Multivariate Analysis of Variance for High Energy Physics Software in Virtualized Environments

Friday 20 Mar 2015 at 11:50 (00h20')

Content:
Cloud Computing is emerging today as the new approach followed by most of computing centres, since the flexibility the Cloud provides is a powerful component to manage their resources. Through the use of virtualization, cloud promise to address, with the same shared set of physical resources, a large user base with different needs. However, virtualization may induce significant performance penalties for the demanding scientific computing workloads. This work presents an evaluation of the usefulness of KVM (Kernel-based Virtual Machine) cloud computing services for High Energy Physics (HEP) applications. The performance of several HEP software running in a KVM under different set-ups is deeply analysed with the use of a multivariate analysis of variance (MANOVA). Our experimental design includes three different aspects: the treatment design with factors such as the number of cores, the memory size, the job length and the software type; the error-control, with software random seeds and multiple executions; and the sampling and data taking, for maximizing the use of CPU resources.

Three HEP software have been selected to make the study. The first one is GAUSS, a Montecarlo simulation program developed by the LHCb collaboration, which includes GEANT4, a software for the simulation of the passage of particles through matter. The second one is AIRES, which was developed for the Auger experiment with the objective of simulating the production of particle showers created by the collision of ultra high energy particles present in cosmic rays with the Earth’s atmosphere. Finally, the last one is a software developed by a group in the University of Santiago de Compostela to study the contamination of jets by background particles in heavy ion collisions. This software makes use of FastJet, which is a software package for jet finding in pp and e+e− collisions, used by major HEP experiments like ATLAS, CMS, D∅, CDF, and others.

Preliminary results of our study had shown a performance lost of about 30% for GEANT4, which was related with the libvirt CPU flags in AMD processors. These results
have motivated us to make a formal statistical study of the influence of these parameters in the software performance when running in virtualized environments. The main conclusion of our work is that, as cloud services gain maturity in HEP experiments, the correct configuration of cores and memory and the set-up of the KVM hypervisor will be an important decision in order to reach an acceptable performance for the specific HEP-Software. Some guidances on how to correctly set these configurations are presented in this work.

**Primary authors**: Mr. FERNANDEZ, Victor (Santiago de Compostela University)

**Co-authors**: Mr. GRACIANI, Ricardo (Universidad de Barcelona) ; Mr. SABORIDO, Juan (Santiago de Compostela University) ; Mr. SECO, Marcos (Santiago de Compostela University) ; Mr. MENDEZ, Victor (Universidad Autónoma de Barcelona) ; Mr. FERNANDEZ, Tomas (Santiago de Compostela University)

**Presenter**: Mr. FERNANDEZ, Victor (Santiago de Compostela University)

**Session classification**: Physics & Engineering

**Track classification**: Physics (including HEP) and Engineering Applications

**Type**: Oral