. It will also present how to speed up of 25 the cold starting of 1000 virtual machines, using tools commonly utilised in all data centres.

Summary

The Compact Muon Solenoid (CMS) experiment at CERN is maintaining a large infrastructure to readout and filter the data from the detector. The High Level Trigger (HLT) online cluster is composed of 16000 cores dedicated to online data and event filtering. However, this resource is used only for about 30% of the time, due to the Large Hadron Collider (LHC) duty cycle and the various maintenance periods thus the rest of time it is free. Only during these unused times an OpenStack cloud is started on top of the cluster allowing it to join the Worldwide LHC Computing Grid (WLCG) for offline data analysis.

Primary author: Mr CHAZE, Olivier (CERN)

Presenter: Mr CHAZE, Olivier (CERN)

Session Classification: Infrastructure Clouds and Virtualisation Session II

Track Classification: Infrastructure Clouds and Virtualisation