Dynamic resource provisioning of the CMS online cluster using a cloud overlay

Monday 10 October 2016 12:15 (15 minutes)

During the past years an increasing number of CMS computing resources are offered as clouds, bringing the flexibility of having virtualised compute resources and centralised management of the Virtual Machines (VMs). CMS has adapted its job submission infrastructure from a traditional Grid site to operation using a cloud service and meanwhile can run all types of offline workflows. The cloud service provided by the online cluster for the Data Acquisition (DAQ) and High Level Trigger (HLT) of the experiment was one of the first facilities to commission and deploy this submission infrastructure. The CMS HLT is a considerable compute resource. It consists currently of approximately 1000 dual socket PC server nodes with a total of ˜25 k cores, corresponding to ˜500 kHEPSpec06. This compares to a total Tier-0 / Tier-1 CMS resources request of 292 / 461 kHEPSpec06. The HLT has no local mass disk storage and is currently connected to the CERN IT datacenter via a dedicated 160 Gbps network connection.

One of the main requirements for the online cloud facility is the parasitic use of HLT which shall never interfere with its primary function as part of the data acquisition system. Hence a design has been chosen where an Openstack infrastructure is overlaid over the HLT hardware resources. This overlay also abstracts the different hardware and networks that the cluster is composed of. The online cloud is meanwhile a well established facility to substantially augment the CMS computing resources when the HLT is not needed for data acquisition, such as during technical stop periods of the LHC. In this static mode of operation, this facility acts as any other Tier-0 or Tier-1 facility. During high workload periods it provided up to ˜40% of the combined Tier-0/Tier-1 capacity, including workflows with demanding I/O requirements. Data needed by the running jobs was read from the remote EOS disk system at CERN and and data produced was written back out to EOS. The achieved throughput from the remote EOS came close to the installed bandwidth of the 4x40 Gpbs long range links.

The next step is to extend the usage of the online cloud to the opportunistic usage of the periods between LHC fills. These periods are a-priori unscheduled and of undetermined length, typically at least 5 hours and once or more a day. This mode of operation of a dynamic usage of the cloud infrastructure requires a fast turn-around for the starting and stopping of the VMs. A more advanced mode of operation where the VMs are hibernated and jobs are not killed is also being explored. Finally, one could envisage to ramp up VMs while the load on the HLT reduces towards the end of the fill. We will discuss the optimisation of the cloud infrastructure for the dynamic operation and the design and implementation of the mechanism in the DAQ system to gracefully switch from DAQ mode to providing cloud resources based on LHC state or server load.

Secondary Keyword (Optional)

DAQ

Primary Keyword (Mandatory)

Cloud technologies

Tertiary Keyword (Optional)

Author: MEIJERS, Frans (CERN)

Co-author: DOBSON, Marc (CERN)

Presenter: DOBSON, Marc (CERN)

Session Classification: Track 6: Infrastructures

Track Classification: Track 6: Infrastructures